

Known unknowns: assessing the impact of instrumental calibration uncertainty on LISA science

Etienne Savalle,¹ Jonathan Gair,² Lorenzo Speri,² and Stanislav Babak¹

¹*AstroParticule et Cosmologie (APC), Université Paris Cité/CNRS, 75013 Paris, France**

²*Max Planck Institute for Gravitational Physics (Albert Einstein Institute),
Am Mühlenberg 1, Potsdam-Golm 14476, Germany[†]*

The primary scientific results of the future space-based gravitational wave interferometer LISA will come from the parameter inference of a large variety of gravitational wave sources. However, the presence of calibration errors could potentially degrade the measurement precision of the system parameters. Here, we assess the impact of calibration uncertainties on parameter estimation for individual sources, focusing on massive black holes, extreme-mass-ratio inspirals (EMRIs), galactic binaries, and stellar origin black hole binaries. Using a Fisher matrix formalism, we investigate how the measurement precision of source parameters degrades as a function of the size of the assumed calibration uncertainties. If we require that parameter measurements are degraded by no more than a factor of two relative to their value in the absence of calibration error, we find that calibration errors should be smaller than a few tenths of a percent in amplitude and 10^{-3} in phase. We also investigate the possibility of using verification binaries and EMRIs to constrain calibration uncertainties. Verification binaries can constrain amplitude calibration uncertainties at the level of a few percent, while both source types can provide constrain phase calibration at the level of a few $\times 10^{-2}$.

I. INTRODUCTION

LISA is a space-based gravitational wave observatory selected by ESA to be the third large mission in the Cosmic Vision programme and will be launched around 2034. Recently it has completed phase A (mission formulation phase) and is now entering phase B1 (preliminary design), which will lead to mission adoption around 2024. LISA will consist of three identical satellites in a heliocentric orbit forming an equilateral triangle with sides of length 2.5 million km. The centre of the constellation will be trailing about 20 deg. behind the Earth. LISA will detect gravitational waves (GWs) by measuring the proper distance between free-falling test masses (located inside each spacecraft) using laser interferometry. The change in the proper distance in the LISA frame (which can be seen as transverse-traceless gauge) translates into a modulation of the frequency (or phase) of the laser light and is directly measured by the phasemeters.

The dominant noise component in LISA measurements is the laser frequency noise. Even with frequency pre-stabilization this is expected to be several orders of magnitude higher than what is required to detect gravitational wave (GW) signals. This problem is further exacerbated by the fact that LISA's arms are not of equal length, and in fact change over time as the constellation "breathes" due to the shape of the orbits of the individual satellites. Fortunately, the laser frequency noise can be cancelled in post-processing. LISA will use *transponding interferometry*, in which incoming laser light is amplified, phase-locked and then transmitted back. The phase of

the light in each of the 6 laser links (one in each direction along each arm of the constellation) will be measured on the satellites using phasemeters. These phase measurements can be delayed in time and linearly combined to form closed loops of equal optical path length. Subtracting such combinations from each other cancels the laser frequency noise to an acceptable level. This technique is called Time Delay Interferometry [1], and for it to work we need to know the distances between the three spacecrafts accurately and we need to synchronize their on-board clocks.

Data decimation and post-processing (that includes filtering and resampling the data during clock correction and building TDI) is a complex procedure which leads to a non-trivial response of the detector to the incident GW signal. Inaccurate modelling of the signal chain could potentially distort the GW signal in the data that is eventually analysed. We will refer to differences between the actual GW signal present in the data and how we model it as calibration error (or mis-calibration). The calibration process, by which the mapping between an incident physical signal and a detector output is learnt or measured, plays an important role in physical experiments. For example, in the LIGO detector what is actually measured is the size of the electrostatic force that must be applied to the end mirrors to keep the detector at a dark interference fringe. The transformation between the force (or voltage) and the incident GW strain requires a calibration process, in which the parameters of a calibration model are constrained by injecting monochromatic signals into the detector using laser modulation and measuring the system's response [2]. Inevitably, the calibration cannot be perfect and leads to some residual uncertainty that impacts our ability to measure the parameters of the incident GW signal. In the case of LISA we will have more direct measurement of the GW strain and so we do not

* etienne.savalle@apc.in2p3.fr

† jgair@aei.mpg.de

expect to require a complex calibration procedure, or to have a large calibration error. However, it is still important to estimate the possible impact of any remaining calibration errors on our ability to detect and estimate the parameters of sources in the data.

In this analysis, we will treat the GW signal in the frequency domain and decompose the calibration error into parts acting on the amplitude and on the phase. The calibration error can be seen as a biased measurement of the amplitude and/or phase, and this bias could be both time and frequency dependent. In practice we will not know this bias, but it should be possible to bound it by an envelope. Here we will place limits on the size of that envelope so that we ensure LISA science exploitation is not affected. Ignoring calibration error will lead to biases in parameter estimation, similar to the biases that can appear due to errors in waveform models [3]. In common with that case, these biases will tend to be independent of the amplitude of the gravitational wave signal and hence become proportionally more important and even limiting for loud signals. In contrast to the waveform error case, time-independent calibration errors will affect all GW signals in the same way, which will have implications for inference on populations.

We adopt a particular model for the LISA calibration uncertainties, which is similar to models previously employed in the analysis of LIGO data [4]. This model is based on a natural cubic spline defined at four frequency knots, chosen to roughly split the full LISA band into low-frequency (0.1 – 1 mHz), mid-frequency (1 – 10 mHz) and high-frequency (10 – 1000 mHz) ranges. We place a prior on the weight at each knot of the spline that is a Gaussian distribution with zero mean and a width which characterises the calibration error envelope.

We use this spline model to determine the impact of calibration uncertainties on detection and parameter estimation of LISA sources, focusing primarily on the latter. We use a Fisher matrix formalism to assess our ability to simultaneously measure the calibration uncertainty and the source parameters, including the Gaussian prior on the calibration parameters to represent our knowledge of the calibration envelope. By changing the variance of the Gaussian prior we can identify the size of uncertainty at which calibration uncertainties become limiting. To make this concrete, we define this as the point at which the parameter estimation uncertainty doubles relative to its value in the absence of calibration uncertainty. We consider all classes of resolvable source expected to be observed by LISA — galactic binaries, (super-) massive black hole binaries, stellar-origin black hole binaries (SBHB) and extreme-mass-ratio inspirals (EMRIs). Galactic binaries are binary systems in the Milky Way containing two compact objects, typically white dwarfs (WDs), but also neutron stars (NSs) and black holes (BHs), with orbital periods of the order of an hour. The GW signals are nearly monochromatic, modulated by the LISA motion. Some of these systems are already known through electromagnetic observations, and are termed

verification binaries. For these binaries the amplitude and phase evolution of the systems are known to some extent, albeit relatively poorly. Massive black hole binaries are the strongest sources expected in the LISA data. These signals arise from the inspiral and merger of massive black holes in the centres of galaxies, following mergers between their host galaxies. The GW signals are broadband and are in the LISA band for between a few hours and a month, depending on the masses of the binary components. Stellar-origin black hole binaries are binaries of two black holes, of the type being observed merging by ground-based gravitational wave detectors. These systems are observable in the LISA band in the early-inspiral stage, between a few and a few hundred years before merger. Finally, EMRIs are the inspirals of stellar-origin compact objects (typically BHs, but possible NS or WD) into massive black holes in the centres of galaxies. These arise following the capture of a compact object as a result of relaxation process in the stellar cluster surrounding the massive black hole. These sources can be observed in LISA data for up to several years, giving the potential for ultra-precise measurements of the parameters of the system by tracking the GW phase over $\sim 10^6$ cycles.

For all source types, we find that calibration uncertainties typically become important when they are comparable to the precision with which the source parameters can be determined. Amplitude calibration uncertainties become limiting, although phase uncertainties do not in most cases, since the phase evolution of individual sources is typically distinct from the slowly-varying calibration uncertainty we assume here. We find that if the amplitude calibration is at the level of 0.1% and phase calibration is at the level of 10^{-3} then parameter estimation will not be adversely affected. The paper is organised as follows. In Section II we describe the formalism we will use to set calibration errors, including the model we use for calibration uncertainties and how we can assess the impact of calibration errors on both source detection and parameter estimation. In Section III we assess the impact of calibration uncertainties on parameter estimation for all resolvable LISA source types and discuss qualitatively how these results can be extrapolated to understand the implications for population inference. Finally, in Section IV we discuss how LISA GW sources can be used as calibrators to measure the LISA calibration uncertainty, before finishing in Section V with a summary of our conclusions.

II. SETTING CALIBRATION REQUIREMENTS

The typical assumption used in data analysis of gravitational wave (and other) detectors is that the observed data stream, $d(t)$, is a linear combination of signal and noise

$$d(t) = s(t) + n(t).$$

We assume a model for how the signal component of the data, $s(t) = h(t|\vec{\theta})$, depends on the physical parameters, $\vec{\theta}$, that characterise the system and which we wish to determine. The likelihood for the observed data is then $p(d(t)|\vec{\theta}) = p(n(t) = d(t) - h(t|\vec{\theta}))$. In a gravitational wave context it is usual to further assume that the instrumental noise follows a Gaussian distribution characterised by a power spectral density, $S_h(f)$. In that case the likelihood becomes

$$p(d(t)|\vec{\theta}) \propto \exp \left[-\frac{1}{2} (d(t) - h(t|\vec{\theta}) | d(t) - h(t|\vec{\theta})) \right], \quad (1)$$

where

$$(a(t)|b(t)) = 4\text{Re} \int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_h(f)} df. \quad (2)$$

In reality this model will be imperfect. The noise distribution might not be Gaussian, or the PSD might be different to that assumed. We will call this **noise modelling uncertainty**. Additionally the response might be different to what we have modelled, which can be represented by writing $\tilde{s}(f) = C(f)\tilde{h}(f|\vec{\theta})$, with $C(f) \neq 1$, and then

$$\tilde{d}(f) = C(f)\tilde{h}(f|\vec{\theta}) + \tilde{n}(f). \quad (3)$$

We will call the existence of $C(f) \neq 1$ a **calibration error**. While both noise modelling uncertainty and calibration error can have an impact on scientific inference and must be minimized, in this paper we will focus on understanding the latter.

The above expression is not completely general, since it assumes that the data stream depends linearly on the incident gravitational wave field and the function $C(f)$ does not depend on the properties of the gravitational wave sources in the data. Non-linearities that scale like the square of the gravitational wave amplitude will be sub-dominant, but there could be effects, for example filtering errors or the presence of data gaps, that lead to leakage between frequencies such that $\tilde{d}(f)$ depends on $h(f'|\vec{\theta})$ for $f' \neq f$. Calibration errors could also arise differently in different spacecraft, which would mean the calibration errors in the TDI channels would depend on the sky locations of the sources. An additional complication arises because the LISA data stream will contain multiple sources of different types, and so in reality

$$\tilde{h}(f) = \sum_{i=1}^m \tilde{h}_i(f|\vec{\theta}_i) \quad (4)$$

where i enumerates the different sources, with parameters $\vec{\theta}_i$ and waveforms $\tilde{h}_i(f|\vec{\theta}_i)$. As these sources will occur at different times, they may be affected by slightly different calibration errors. Some of these effects can be accounted for by writing down the calibration error in a different way. For dealing with data gaps, a model

similar to Eq. (3) but applied to the time domain data might be more appropriate, since gaps are multiplicative filters in the time domain, whereas they are convolutions in the Fourier domain. To handle the presence of multiple sources in the data, a model that includes more than one calibration error function could be used, for example

$$\tilde{d}(f) = C(f) \sum_{i=1}^m C_i(f)\tilde{h}_i(f) + \tilde{n}(f). \quad (5)$$

Here $C(f)$ and $C_i(f)$ can be interpreted as the global and source- i specific calibration error respectively.

Regardless of how complex the true dependence of the data stream on the signal parameters is in reality, we can always define a total calibration error $C(f|\{\vec{\theta}_i\})$ as the ratio of the signal-dependent part of the instrumental data stream to the model assumed for that component of the data. Doing this would introduce a dependence into $C(f)$ on the parameters of the sources in the data. In this paper we will ignore these complications, and use model (3) to set limits on how much $C(f)$ can deviate from unity before calibration error starts to impact scientific inference for individual sources. These results can then be used with more complete models of the LISA data to assess if the calibration requirement is met. Provided that the modelled $C(f)$ lies within the derived calibration envelope, for all sources and for any reasonable realisation of the full population of sources in the data, the conclusion that science is unaffected should be robust. If the calibration requirement is not satisfied, using a more complex model for the calibration error might partially mitigate the problem. However, those more complex models are not needed for setting a baseline requirement which is the goal of this paper.

Eq. (3) applies to a single data stream, but in reality LISA data will comprise three separate TDI data channels, (A, E, T). The response in each one can be modelled in a similar way, but an assumption must be made as to whether the calibration uncertainties are independent or correlated between different channels. There are other TDI combinations that can be constructed and it may be that certain TDI combinations can be better represented by the calibration model assumed here than others. For the purpose of this paper we will use (A, E, T) and will additionally assume that the calibration error is common to the three channels. This should be a conservative assumption as it makes the calibration error maximally degenerate with the signal waveform.

Calibration errors affect our ability to identify sources in the data (see Section IIB), and our ability to accurately measure the parameters of the sources (see Section IIC). The latter drives the LISA calibration requirement since the relatively small fluctuations in LISA event rates arising from calibration uncertainties are much smaller than intrinsic astrophysical uncertainties in the LISA source population. However, for completeness, we will describe in the following sections how to assess the impact of calibration errors on both. Before doing that,

in Section II A, we will first describe the calibration error model that we will use to produce later results in this paper.

A. Calibration model

In practice, the statistical properties of the calibration error might be estimated through simulation or measurements on board the LISA satellites. Given a distribution of calibration errors these can be folded into scientific inference through, for example, reweighting of posterior samples computed ignoring calibration error¹. For the purpose of setting/assessing calibration requirements we propose to take a different approach and write down a simple model for $C(f)$, the parameters of which can be interpreted as the accuracy of (amplitude and phase) calibration at different reference frequencies. First we decompose

$$C(f) = (1 + \delta A(f)) \exp(2\pi i \delta \phi(f)),$$

in which $\delta A(f)$ can be interpreted as the amplitude calibration uncertainty at frequency f , and $\delta \phi(f)$ can be interpreted as the corresponding phase calibration uncertainty. We further assume that these uncertainties are smoothly varying over a relatively wide range of frequency. At frequencies below 1mHz, the LISA noise is dominated by acceleration noise, while above 10mHz it is dominated by noise in the optical metrology system. The region 1–10mHz is a cross-over region [6]. These frequencies define the scale over which we would expect the LISA calibration error to vary. We therefore model them as natural cubic splines in $\log_{10}(f)$, with nodes at the frequencies

$$\{0.1, 1, 10, 1000\}.$$

In this model, between any two knots (f_i and f_{i-1}) the amplitude or phase calibration error function can be written as:

$$\delta X(f) = M_{i-1} \frac{(f_i - f)^3}{(f_i - f_{i-1})} + M_i \frac{(f - f_{i-1})^3}{(f_i - f_{i-1})} \quad (6)$$

$$+ \left(\delta X_{i-1} - \frac{1}{6} M_{i-1} (f_i - f_{i-1})^2 \right) \frac{f_i - f}{f_i - f_{i-1}} \quad (7)$$

$$+ \left(\delta X_i - \frac{1}{6} M_i (f_i - f_{i-1})^2 \right) \frac{f - f_{i-1}}{f_i - f_{i-1}} \quad (8)$$

where the four amplitude calibration parameters will be noted as $\{\delta A_0, \delta A_1, \delta A_2, \delta A_3\}$ and the four phase calibration parameters as $\{\delta \phi_0, \delta \phi_1, \delta \phi_2, \delta \phi_3\}$. The variables

¹ This is the approach that is now being taken in the analysis of data from the ground-based gravitational wave detectors. See, for example [5].

M_i are unknowns that can be easily evaluated by solving a system of equations defined by the boundary conditions imposed on the interpolation function. To avoid calibration-error-induced biases we assume that inference on the LISA data will include the parameters describing the calibration error as additional model parameters to be marginalised over during inference. We represent our knowledge of the calibration accuracy by the priors on the spline model. For simplicity we will use Gaussian uncertainties σ_C , and assume these are independent between the amplitude and phase errors and between different frequencies. Example splines constructed in this way are shown in Figure 1.

B. Impact of calibration error on source detection

It can be seen from Eq. (3) that the effect of calibration error is to modify the effective signal component of the data. As such, it is clear that this will have some bearing on our ability to identify sources in the data. The methods by which sources will be identified in the LISA data are not yet fully specified, but for ground-based detectors source detection relies mostly on *matched filtering* [7]. A *bank of templates* is constructed, which are waveforms, $h(t|\vec{\theta}_i)$, for a set of parameter choices, $\vec{\theta}_i$, for potential sources. The waveforms are then *normalised*,

$$\hat{h}(t|\vec{\theta}) = \frac{h(t|\vec{\theta})}{\sqrt{(h(t|\vec{\theta})|h(t|\vec{\theta}))}}, \quad (9)$$

where the inner product $(a(t)|b(t))$ was defined in Eq. (1). The number of templates in the bank is chosen to ensure a sufficiently dense coverage of parameter space. The maximum *signal-to-noise ratio* (SNR) across the template bank

$$\hat{\rho} = \max_{\vec{\theta}} (d(t)|\hat{h}(t|\vec{\theta})),$$

is computed and compared to a threshold. Events above the threshold are regarded as potential gravitational wave candidates that are subsequently followed up. In the presence of calibration errors, the search would be constructed and carried out in the same way, but the observed data would be modified according to (3). We assume that the observed data contains a source with true parameters $\vec{\theta}_0$, true signal-to-noise ratio $\rho_0^2 = (h(t|\vec{\theta}_0)|h(t|\vec{\theta}_0))$ and has a true frequency-domain calibration uncertainty $C_0(f)$, so that $\tilde{d}(f) = \rho_0 C_0(f) \hat{h}(f|\vec{\theta}_0) + \tilde{n}$. Expanding the likelihood in $\Delta \vec{\theta} = \vec{\theta} - \vec{\theta}_0$, and dropping the explicit dependence of the various quantities on

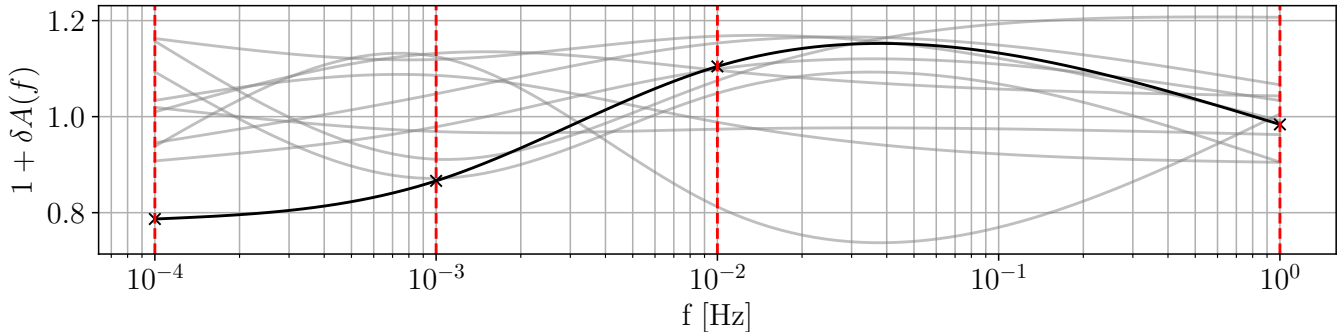


FIG. 1: Example calibration envelopes. Each line is a random draw of the amplitude calibration uncertainty, $1 + \delta A(f)$, assuming equal size independent uncertainties at each knot of magnitude $\sigma_C = 10^{-1}$. The vertical dashed lines show the locations of the knots.

time or frequency for ease of notation, we obtain

$$\begin{aligned}
 (d|\hat{h}(\vec{\theta})) &= (d|\hat{h}(\vec{\theta}_0)) + (d|\partial_i \hat{h}(\vec{\theta}_0)) \Delta \theta^i \\
 &\quad + \frac{1}{2} (d|\partial_{ij} \hat{h}(\vec{\theta}_0)) \Delta \theta^i \Delta \theta^j + \dots \\
 &= \frac{1}{2} (\Delta \theta^i + D_k (H^{-1})^{ki}) H_{ij} (\Delta \theta^j + (H^{-1})^{jk} D_k) \\
 &\quad + (d|\hat{h}(\vec{\theta}_0)) - \frac{1}{2} D_i H_{ij}^{-1} D_j + \dots \\
 \max_{\vec{\theta}} (d|\hat{h}(\vec{\theta})) &= (\rho_0 C_0 \hat{h}(\vec{\theta}_0) + n|\hat{h}(\vec{\theta}_0)) \\
 &\quad - \frac{1}{2} D_i H_{ij}^{-1} D_j + \dots
 \end{aligned}$$

$$\text{where } D_i = (d|\partial_i \hat{h}(\vec{\theta}_0)), \quad H_{ij} = (d|\partial_{ij} \hat{h}(\vec{\theta}_0)). \quad (10)$$

Since $(\hat{h}|\hat{h}) \equiv 1$ by construction, we deduce $(h|\partial_i \hat{h}) = 0$ and hence that the second term is quadratic in small quantities, while the first term has a linear correction in the size of the calibration uncertainty. We deduce that the expected value (i.e., the average value over realisations of the noise) of the maximum SNR in the presence of calibration uncertainties is $(\rho_0 C_0 \hat{h}(\vec{\theta}_0) | \hat{h}(\vec{\theta}_0)) = (C_0 h(\vec{\theta}_0) | h(\vec{\theta}_0)) / \rho_0$, which is

$$\frac{1}{\sqrt{h(\vec{\theta}_0) | h(\vec{\theta}_0)}} \int_0^\infty \frac{(1 + \delta A(f)) e^{2\pi i \delta \phi(f)} |\tilde{h}(f | \vec{\theta}_0)|^2}{S_n(f)} df. \quad (11)$$

The corresponding result in the absence of calibration uncertainties can be found by setting $\delta A(f) = \delta \phi(f) = 0$ in the above. The fraction of events that pass the specified SNR threshold could be evaluated using the above model combined with a population model and a model for the distribution of calibration uncertainties. However, we will not do this here for three reasons. Firstly, the modified SNR is linear in $(1 + \delta A(f))$ and so will tend to average to 1 if we use a symmetric amplitude calibration uncertainty model. This does not mean the number of detected events will be unchanged, as there are more events further away, but it will partially mitigate

the impact. Secondly, it is not expected that the calibration of LISA will be terrible, but uncertainties should be much smaller than unity. Typical uncertainties in the astrophysical rates of LISA events are one or more orders of magnitude, which will completely dominate over any impact of calibration errors. Finally, the measured SNR also depends on the assumed PSD, $S_h(f)$, and so will also be significantly affected by noise modelling uncertainties, which are likely to be much larger. It therefore does not make much sense to use fluctuations in the measured SNR to place a requirement on calibration if the noise model is fixed. A more robust way to place requirements on calibration is to limit their impact on parameter estimation for detected sources. We will describe how this can be done in the next section.

C. Impact of calibration error on parameter estimation

The impact of calibration errors on parameter estimation can be assessed in two different ways. One approach is to assume that the analysis of the data is based on taking $C(f) \equiv 1$, while the true data generating process is described by Eq. (3). When the model used in data analysis differs from reality you expect to obtain biases. Calibration error requirements can be set by estimating the induced biases and requiring them to be smaller than uncertainties arising from instrumental noise fluctuations. A second approach is to include additional calibration parameters as part of the model used to fit the data. In that case, provided the model is sufficiently flexible, we would not expect to see biases in parameter estimates, but we would expect that parameter posteriors would be broadened to reflect the additional degrees of freedom coming from the calibration model. We will take the second approach in this paper, but we describe the first approach in Appendix A, and show how it coincides with the second approach when calibration errors are small.

To implement the second approach, we modify the log-likelihood used in parameter estimation so that it in-

cludes the calibration error

$$\log \mathcal{L} = -\frac{1}{2} \int_0^\infty \frac{|\tilde{d}(f) - C(f|\vec{\mu})\tilde{h}(f|\vec{\theta})|^2}{S_n(f)} df, \quad (12)$$

where $\vec{\theta}$ are the parameters of the gravitational wave source, as before, and $\vec{\mu}$ are the parameters describing the calibration error model, in this case the parameters of the amplitude and phase calibration splines. We assume that the noise spectral density is known and this allows us to drop constant terms. If we wanted to model the noise as well we would need to include those terms. When analysing observed LISA data using Bayesian inference methods, the calibration error model parameters can be sampled at the same time as the parameters describing the source. The resulting posteriors on the source parameters will be marginalised over calibration errors. This procedure is computationally expensive and therefore impractical for setting calibration requirements over a wide range of source types and source parameters. However, we will use this in Appendix B to verify the results we obtain using computationally cheaper methods.

A cheaper way to assess the impact of calibration errors is to use the Fisher matrix formalism. In the absence of calibration uncertainties, the Fisher matrix is defined by

$$\Gamma_{ij}^\theta = \left(\frac{\partial h}{\partial \theta_i} \middle| \frac{\partial h}{\partial \theta_j} \right). \quad (13)$$

The Fisher matrix, Γ_{ij}^θ , provides an estimate of the precision to which the parameters of the source model can be determined from the data. Specifically, the uncertainty in parameter θ^i , $\Delta\theta^i$, can be estimated as $\langle \Delta\theta^i \Delta\theta^j \rangle = (\Gamma^\theta)_{ij}^{-1}$. In the presence of calibration errors, we can now evaluate the Fisher matrix for the modified likelihood given by Eq. (12). This Fisher matrix is over the expanded parameter set that includes both $\vec{\theta}$ and $\vec{\mu}$. The modified likelihood is related to the standard likelihood through the replacement $\tilde{h}(f) \rightarrow C(f|\vec{\mu})\tilde{h}(f|\vec{\theta})$. We can thus directly write down the Fisher Matrix for the combined estimation of $\vec{\theta}$ and $\vec{\mu}$

$$\begin{aligned} \Gamma &= \begin{pmatrix} \Gamma^{\theta\theta} & \Gamma^{\theta\mu} \\ (\Gamma^{\theta\mu})^T & \Gamma^{\mu\mu} \end{pmatrix} \\ \text{where } \Gamma_{ij}^{\theta\theta} &= \left(C(\vec{\mu}) \frac{\partial h}{\partial \theta_i} \middle| C(\vec{\mu}) \frac{\partial h}{\partial \theta_j} \right), \\ \Gamma_{ij}^{\theta\mu} &= \left(C(\vec{\mu}) \frac{\partial h}{\partial \theta_i} \middle| h(\vec{\theta}) \frac{\partial C}{\partial \mu_j} \right), \\ \Gamma_{ij}^{\mu\mu} &= \left(h(\vec{\theta}) \frac{\partial C}{\partial \mu_i} \middle| h(\vec{\theta}) \frac{\partial C}{\partial \mu_j} \right). \end{aligned} \quad (14)$$

We anticipate that calibration will be good and therefore that $C(f) \approx 1$. So, we can evaluate this expression for $C(f) = 1$ when assessing and setting calibration requirements.

Knowledge of the accuracy of calibration can be incorporated by imposing a prior on the calibration parameters, $\vec{\mu}$. When doing numerical marginalisation any prior

can be imposed, but in the Fisher matrix formalism it is easiest to work with a Gaussian prior. Using the prior $\vec{\mu} \sim \mathcal{N}(0, \Sigma^{\mu\mu})$ (assuming that we centre the parameters such that $\vec{\mu} = 0$ corresponds to $C(f) = 1$), the posterior covariance is given by the inverse of the modified Fisher matrix

$$\Gamma = \begin{pmatrix} \Gamma^{\theta\theta} & \Gamma^{\theta\mu} \\ (\Gamma^{\theta\mu})^T & \Gamma^{\mu\mu} + \Sigma^{\mu\mu} \end{pmatrix}. \quad (15)$$

The diagonal elements of the inverse of this matrix provide estimates for the precision with which the corresponding parameters can be measured. The estimated precision of measurement of the waveform parameters accounting for calibration model uncertainty is thus given by the diagonal elements of the matrix

$$(\Gamma^{\theta\theta} - \Gamma^{\theta\mu}(\Gamma^{\mu\mu} + \Sigma^{\mu\mu})^{-1}(\Gamma^{\theta\mu})^T)^{-1}. \quad (16)$$

Given a specification for the calibration model uncertainties, $\Sigma^{\mu\mu}$, these parameter estimation uncertainties can be evaluated and compared to the diagonal elements of $(\Gamma^{\theta\theta})^{-1}$, which are the estimated uncertainties with perfect calibration. If the uncertainties are significantly larger then our lack of knowledge of the calibration is having an impact on our ability to measure the parameters of the system. For the purpose of concreteness we will say that the calibration knowledge, $\Sigma^{\mu\mu}$, is not good enough if the error predicted by Eq. (16) is more than double that predicted by $(\Gamma^{\theta\theta})^{-1}$ for any parameter. Although chosen arbitrarily, this threshold corresponds to the point at which the systematic biases that arise from ignoring calibration uncertainties become comparable to the size of statistical errors, as described in Appendix A. This is the standard criterion for concluding that a systematic bias is unacceptably large in the context of assessing the impact of waveform modelling errors [3].

In section III we will use this formalism to evaluate the impact of calibration uncertainty on parameter estimation for all different types of LISA source. In all cases we will take $\Sigma^{\mu\mu}$ to be a diagonal matrix, so that we assume the calibration uncertainty at each frequency knot is independent of the others, and amplitude and phase uncertainties are also independent. Additionally, we will usually assume that the size of the uncertainties are the same at all knots, so that there is a single parameter $\sigma_C = \sqrt{\Sigma^{\mu\mu}}$ that characterises the size of the amplitude uncertainty and likewise for the phase. This makes the presentation of the results easier, as we can then plot the ratio of the parameter measurement uncertainties with and without calibration uncertainty as a function of the size of the calibration uncertainties σ_C , making it easier to identify the point at which the error-doubling threshold is reached. We will present results that vary only the amplitude uncertainties or only the phase uncertainties. For all source types, we have also computed results varying both amplitude and phase calibration uncertainties simultaneously. These are not presented explicitly, as in all cases the results were the same. This is because

there are certain parameters that are sensitive to amplitude calibration uncertainties and other parameters that are sensitive to phase calibration uncertainties, but none appear to be sensitive to both.

As a final remark, we note that we can evaluate Eq. (16) with $\Sigma^{\mu\mu} \rightarrow 0$, and obtain finite uncertainties for many parameters. This corresponds to the limit in which we assume that the calibration is completely unknown and that our data analysis procedure attempts to estimate the calibration simultaneously with the source parameters. The fact that source parameters can still be estimated is because the assumed calibration uncertainty does not look very much like a gravitational waveform and hence can be distinguished from a gravitational wave source. This will be discussed further in section IV.

III. CALIBRATION REQUIREMENTS FOR INDIVIDUAL SOURCES

In this section we will use the formalism described in the preceding section to assess the calibration requirement that arises from requiring that we are able to do accurate parameter estimation on individual resolvable sources. We will consider each of the main categories of resolvable source expected to be observed by LISA.

A. Galactic binaries

1. Definition

Among the gravitational wave sources expected to be observed by the LISA mission, the most common will come from galactic binaries (GB). Binaries of compact stellar remnants, typically white dwarfs but also neutron stars and black holes, with orbital periods of around 1 hour in the Milky Way emit gravitational waves at millihertz frequencies that LISA will be able to observe. Several tens of millions of these binaries are expected to be present in the LISA data. The majority of these will form an unresolvable astrophysical foreground between 0.5 mHz and 3 mHz, but we still expect to individually resolve as many as several tens of thousands of sources. Various approaches have been proposed to identify and characterise these sources [8–11]. As for all sources, we will assess the impact of calibration errors on parameter estimation for individual galactic binaries. In principle, calibration errors could also lead to the incomplete subtraction of binaries from the data and hence lead to an increase of the galactic confusion noise, with a knock-on impact on the characterisation of other LISA sources. We will not consider this here, as we expect the effect to be small in the regime where individual source parameter estimation is unaffected by calibration error.

To represent gravitational waves from galactic binaries we will use the model provided by the LISA Data Challenge <https://lisa-ldc.lal.in2p3.fr/>, including all

three TDI data channels, A, E, T . The GW signal from a GB is characterized by 8 parameters. The ecliptic latitude λ , the ecliptic longitude β , the inclination ι , the polarization ψ and the amplitude \mathcal{A} are extrinsic parameters that describe the position and orientation of the source with respect to LISA. The remaining three parameters are intrinsic and determine the temporal evolution of the plus h_+ and cross h_x polarizations of the GW as follows:

$$\begin{aligned} h_+^S &= \mathcal{A} (1 + \cos^2 \iota) \cos(\Phi(t)), \\ h_x^S &= 2\mathcal{A} \cos \iota \sin(\Phi(t)), \\ \Phi(t) &= 2\pi f_0 t + \pi \dot{f} t^2 - \phi_0 \end{aligned}$$

where f_0 is the GW frequency given at an initial moment t_0 , \dot{f}_0 is the frequency first order derivative and ϕ_0 is the initial GW phase. The amplitude depends on the binary chirp mass \mathcal{M}_c , the frequency f_0 and the distance D :

$$\mathcal{A} = \frac{(G\mathcal{M}_c)^{\frac{5}{3}} (\pi f_0)^{\frac{2}{3}}}{c^4 D}. \quad (17)$$

The chirp mass is related to the masses of the individual components of the binary, m_1 and m_2 , via $\mathcal{M}_c = m_1^{\frac{3}{5}} m_2^{\frac{3}{5}} / (m_1 + m_2)^{\frac{1}{5}}$. For a binary evolving purely under GW emission, the frequency derivative and chirp mass are related through

$$\dot{f}_0 = -\frac{96 \pi^{\frac{8}{3}}}{5 c^5} (G\mathcal{M}_c)^{\frac{5}{3}} f_0^{\frac{11}{3}} \quad (18)$$

but there are other processes, such as mass-transfer, that can drive frequency evolution in galactic binaries, so we do not use this relation to interpret \mathcal{A} in terms of D .

For typical galactic binaries, $\dot{f}_0 \lesssim 10^{-15}$, so the GW signal from a GB is almost monochromatic with a slow drift in the frequency over years of observation, although it is modulated (time-varying Doppler shift) due to LISA's orbital motion. Some galactic binaries, referred to as verification galactic binaries (VGB), have distances that have been well measured using electromagnetic observations. These systems could be used to directly measure the amplitude calibration error and its evolution with time. This will be discussed further in section IV A.

2. Calibration uncertainty Fisher matrix for galactic binaries

In order to understand which of the galactic binary parameters is most likely to be affected by calibration uncertainty, we first look at the structure of the Fisher matrix for the simultaneous estimation of the calibration uncertainty and GB parameters, computed using the formalism described in section II C. We consider a single choice for the GB parameters, but the structure of the Fisher matrix does not depend on this choice.

We consider a GB with the same parameters as the first VGB (called AM-CVn) present in the LDC catalog (<https://gitlab.in2p3.fr/korol/lisa-verification-binaries> and references therein). The parameters are $f_0 = 1.9 \times 10^{-3}$, $\dot{f}_0 = 6.1 \times 10^{-18}$, $\mathcal{A} = 2.8 \times 10^{-22}$, $\iota = 0.8$, $\lambda = 0.7$, $\beta = 3.0$, $\psi = 2.2$ and $\phi_0 = 6.1$. The waveform is generated for a 4 year duration with LISA analytical orbits given in the LDC manual (<https://lisa-ldc.lal.in2p3.fr/static/data/pdf/LDC-manual-001.pdf>), noise according to the first version of the Scientific Requirement Document (SciRDv1) <https://www.cosmos.esa.int/documents/678316/1700384/SciRD.pdf> and with instrumental response computed using second order time delay interferometry (TDI2.0).

As discussed in section II A, the full calibration error model we use here consists of two cubic spline functions, defined by four amplitude error parameters $\{\delta A_0, \delta A_1, \delta A_2, \delta A_3\}$ and four phase error parameters $\{\delta \varphi_0, \delta \varphi_1, \delta \varphi_2, \delta \varphi_3\}$. The cubic spline is constructed as a function of the logarithm of frequency to ensure smoothness over the full LISA frequency range. The elements of the inverse Fisher matrix, with and without the inclusion of calibration errors, are given in figure 2. The results with calibration errors take the calibration error uncertainty to be $\sigma_C = 0.1$. The diagonal elements of the inverse Fisher matrix provide an estimate of the precision with which the corresponding parameter can be determined, while the off-diagonal elements indicate the amount of correlation in the uncertainties between pairs of parameters.

Comparing the diagonal elements in the left and right panels we see that the only two parameters for which the uncertainties change significantly when calibration uncertainties are included are \mathcal{A} and ϕ_0 . This conclusion is supported by the off-diagonal elements which show that these two parameters are the only ones that are highly correlated with the calibration model parameters, specifically δA_1 and $\delta \varphi_1$. This makes sense physically, since a change in \mathcal{A} represents a constant change in the amplitude of the signal, and δA_1 does the same thing, at the frequency of this source. Similarly a change in ϕ_0 is a constant change in phase, which can also be accomplished by a change in $\delta \varphi_1$.

3. Setting a calibration requirement for galactic binary parameter estimation

In the previous section we saw the precision of parameter estimation changed for a particular choice of calibration uncertainty. To set a calibration requirement, we must vary the size of the calibration uncertainty to see how large this can be before parameter estimates become significantly degraded. These results are shown in Figure 3 when varying only the amplitude calibration uncertainty or only the phase calibration uncertainty. Results varying both simultaneously are the same and are there-

fore not shown. Consistent with the form of the inverse Fisher matrix in Figure 2, we find that the only parameter impacted by phase calibration uncertainty is the initial phase, ϕ_0 , and the only parameter affected by amplitude calibration uncertainty is the amplitude/distance. In each case, the precision of parameter estimation begins to degrade when the size of the calibration uncertainty is comparable to the precision with which the affected parameter can be measured in the absence of these uncertainties. Requiring that the parameter uncertainty no more than doubles sets a requirement that the amplitude calibration uncertainty is no bigger than a few percent and the phase calibration uncertainty is no bigger than 3×10^{-2} .

Once the uncertainty starts to increase it varies roughly linearly with the size of the calibration uncertainty. This behaviour can be understood analytically, which is discussed further in appendix B. While these results were based on the Fisher matrix, they can also be verified by doing full Bayesian parameter estimation with MCMC methods. We have done this for the case of galactic binaries and the MCMC results verify the conclusions of the Fisher matrix. See Figure 14 in appendix B.

The stated calibration requirements are valid for this particular choice of GB parameters, but we would find a different requirement if we considered different parameter choices. This dependence on the source parameters can be understood from the analytic results described in appendix B, but in general we expect the impact of calibration uncertainties to be greatest on sources for which we can determine the parameters most precisely. These are typically the highest signal-to-noise ratio systems. We expect the system considered here to lead to one of the most stringent constraints, so we would not expect to need more than a factor of a few more stringent calibration requirement in the worst case.

B. Massive black hole binaries

1. Definition

The mergers of massive black hole binaries (MBHBs) are among the loudest gravitational wave sources expected to be observed by the LISA mission. Most galaxies contain massive black holes at their centres and MBHBs are expected to form following the merger of their host galaxies. Such systems are believed to form even very early in the history of the Universe, and LISA can observe systems of suitable mass ($\sim 10^4 M_\odot - 10^7 M_\odot$) to very high redshift ($z \sim 20$). LISA is expected to observe several tens of sources and these observations will enable us to better understand the formation processes of these binary systems, their environments and to discriminate between different scenarios of seed MBH formation.

In this analysis, we model the source in the Fourier domain using the IMRPhenomD model [12], which represents coalescing black holes in which the spins are par-

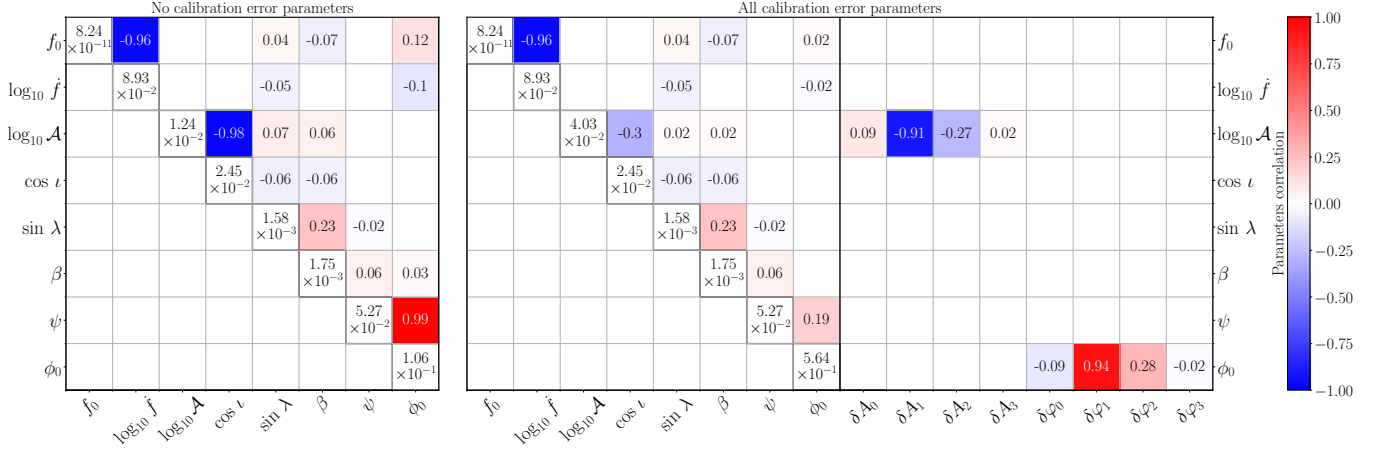


FIG. 2: Inverse of the Fisher matrix for a single GB with (right) and without (left) a calibration error function. The diagonal elements provide uncertainties in the corresponding parameters, while the off-diagonal elements indicate the degree of correlation in the estimates of pairs of parameters. As we are primarily interested in the impact of calibration on the determination of the source parameters, for clarity we do not show the calibration-calibration sub-matrix. Off-diagonal elements that are smaller than 0.01 are left blank.

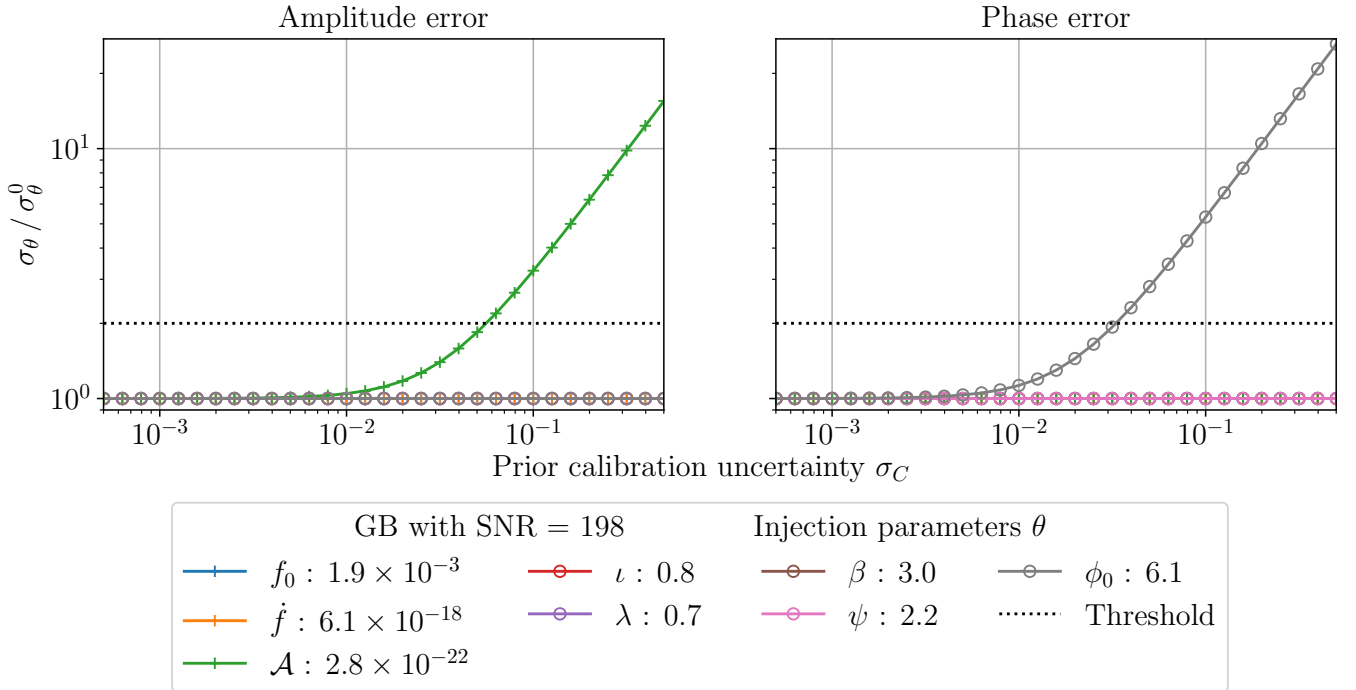


FIG. 3: Ratio between the uncertainty of a waveform parameter with (σ_θ) and without (σ_{θ_0}) the calibration error function as a function of the size of the calibration error uncertainty σ_C . The red line corresponds to a doubling of the parameter uncertainty which is our threshold for the calibration uncertainty being unacceptably large. In the left hand panel we vary only the amplitude calibration uncertainty, and in the right panel we vary only the phase calibration uncertainty.

allel to the orbital angular momentum. We construct the LISA TDI response in the same way as for the galactic binaries. The GW signal from a MBHB is characterized by 11 parameters. There are the same five extrinsic parameters that were described for the GB model: the ecliptic latitude β , the ecliptic longitude λ , the inclination ι , the polarization ψ and the distance D . The additional intrinsic parameters are the chirp mass, \mathcal{M}_c , the symmetric mass ratio, $\eta = m_1 m_2 / (m_1 + m_2)^2$, the spins of the two components, $\chi_{1,2}$, the phase at coalescence, ϕ , and the coalescence time, τ . The bandwidth of an MBHB signal in the frequency domain is broad, in contrast to the narrow bandwidth of a GB, which we expect to help us distinguish MBHB signals from calibration errors.

2. Calibration uncertainty Fisher matrix for massive black hole binaries

As for the GB case, we first look at the correlations in the Fisher matrix for simultaneous estimation of calibration errors and MBHB parameters. The inverse Fisher matrix is shown in Figure 4 for a typical MBHB source with parameters $\mathcal{M}_c = 7.82 \times 10^5 M_\odot$, $\eta = 0.2$, $\chi_1 = \chi_2 = 1.0$, $D = 1.42 \times 10^5 \text{Gpc}$, $\iota = 2.3$, $\lambda = -0.6$, $\beta = 0.6$, $\tau = 4.8 \times 10^6 \text{s}$, $\psi = 3.1$ and $\phi_0 = 4.3$. We again take the prior on the calibration uncertainty to have a width $\sigma_C = 0.1$.

Comparing the diagonal elements in the left and right hand figures we see that the inclusion of calibration uncertainties has a moderate impact on the measurements of several of the intrinsic parameters – \mathcal{M}_c , η , χ_1 , χ_2 – and a more significant impact on the measurement of ϕ_0 and D . This is to be expected from looking at the off-diagonal elements, which show that the measurement of these four intrinsic parameters is mildly (~ 0.3) correlated with the measurement of the calibration uncertainty parameters, while the measurements of ϕ_0 and D are strongly (~ 0.9) correlated with the phase and amplitude calibration uncertainty parameters respectively. These results make sense – changes in the amplitude and ϕ_0 impact the MBHB signal in the same way that they impact the GB signal and so we expect the same degeneracy with the calibration parameters that we saw in the GB case. The other intrinsic parameters primarily affect the phase evolution of the MBHB signal and so we might expect a mild degeneracy with the phase calibration uncertainty.

3. Setting a calibration requirement for MBHB parameter estimation

We now explore the variation in the parameter measurement uncertainty for MBHBs as we vary the size of the calibration uncertainty. These results are shown in Figure 5. As expected we see that once the amplitude calibration uncertainty is comparable to the precision with

which the distance can be measured, it starts to limit our ability to measure distance. Once the phase calibration uncertainty is sufficiently large, it has a similar impact on measurements of the phase offset, ϕ_0 . In this case we also see a mild impact on the other intrinsic parameters – \mathcal{M}_c , η , χ_1 and χ_2 . Once again, our ability to measure these parameters starts to degrade once the calibration uncertainties become comparable with the expected precision on these parameters in the absence of calibration uncertainty. For these parameters, however, the uncertainty does not continue to increase, but levels off at about twice the uncertainty in the absence of calibration errors. This happens because the degeneracy is not perfect in this case – the phase evolution of an MBHB waveform does not look like a calibration uncertainty of the form modelled here and so the two effects can be distinguished even with no prior on the size of the calibration error, with only mild confusion.

For this source we deduce a calibration requirement of $\sim 8\%$ in amplitude and 2×10^{-2} in phase to ensure the measurement uncertainties are no more than doubled. Once again, this conclusion will be source dependent but this source is relatively high signal-to-noise ratio and so we would not expect the results to change by more than a factor of a few for other sources.

C. Stellar-origin black hole binaries

1. Definition

Stellar-origin black hole binaries (SBHBs) refer to binaries of compact objects formed as the end point of binary stellar evolution. These are the systems from which ground-based detectors are observing gravitational waves emitted during the final stages of inspiral and merger. For binaries with sufficiently large component masses ($\sim 30 M_\odot$), LISA will observe gravitational waves emitted by these systems as the gradually inspiral, ~ 10 years before merger. The number of these events that will be observed is somewhat uncertain, as it depends critically on the high-mass end of the SBHB mass spectrum, which is not yet well constrained. However, they are of great scientific interest because of the prospect of multi-band observations of the same system with both LISA and ground-based instruments.

In this work we use the same model for gravitational waves from SBHBs as we do for MBHBs, the aligned-spin IMRPhenomD [12] model, as implemented in the LISA Data Challenge. The parameters characterising these systems are the same as those for the MBHBs, the only difference is that the chirp mass, \mathcal{M}_c , and distance, D , take different ranges. As for MBHBs, these systems are relatively broad in the frequency domain which should help to distinguish these systems from calibration errors. Unlike MBHBs, LISA is expected to observe SBHBs with relatively low signal-to-noise ratio and measure the parameters with relatively poorer precision. This should

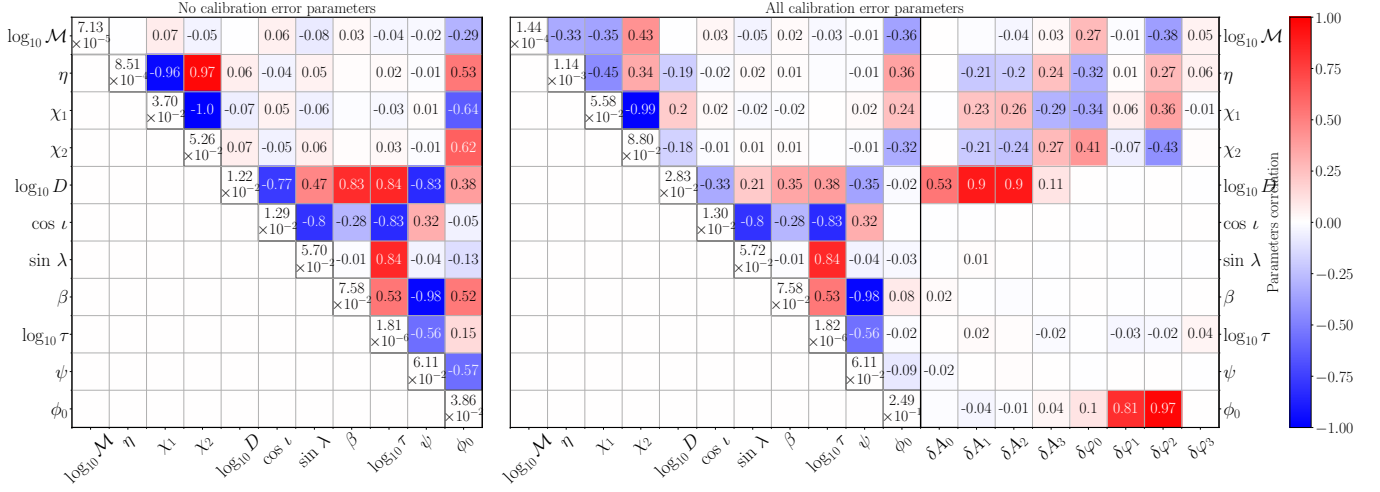


FIG. 4: As Figure 2 but now for a MBHB source.

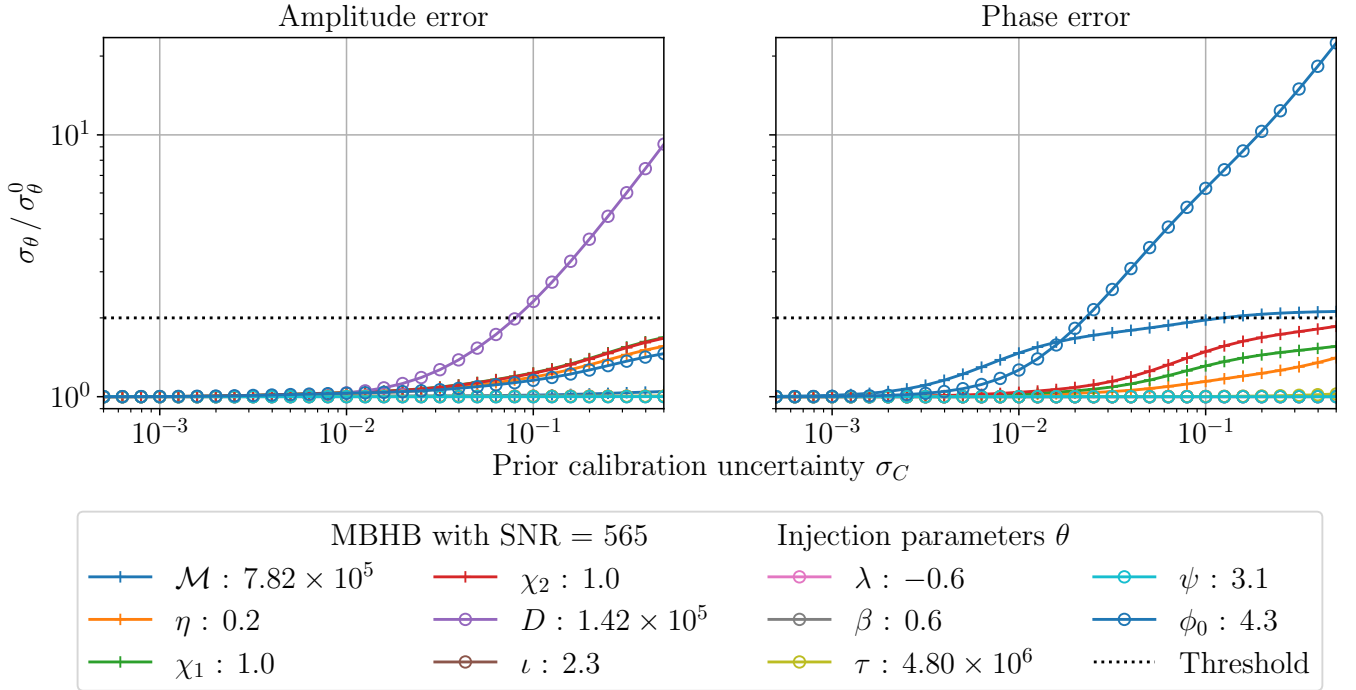


FIG. 5: As Figure 3, but now for the MBHB source shown in Figure 4.

also limit the impact of calibration errors.

2. Calibration uncertainty Fisher matrix for stellar-origin black hole binaries

In Figure 6 we again show the elements of the inverse Fisher matrix with and without the inclusion of the calibration uncertainty parameters. These are shown for a SBHB source with parameters $\mathcal{M}_c = 47.9M_\odot$, $\eta = 0.2$, $\chi_1 = \chi_2 = 0.0$, $D = 742.82\text{Mpc}$, $\iota = 0.4$, $\lambda = 2.0$, $\beta = 1.2$, $\tau = 2.1 \times 10^8\text{s}$, $\psi = -3.3$ and $\phi_0 = 2.7$. Once again we take $\sigma_C = 0.1$. Comparing the diagonal elements of the Fisher matrix in the left and right hand panels we find in this case that there is essentially no change in the precision of parameter estimation when simultaneously fitting for the calibration model. This is supported by the off-diagonal elements of the Fisher matrix which show that there is almost no correlation between the parameters of the signal model and those of the calibration error model.

3. Setting a calibration requirement for SBHB parameter estimation

As before we now vary the size of the calibration uncertainty, σ_C . The results are shown in Figure 7. We see that even for relatively large calibration uncertainties, the parameter measurement precisions remain almost unchanged relative to their values in the absence of calibration uncertainty. The inset panels show that for calibration uncertainties of order unity we begin to see an impact on the measurement of distance and initial phase, but the error increases by less than 1%. This result is driven by the fact that SBHBs are observed with relatively low signal-to-noise ratio and with relatively poor parameter estimation precision. Artificially increasing the signal-to-noise we observe more of an effect (see appendix B). Based on these results, the characterisation of SBHBs does not appear to impose a requirement on the calibration of the LISA instrument.

D. Extreme-mass-ratio inspirals

1. Definition

Extreme-mass-ratio inspirals (EMRIs) are the inspirals of stellar-origin compact objects (COs) into massive black holes (MBHs) in the centres of galaxies. EMRI observations with LISA have great scientific potential, since the compact object typically generates $10^4 - 10^5$ detectable waveform cycles in band, during which time it is orbiting in the strong field region close to the central black hole. The emitted gravitational waves encode a detailed map of the space-time outside the central MBH and offer a unique opportunity to measure the properties,

evolution and environment of MBHs [13–18], to test for deviations from General Relativity (GR) [19, 20] and to constrain cosmological parameters [21, 22]. Event rates are highly uncertain, but LISA could observe between a few and a few hundred EMRIs over the mission duration [18, 23–27].

The great scientific potential of these astrophysical sources relies on being able to make very precise measurements of their parameters and this could therefore be significantly impacted by calibration uncertainties. To assess this we use two different waveform models: the Analytical Kludge [13] and the Augmented Analytical Kludge as implemented in the Fast EMRI waveform package [28–33]. We use the low-frequency approximation to implement LISA response, as first described in [13, 34]. The full EMRI parameter space is seventeen dimensional [29], but for the purposes of this study we restrict attention to seven parameters: the central black hole mass, $M [M_\odot]$, the dimensionless spin parameter of the central MBH, a , the compact object mass, $\mu [M_\odot]$, the initial semi-latus rectum, p_0 , eccentricity, e_0 , and cosine of the inclination, Y_0 , of the orbit, and the luminosity distance, $d_L [\text{Gpc}]$.

2. Calibration uncertainty Fisher matrix for EMRIs

In Figure 8 we show the inverse of the Fisher matrix with and without including calibration uncertainties for an EMRI source with parameters $M = 10^6 M_\odot$, $a = 0.9$, $\mu = 23.1 M_\odot$, $p_0 = 12.0$, $e_0 = 0.6$, $Y_0 = 0.7$ and $d_L = 1.1\text{Gpc}$, observed for 2 years. The EMRI model used for these results was the AAK, but we found the same quantitative results when using the AK model. We again fix $\sigma_C = 0.1$. Comparing the diagonal elements of the Fisher matrix on the left and the right we see that distance measurement precision is significantly impacted by calibration uncertainties and there is also a minor impact on measurements of the intrinsic parameters. As in previous cases, this behaviour is reinforced by looking at the off-diagonal elements of the Fisher matrix, which shows weak correlations between measurements of the intrinsic parameters of the source and the phase calibration parameters, and a very strong correlation between measurements of the distance and the amplitude calibration parameters.

3. Setting a calibration requirement for EMRI parameter estimation

We now vary σ_C , which determines the size of the calibration uncertainties, and determine how the parameter measurement precisions change. These results are shown in Figure 9 and were again computed using the AAK EMRI model. Results from the AK model show the same quantitative behaviour. These results are consistent with the conclusions from looking at the elements of the inverse Fisher matrix. Once the amplitude calibration un-

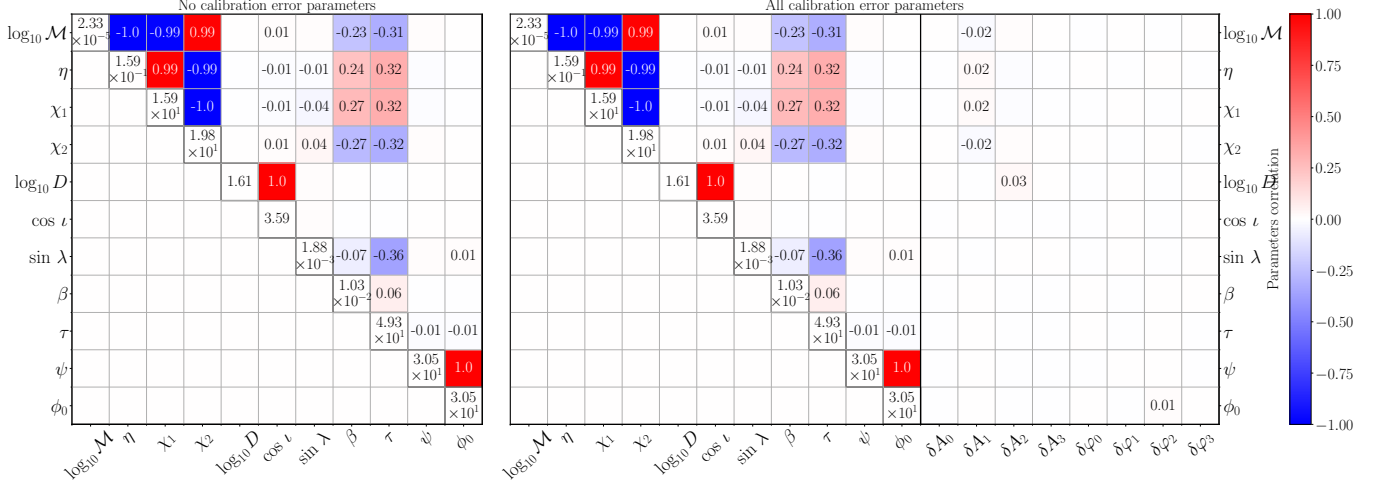


FIG. 6: As Figure 2 but now for a SBHB source.

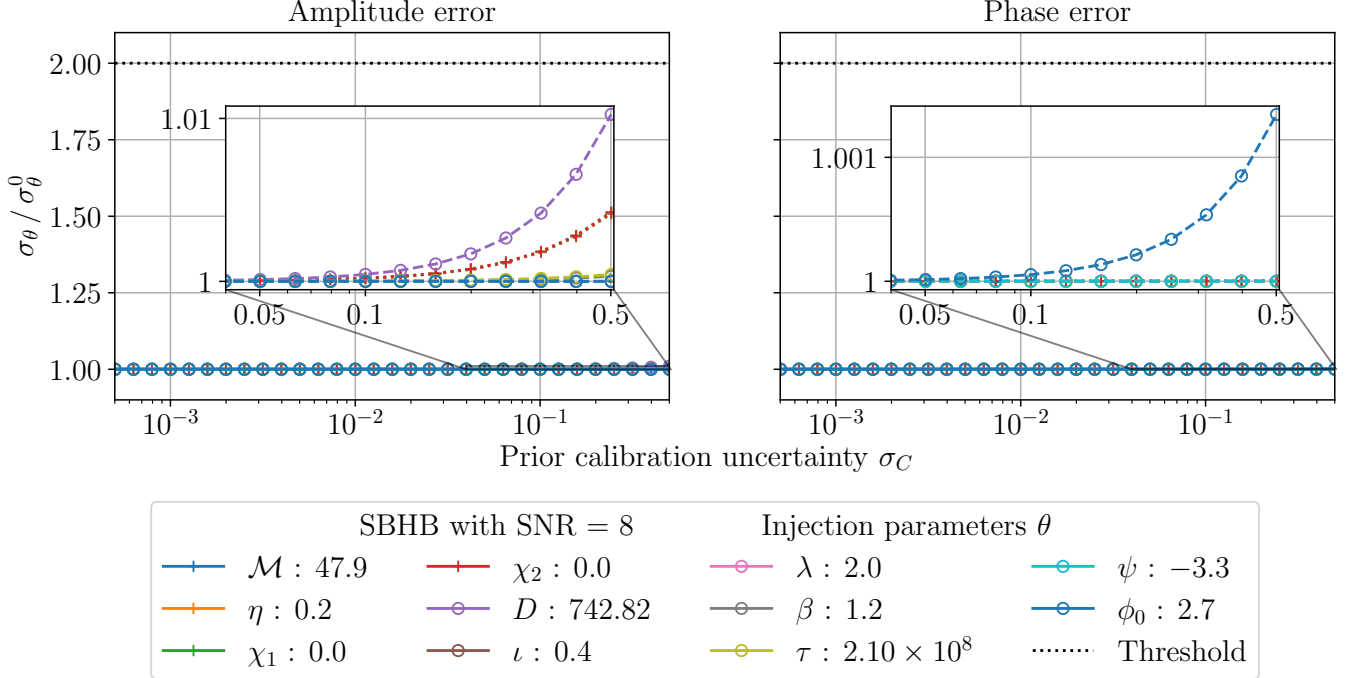


FIG. 7: As Figure 3 but now for the SBHB source shown in Figure 6.

certainty becomes too large, we see that the distance measurement precision rapidly degrades, ultimately linearly with the size of the calibration uncertainty. The measurement precision of the intrinsic parameters also starts to degrade once the phase calibration uncertainty becomes large enough, but the degradation is limited and the measurement precision is never more than $\sim 25\%$ worse than in the absence of calibration uncertainty.

These results are consistent with results for other sources and are readily understood. The effect of amplitude calibration error in the observed data is degenerate with the effect of a change in the source distance, so once the uncertainty in the former becomes comparable with the expected measurement precision in the latter it starts to dominate the uncertainty. The intrinsic parameters are measured primarily through their influence on the phase evolution of the source and so we expect them to be somewhat degenerate with phase calibration uncertainty. However, the complex evolution of the phase with frequency in an EMRI is very unlike the slowly varying phase calibration error we assume here and so the degeneracy is limited, allowing both effects to be simultaneously constrained from the observed data.

One of the primary scientific applications of EMRI observations will be to provide stringent tests that the spacetime structure outside the MBH is consistent with the Kerr metric, as predicted by the no-hair theorem. Such tests will be carried out by fitting EMRI models that include additional parameters that represent deviations from this assumption. We can thus assess the impact of calibration uncertainties on these tests of general relativity using the same Fisher matrix approach, applied to the extended parameter space that includes standard EMRI parameters, calibration uncertainty parameters and parameters characterising deviations from general relativity. We do this using the extended AK model described in [20, 25], which includes one additional parameter, Q , that represents an “excess quadrupole”, i.e., a difference in the quadrupole moment of the spacetime relative to the value predicted from its mass and spin in the Kerr metric. In Figure 10 we show how the measurement precision of Q varies as we increase the size of the calibration uncertainty. This was evaluated for a different EMRI system to the one considered above, with smaller mass for the inspiraling body, $\mu = 10M_{\odot}$, smaller eccentricity ($e_{\text{pl}} = 0.06$ at plunge) and a shorter observation time (1 year). The Fisher matrix was evaluated with the true excess quadrupole moment set to $Q = 0$. The conclusions are not significantly affected by the specific choice of EMRI parameters. This quadrupole moment is again measured through its impact on the phase evolution of the binary and hence it is not surprising that the measurement precision behaves in a similar way to the other intrinsic parameters. We see a slight degradation in precision for large calibration uncertainties, but even with uncertainties of order unity, the measurement precision is within 5% of its value in the absence of calibration error.

Based on this source, we conclude that EMRIs do not place a requirement on phase calibration uncertainty, but place a requirement on amplitude calibration uncertainty of $\sim 8\%$. Once again we expect this to be source specific, and this EMRI has signal-to-noise ratio of 30, which is close to the threshold needed for EMRI detection. In the best case EMRIs could have signal-to-noise ratio of a few hundred, so we might expect to require amplitude calibration uncertainty of $\sim 1\%$ to ensure calibration has no impact for any individually resolved EMRI.

E. Calibration requirements for inference on source populations

All of the results described in this section have been based on measurements of the parameters of individual sources. LISA observations of many individual sources will also be combined to make statements about the properties of the underlying population and we also do not want population inference to be calibration limited. What happens when sources are combined will depend critically on whether the calibration errors for each source are independent of one another or not, which itself depends on how much the LISA calibration error might vary with time. With reference to Eq. (5), it is important to know if $|1 - C_i(f)| \ll |1 - C(f)|$ for all i . If all sources are affected by the same calibration error then combining the sources can help to measure the calibration error and hence reduce uncertainties for individual events. This would rely on the fact that different sources have different frequency dependencies, and does not apply to an overall scaling, which would impact distance measurements and hence constraints on cosmological parameters. If individual sources have independent calibration errors then the impact on population parameters will at least partially average out as sources are combined and so the uncertainties in the population might well be comparable to those on the individual events.

Without doing detailed simulations, we can place a conservative bound by considering a worst case scenario. In the absence of calibration errors we expect that the constraint on a population parameter improves like $1/\sqrt{N}$ as the number of observed sources, N , increases. In the worst case, the bias in the population parameter arising from calibration uncertainties will be common to all events and hence does not change as the number of events is increased. Therefore, we need the uncertainty arising from calibration error to be a factor of $1/\sqrt{N}$ smaller than the instrumental-noise induced uncertainty for a single event if we want to ensure that the calibration-error-induced uncertainty does not dominate after we combine N events. Typical LISA source populations contain $O(100)$ events [6]. Thus, if the parameter uncertainties for individual events are increased by no more than 10% when accounting for calibration uncertainties, we would expect population uncertainties to be no more than doubled. This corresponds to placing a

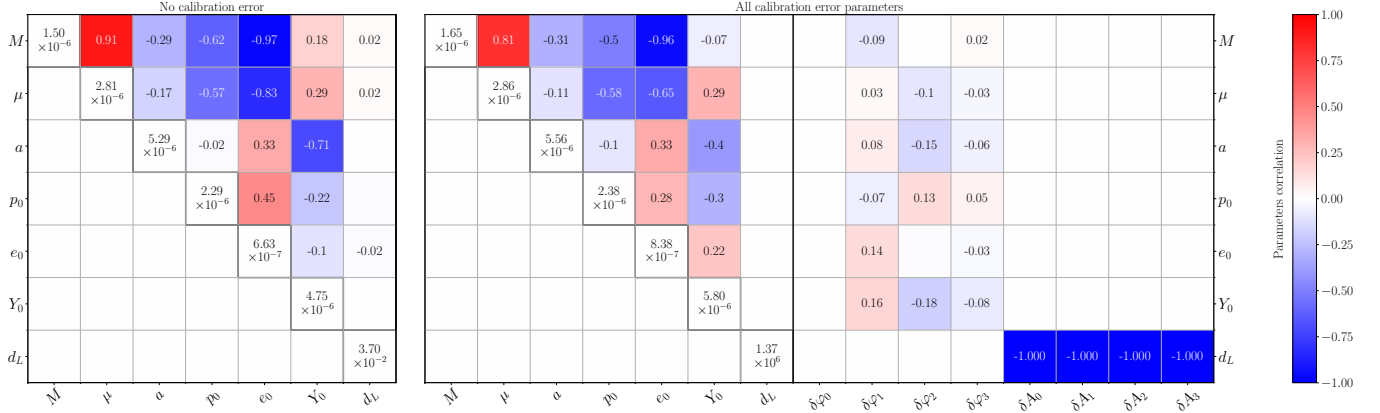


FIG. 8: As Figure 2 but now for an extreme-mass-ratio inspiral source.

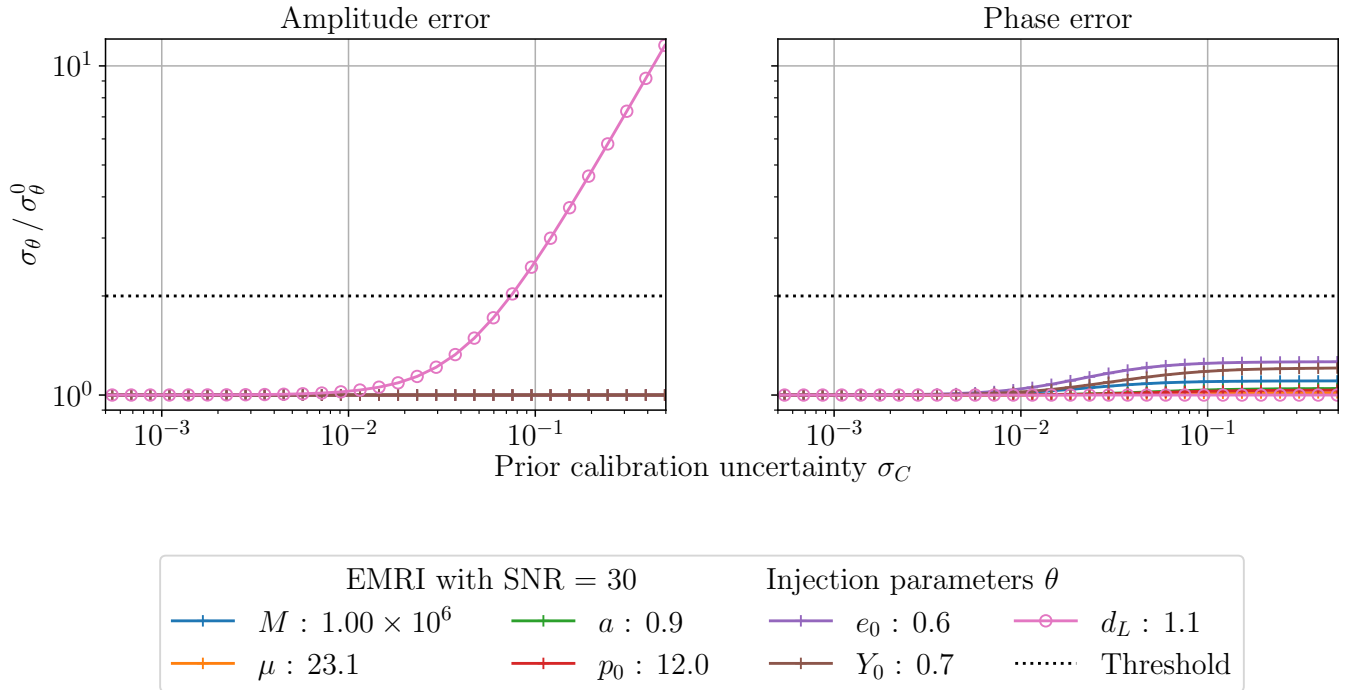


FIG. 9: As Figure 3 but now for the EMRI source shown in Figure 8.

threshold $\sigma_\theta/\sigma_{\theta_0} < 1.1$ on individual event parameter inference. Given the roughly linear growth of parameter uncertainties with the size of calibration uncertainties, we might expect to need a calibration requirement a factor of 10 more stringent for population inference. Looking more carefully at the earlier results we find this constraint imposes amplitude calibration requirements of 10^{-2} , 2×10^{-2} , > 1 and 10^{-2} for population inference with galactic binaries, MBHBs, SBHBs and EMRIs respectively. The corresponding phase calibration requirements are 8×10^{-3} , 2×10^{-3} , > 1 and 2×10^{-2} . Including an additional factor of a few to account for the fact that these sources may not be completely representative of the population, we conclude that if LISA amplitude calibration is better than a few $\times 0.1\%$ and phase calibration is better than 10^{-3} we would not expect to see any impact of calibration uncertainty on the LISA science output. If this more stringent requirement could not be met, then a more careful study would be required to understand if there would be any impact on population inference, which would have to account for possible time-variability of the LISA calibration error. This is beyond the scope of this paper.

IV. USING GRAVITATIONAL WAVE SOURCES AS CALIBRATORS

A. Known galactic binaries as amplitude calibrators

For all source types we have seen that there is a degeneracy between measurement of distance and uncertainty in amplitude calibration, which ultimately limits our ability to measure distances. However, if distance is already known, the same effect can be used to directly measure the amplitude calibration error. Most of the sources LISA will observe will not have known distances, but the verification galactic binaries (VGBs) are an exception. These are galactic binaries which are already known from electromagnetic observations [36]. There are a few tens of such systems for which the sky position is known near-perfectly, the period is known with a fractional uncertainty of a few $\times 10^{-6}$, the inclination is known to $\sim 10\%$ and the distance is known to between a percent and a few tens of percent. Combining these uncertainties, the GW amplitude in Eq. (17) is known to between 1% and several tens of percent. Figure 11 shows the GW amplitude uncertainties for all currently known VGBs. By comparing the observed amplitude of these sources to these expected amplitudes, we will be able to directly measure the amplitude calibration error as a function of frequency. The best VGBs for doing this (giving the tightest constraint over frequencies in the range [0.3, 5] mHz) are joined by a blue line in Figure 11. Unfortunately, the information on the initial phase of the VGBs is missing. Subsequently, we make the choice to consider the phase error as small in order to demonstrate

the significant improvement of the calibration error parameters estimation when multiple VGBs are considered.

To illustrate this procedure, we simulate the observation of a set of 8 VGBs and then recover posterior distributions on the source amplitude and the amplitude calibration spline parameters, keeping the other sources parameters fixed to their true values. We impose Gaussian priors on the weights at each knot of the calibration spline, with width of 0.1 in strain, and we consider the case either that there is no prior electromagnetic information on the source distance, or that we use EM prior information from the best 8 VGBs identified in Figure 11. The waveforms were generated for a 4 year LISA observation, using a LISA response based on analytical orbits and computed using second order time delay interferometry (TDI2.0), and with the instrumental noise generated and modelled according to the first version of the Scientific Requirement Document (SciRDv1). We include a calibration error in the simulated data, drawn randomly from the assumed prior distribution.

The results are shown in the left-hand panel of Figure 12a, along with the prior distribution on the calibration parameters for comparison. We see that in the absence of a prior from the EM data, the posterior recovered for the amplitude calibration parameters coincides with the prior, as we might expect. We cannot decouple the unknown distance from the unknown amplitude calibration without additional information. However, once the EM prior is added, the posterior on δA_1 and δA_2 are narrower than the prior distribution, indicating that we are successfully measuring these quantities. These are the weights of the amplitude calibration spline at $f = 1\text{mHz}$ and $f = 10\text{mHz}$. All the VGBs lie at frequencies between these two knots, so it makes sense that this is where the amplitude calibration uncertainty can be measured. We see that VGBs could provide amplitude calibration at the level of a few percent.

GBs also show a similar degeneracy between the phase offset of the GW signal and the phase calibration uncertainty, and so we might hope to use EM measurements of the former to determine the latter. Unfortunately, the VGB phase offset is not normally well constrained by EM observations. However, to illustrate the possibility, we assume that we do have an EM prior on ϕ_0 and repeat the above procedure, now fitting for the system phase and the four phase calibration parameters, and imposing a Gaussian prior on the latter with width of 0.1 radians. These results are shown in the right-hand panel of Figure 12a. In this case, the degeneracy between φ_0 and the calibration parameters is not complete and so even in the absence of EM information the posterior on the phase calibration parameters, in the region where the signals lie, is different to the prior. This will be discussed further for EMRIs in the next subsection. When the EM prior is added the constraints on all the phase calibration parameters becomes tighter, with measurement precisions less than one hundredth of a radian.

We include that EM observations of VGBs can be used

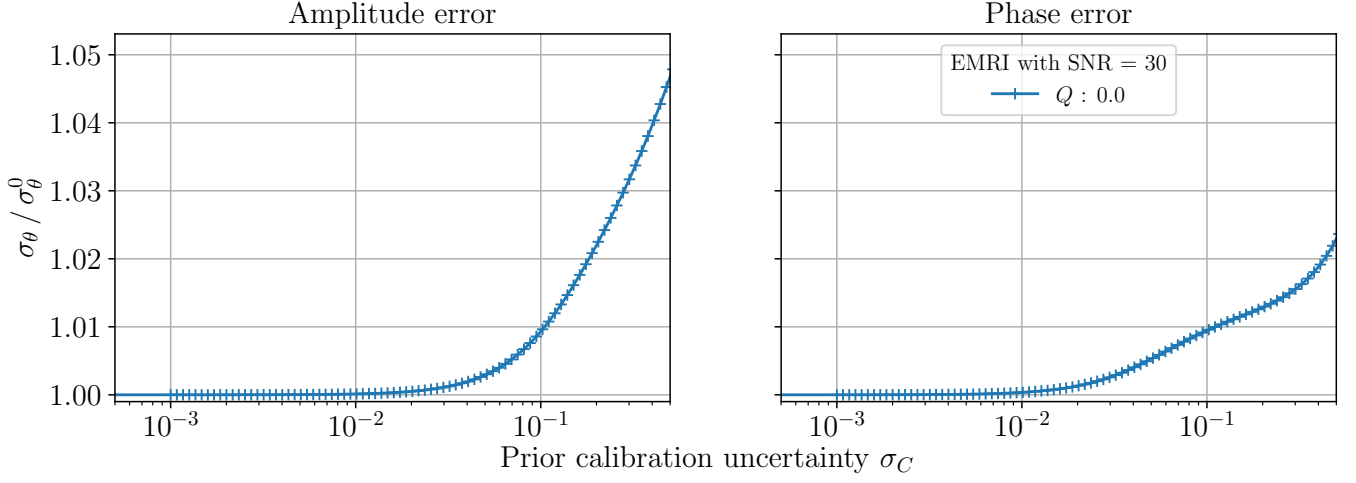


FIG. 10: As Figure 9 but now for the excess quadrupole moment of the extended AK model described in [20]. The waveform parameters used here are different to those used for Figure 9, but the results are similar for all choices of EMRI parameters.[25] [35]

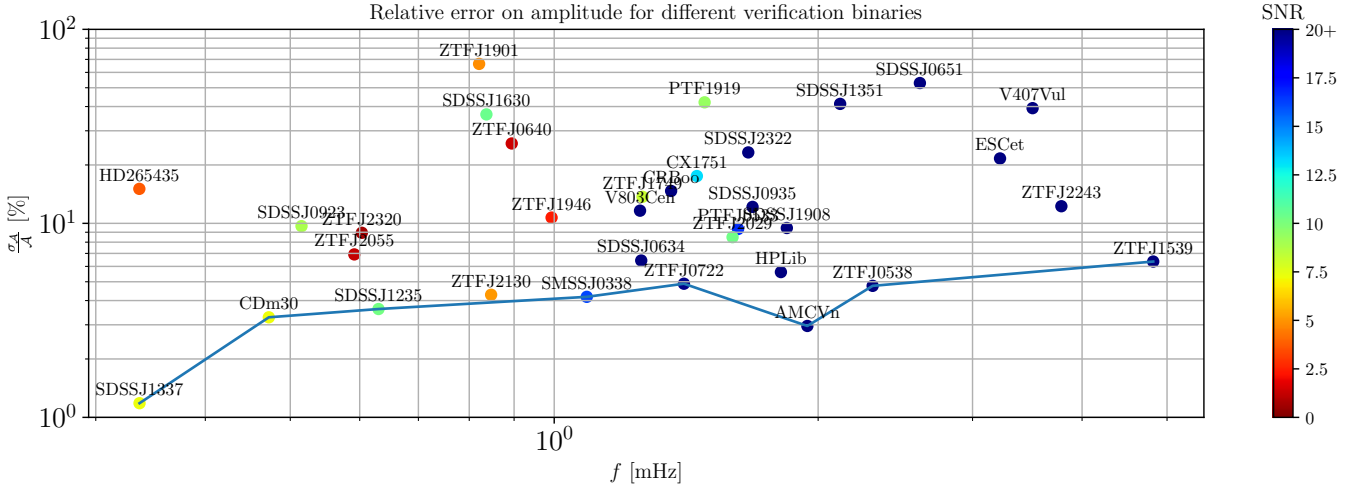


FIG. 11: Relative error on GW amplitude for different verification binaries as a function of the VGB central frequency. The SNR of each source is indicated by the color of the dot. The blue line shows the optimal choice of VGB to cover the largest frequency interval with the lowest relative error in amplitude and highest SNR.

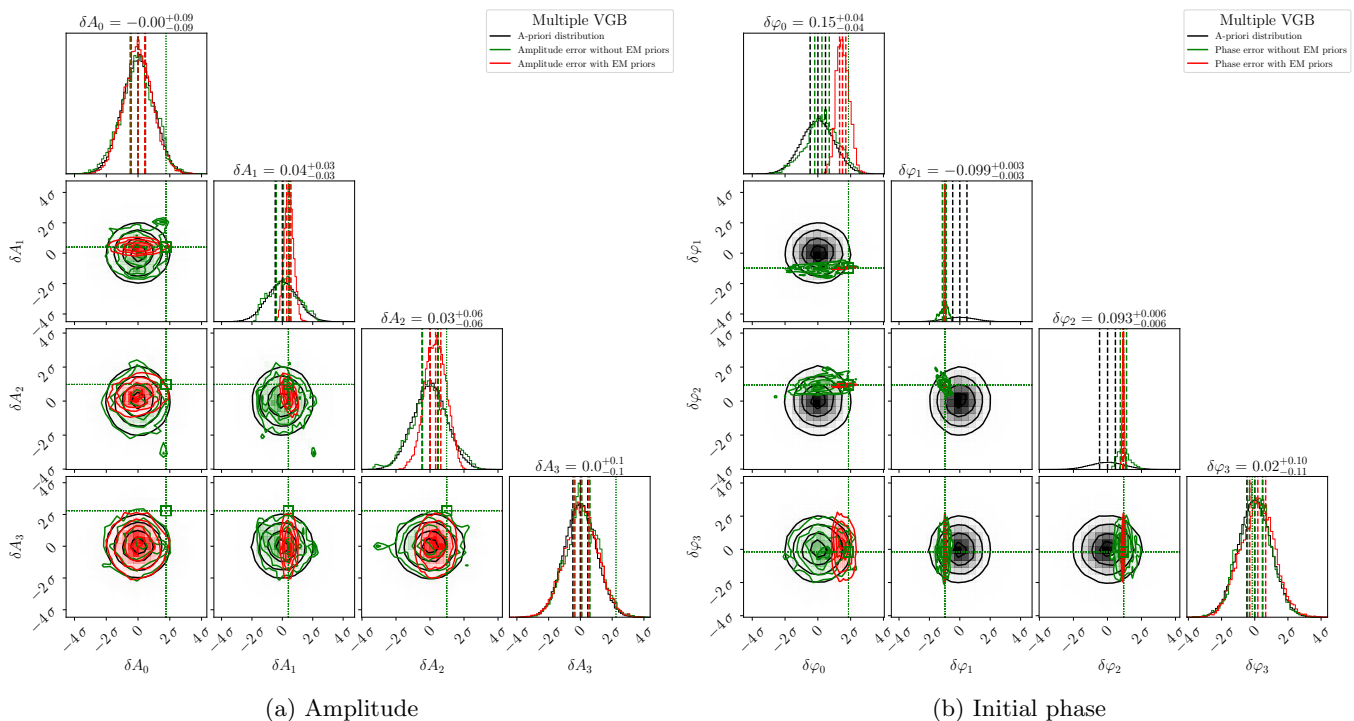


FIG. 12: Posterior distribution for the calibration error parameters in three scenarios : without any calibration error (black), with the calibration error without (green) and with (red) the electromagnetic counterparts a-priori. Left corresponds to the amplitude error parameters ($\delta A_0, \delta A_1, \delta A_2, \delta A_3$) while right corresponds to the phase error parameters ($\delta \varphi_0, \delta \varphi_1, \delta \varphi_2, \delta \varphi_3$). The dashed green lines indicate the injected value of the calibration error, which was drawn randomly from the assumed prior.

to provide information on calibration of the LISA instrument. This calibration will be limited to the frequency range where the VGBs lie, which is 0.1 to 5 mHz, and amplitude calibration will be limited by how well distances to the VGBs are known, roughly 1% at present. The publication of the third edition of the GAIA catalog <https://www.cosmos.esa.int/web/gaia/earlydr3>, expected soon, should significantly improve distance measurements of several VGBs and provide new VGB candidates. By the time LISA flies we might therefore be in a position to do somewhat better than this, although it is unlikely we will meet the necessary amplitude calibration precision of $\sim 0.1\%$ identified above. Nonetheless, it is reassuring to know that even if the calibration of LISA was completely unknown we could use this procedure to minimize the impact on LISA science objectives.

B. Inspiring binaries as phase calibrators

The previous analysis relied on using additional EM information to calibrate the LISA data stream. However, we have also seen that for many of the source types, parameter uncertainties do not grow without bound as the phase calibration uncertainties are made arbitrarily large. This is because the way in which (the assumed model

of) phase calibration errors affects the waveform phase is distinct and distinguishable from the phase evolution of the signal. This means that it should be possible to measure both simultaneously from the data, “piggy-backing” the measurement of the calibration uncertainty onto the measurement of the astrophysical signal. We explore this here, using EMRI systems, which show the most limited degradation in parameter measurements of all the sources considered here. EMRIs are expected to be on eccentric and inclined orbits, so their GW signals contain a large number of significant harmonic modes. These evolve slightly differently to one another, but are all affected by the calibration error in the same way, which is likely what facilitates separation of the two effects.

To investigate this, we repeat the same Fisher matrix analysis as before, but now look at the other diagonal elements of the inverse Fisher matrix, which correspond to the calibration error model parameters, $\{\delta \phi_i, \delta A_i\}$. These elements of the inverse Fisher matrix characterise how well we can measure the calibration parameters. If they are comparable to the assumed prior width, σ_C , then we have not gained any information from fitting these parameters at the same time as the source. If the posterior is narrower than σ_C then it shows we have measured the calibration uncertainty. We show these results in Figure 13, as a function of the assumed calibration uncertainty. We see that EMRIs cannot be used to mea-

sure the amplitude calibration uncertainty. This is to be expected as that degeneracy is perfect. Only by adding external information, as described in the previous subsection, can we hope to measure the amplitude calibration. However, we see that we can use EMRI observations to measure the phase calibration. Once the phase calibration uncertainty is greater than 0.01 radians the measured uncertainty starts to flatten off, showing that it is being determined from the data. It only flattens off for $\delta\phi_1$ and $\delta\phi_2$, but that is because these are the parameters that affect the calibration in the frequency range where the EMRI is generating GWs. We conclude that we will be able to measure the phase calibration uncertainty to better than a few times 10^{-2} radians, by fitting it simultaneously with the parameters of our sources. Once again, this is not quite at the level we argued was required to ensure no impact on science, but it is not significantly worse. So even under the most pessimistic assumptions about calibration uncertainties, LISA should still be able to produce high quality scientific results.

V. SUMMARY

We have studied the impact of calibration uncertainties, i.e., differences between the true signal component of a data set and the model used to describe that signal component in data analysis, on LISA science output. We have focused on the impact of calibration uncertainties on parameter estimation for individual sources, and have considered all the major source types that LISA is expected to observe. These results show that amplitude calibration uncertainties directly affect our ability to measure the distances to all types of gravitational wave source. Amplitude calibration uncertainty becomes limiting when it becomes comparable to the typical precision with which amplitudes can be measured, which is typically 1%. In a similar way, phase calibration uncertainties directly limit our ability to measure the absolute phase of binaries, but the initial phase is typically not a parameter that conveys a significant amount of astrophysical information. Phase calibration uncertainties also affect our ability to measure other physical parameters of the system, such as masses and spins, that determine the phase evolution of the gravitational waveform. However, in that case the impact of calibration uncertainty is limited and even large phase calibration uncertainties at worst degrade the determination of the system parameters by a factor of 2.

The impact of phase calibration uncertainties is limited because the slowly-varying (as a function of frequency) calibration error that we assume here is not degenerate with the more rapidly varying phase of a typical gravitational wave source. This effect allows us to simultaneously fit for the gravitational wave signal and the calibration error, with the gravitational wave acting as a carrier beat note for measuring the calibration uncertainty. Using extreme-mass-ratio inspirals we find that phase cali-

bration errors of the level of a few percent could be measured. Amplitude calibration uncertainty is degenerate with the unknown distance to the source and so cannot be measured in the same way. However, some galactic binary sources have distances measured as a percent level through electromagnetic observations. These verification binaries could be used to measure amplitude calibration uncertainties with a precision of a few percent.

We have focused on the impact of calibration uncertainties on parameter estimation in this paper, because it is our ability to measure the parameters of sources with high precision that drives most of the scientific objectives of the LISA mission [6]. This is not the only thing that calibration uncertainties will impact, however. Calibration errors will also impact our ability to detect sources in the data, because mismatches between the detection templates and the signal component of the data will lead to losses in signal-to-noise ratio. For the size of calibration uncertainties that are limiting for parameter estimation, this effect will be at the level of tens of percent of the total number of events. This is much smaller than the intrinsic uncertainties in the astrophysical rates of LISA events, and uncertainties arising from weak lensing of sources at cosmological distances. Additionally, the observed signal-to-noise ratio will fluctuate due to uncertainties in the noise spectral density of the detector, and these uncertainties are expected to be much larger than those in the calibration model. Calibration uncertainties will also affect our ability to infer the parameters that characterise the astrophysical population of sources, obtained by combining observations of multiple sources hierarchically. To fully understand the impact of calibration uncertainties on population inference, we would need to model the time dependence of the calibration error, making assumptions about the degree of correlation between the calibration uncertainty in different events. For the moment, we can crudely estimate that calibration uncertainties would have to be about one order of magnitude smaller to be sure they would not limit typical LISA population inference studies. Finally, we have not considered the impact of calibration uncertainties on the characterisation of stochastic gravitational wave backgrounds, as these require a different analysis approach to the one described here. However, phase calibration uncertainties should not affect our ability to characterise stochastic backgrounds, as the expected phase is random. Amplitude calibration uncertainties would be degenerate with stochastic background spectra that had a similar form, but the spline model assumed here does not resemble the power-law form of a typical stochastic or cosmological background, and the calibration uncertainty would be significantly sub-dominant to uncertainties arising from lack of knowledge of the spectral density of the instrumental noise.

The results in this paper give a baseline target for the calibration of the LISA instrument, but the analysis could be extended in a number of ways if necessary. Firstly, we have used a very slowly varying model for the

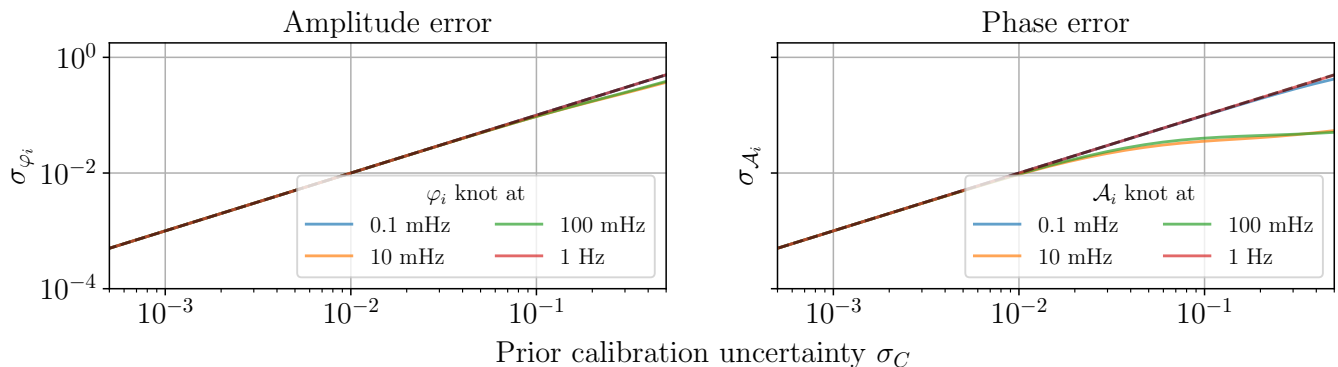


FIG. 13: Estimated uncertainty σ_θ of the calibration parameters $\delta\varphi_i, \delta A_i$ as a function of the prior calibration uncertainty σ_C , when fitting these parameters simultaneously with those of the EMRI signal represented in Figure 9. The black dashed diagonal line shows the case where the estimated uncertainty σ_θ equals the prior calibration uncertainty σ_C . Deviations from this line indicate that we are able to constrain the calibration parameters $\delta\varphi_i, \delta A_i$ using the observation.

calibration uncertainty. This was well motivated from instrumental considerations, but it does mean that the calibration uncertainty is not very degenerate with the gravitational waveform, and this is one of the reasons that phase uncertainties are not limiting. Increasing the number of spline points, or adopting an alternative model such as representing the uncertainty as a realisation of a Gaussian process, would provide the flexibility to model more rapidly varying uncertainties. Secondly, we have taken a single calibration spline to represent the whole data set, while in reality the calibration might be expected to vary over time. This could have an impact on the determination of sky position, which relies on the variation in amplitude over the course of a year. We have also ignored realistic features of the instrumental data, such as data gaps. The assumed model of the time-dependence of the calibration error will influence conclusions about population inference in particular. Finally, in this work we have only placed limits on the calibration uncertainty, but LISA science will also be limited by our degree of knowledge of the noise model. Understanding the impact of and placing requirements on our noise-model uncertainty is also important to ensure the mission delivers on its high scientific promise.

Appendix A: Setting calibration requirements based on induced parameter errors

The Fisher matrix-based approach that we use in this paper implicitly assumes that the calibration model is being fitted for at the same time as the individual sources when the data is analysed. Another approach to setting calibration requirements is to ask how large the uncor-

rected calibration error can be before it starts to lead to biases in the recovered parameters.

The bias can be estimated in the linear signal approximation using the formula

$$\Delta\theta_b^i = ((\Gamma^{\theta\theta})^{-1})_{ij}(\partial_j h | ((\mathbf{C}(f) - 1)\mathbf{h}(\vec{\theta}))). \quad (\text{A1})$$

This expression can be used to assess whether a particular calibration error, described by $\mathbf{C}(f)$, would lead to significant biases in parameter estimation, by comparing the calibration error-induced parameter uncertainty, $\Delta\theta_b^i$, to the corresponding statistical uncertainty arising from instrumental noise, which is given by the square root of the corresponding diagonal element of the inverse Fisher Matrix, $\sqrt{(\Gamma^{\theta\theta})_{ii}^{-1}}$.

This approach can be related to the joint Fisher Matrix approach adopted here by working in the linear signal approximation. We assume that the calibration model parameters are close to the reference value of zero and write

$$\mathbf{C}(f) - 1 \approx \frac{\partial \mathbf{C}}{\partial \mu_k} \Delta\mu_k.$$

The prior that the calibration errors are centred around zero with uncertainties characterised by $\Sigma^{\mu\mu}$ is equivalent to the assumption that $\Delta\mu_k \sim N(0, (\Sigma^{\mu\mu})^{-1})$. As the precise calibration error is unknown, the parameter uncertainty induced by the calibration error

$$\Delta\theta_b^i = ((\Gamma^{\theta\theta})^{-1})_{ij}(\partial_j h | \partial_k \mathbf{C} \mathbf{h}(\vec{\theta})) \Delta\mu_k$$

is a random variable. Here we are using the notation $\partial_k \mathbf{C} = \partial \mathbf{C} / \partial \mu_k$. This random variable has zero mean and variance

$$\begin{aligned} \langle \Delta\theta_b^i \Delta\theta_b^j \rangle &= ((\Gamma^{\theta\theta})^{-1})_{il} (\partial_l h | \partial_k \mathbf{C} \mathbf{h}(\vec{\theta})) \langle \Delta\mu_k \Delta\mu_m \rangle ((\Gamma^{\theta\theta})^{-1})_{jn} (\partial_n h | \partial_m \mathbf{C} \mathbf{h}(\vec{\theta})) \\ &= ((\Gamma^{\theta\theta})^{-1})_{il} (\Gamma^{\theta\mu})_{lk} (\Sigma^{\mu\mu})_{km}^{-1} ((\Gamma^{\theta\theta})^{-1})_{jn} (\Gamma^{\theta\mu})_{nm}. \end{aligned} \quad (\text{A2})$$

As this bias is independent of the noise-induced statistical uncertainty, the total variance in the parameter uncertainties is the sum of the statistical and systematic variances

$$\langle \Delta\theta^i \Delta\theta^j \rangle = (\Gamma^{\theta\theta})_{ij}^{-1} + ((\Gamma^{\theta\theta})^{-1} (\Gamma^{\theta\mu}) (\Sigma^{\mu\mu})^{-1} (\Gamma^{\theta\mu})^T ((\Gamma^{\theta\theta})^{-1})^T)_{ij}. \quad (\text{A3})$$

We can compare this to the expression used here, based on the inverse of the joint Fisher Matrix, Eq. (16), which can also be expressed in the alternative form

$$\begin{aligned} &(\Gamma^{\theta\theta})^{-1} + (\Gamma^{\theta\theta})^{-1} \Gamma^{\theta\mu} (\Gamma^{\mu\mu} + \Sigma^{\mu\mu} - \\ &(\Gamma^{\theta\mu})^T (\Gamma^{\theta\theta})^{-1} \Gamma^{\theta\mu})^{-1} (\Gamma^{\theta\mu})^T (\Gamma^{\theta\theta})^{-1}. \end{aligned} \quad (\text{A4})$$

We see that this has a similar form to expression (A3), but we have replaced $\Sigma^{\mu\mu}$ by $\Sigma^{\mu\mu} + \Gamma^{\mu\mu} - (\Gamma^{\theta\mu})^T (\Gamma^{\theta\theta})^{-1} \Gamma^{\theta\mu}$. This difference reflects the impact of fitting for the calibration model. The inverse of this term can be interpreted as the residual variance (or uncertainty) in the calibration model, which is reduced when it is fitted out of the data.

As a final remark, we note that Eq. (A1) can also be used to assess our tolerance to unmodelled calibration effects. In practice, we will fit for some kind of calibration uncertainty model at the same time as fitting for the parameters of the sources present in the data. Any component of the true calibration uncertainty that can not be represented by that model will remain in the data and lead to biases. These biases can be estimated using the above formula, but with $(\mathbf{C}(f) - 1)$ replaced by the residual calibration error $\mathbf{C}_{\text{true}}(f; \mu) - \mathbf{C}_{\text{mod}}(f; \mu)$ and the set of parameters and Fisher Matrix expanded to include the calibration model parameters. We will not consider this further here, but such biases could be assessed once numerical calibration uncertainty simulations from instrumental modelling are available.

Appendix B: Cross-checks of the Fisher matrix results

The Fisher matrix results presented in this paper were computed numerically. In this appendix we report a num-

ber of cross-checks that were used to verify the validity of these results. We derive the Fisher matrix for Galactic binaries analytically and that for SBHBs semi-analytically. We also verify the GB results by directly evaluating a posterior on the calibration parameters using MCMC methods. To facilitate the analytic calculations, we first introduce an alternative model for the calibration error.

1. Calibration error knots formulae

In the main part of this paper, the calibration error is described using two cubic splines. In order to verify the validity and stability of the Fisher matrix, we will use in this appendix a calibration error that is just a simple cubic to simplify the analytical derivation. While the cubic spline model has considerably more flexibility in general, when using only four spline knots it is not so different to a cubic, and using the latter as an alternative model allows us to derive analytically (respectively semi-analytically) the Fisher matrix for a GB (resp. SBHB).

The amplitude (respectively phase) error function is a cubic that is defined with 4 parameters $\{\delta A_i\}$ (resp. $\{\delta\varphi_i\}$) defined at specific knots $\{0.1, 1, 10, 1000\}$. The total calibration error can be written:

$$C(f, \{\delta A_i, \delta\varphi_i\}) = \left(1 + \sum_{n=0}^3 a_n (\{\delta A_i\}) (\log_{10} f)^n \right) \times e^{2\pi i \sum_{n=0}^3 p_n (\{\delta\varphi_i\}) (\log_{10} f)^n}, \quad (\text{B1})$$

where $\{a_i\}$ and $\{p_i\}$ corresponds to the cubic function parameters and are linear combination of $\{\delta A_i\}$ and $\{\delta\varphi_i\}$

Using $\delta X(f)$ to denote either the amplitude or the phase error function we have:

$$\delta\tilde{X}(f) = x_0 + x_1 (\log_{10} f) + x_2 (\log_{10} f)^2 + x_3 (\log_{10} f)^3 \quad (\text{B2})$$

where $\{x_i\}$ are the cubic spline parameters, which relate to the weights at the knots, $\{\delta X_i\}$, via

$$\begin{cases} \delta X_0 = \delta X(f_0) = x_0 + x_1 (\log_{10} f_0) + x_2 (\log_{10} f_0)^2 + x_3 (\log_{10} f_0)^3 \\ \delta X_1 = \delta X(f_1) = x_0 + x_1 (\log_{10} f_1) + x_2 (\log_{10} f_1)^2 + x_3 (\log_{10} f_1)^3 \\ \delta X_2 = \delta X(f_2) = x_0 + x_1 (\log_{10} f_2) + x_2 (\log_{10} f_2)^2 + x_3 (\log_{10} f_2)^3 \\ \delta X_3 = \delta X(f_3) = x_0 + x_1 (\log_{10} f_3) + x_2 (\log_{10} f_3)^2 + x_3 (\log_{10} f_3)^3 \end{cases}. \quad (\text{B3})$$

In our case, $f_0 = 10^{-4}$, $f_1 = 10^{-3}$, $f_2 = 10^{-2}$ and $f_3 = 1$ Hz. Writing these equations in matrix form, we can

$$\begin{pmatrix} \delta X_0 \\ \delta X_1 \\ \delta X_2 \\ \delta X_3 \end{pmatrix} = \begin{pmatrix} 1 & -4 & 16 & -64 \\ 1 & -3 & 9 & -27 \\ 1 & -2 & 4 & -8 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ -\frac{3}{4} & \frac{8}{3} & -3 & \frac{13}{12} \\ -\frac{5}{8} & 2 & -\frac{7}{4} & \frac{5}{8} \\ -\frac{1}{8} & \frac{1}{3} & -\frac{1}{4} & \frac{1}{24} \end{pmatrix} \begin{pmatrix} \delta X_0 \\ \delta X_1 \\ \delta X_2 \\ \delta X_3 \end{pmatrix}. \quad (\text{B4})$$

2. Galactic binary analytical confirmation

Here we recompute the Fisher matrix for GBs analytically. The noiseless TDI combinations (here X) used to simulate the gravitational wave can be represented with the following formula :

$$\tilde{d}(f) \equiv X(f, \theta, \mu) = C(f, \mu) R_X(f, \theta) \mathcal{A} e^{i\Phi(f, \theta)} \quad (\text{B5})$$

where θ are the waveform parameters, μ the calibration error parameters, C is the calibration error function, R_X is the LISA transfer function for the TDI X combination, \mathcal{A} is the waveform amplitude and $\Phi(t)$ is the waveform phase defined in equation (17). For GBs the LISA transfer function R_X only depends on the sky location and is independent from the intrinsic waveform parameters.

In the main part of the paper, we numerically simulated the 3 TDI combinations and then calculated the Fisher matrix via equation (13). To do this, it is necessary to calculate the derivative of each TDI combination with respect to the different parameters (waveform + calibration). Since the calibration error only affects the amplitude and initial phase of the gravitational wave (as seen in Figure 2), we will concentrate on these waveform parameters only :

$$\begin{aligned} \frac{dX(f, \theta, \mu)}{d\mathcal{A}} &= \frac{1}{\mathcal{A}} X(f, \theta, \mu) \\ \frac{dX(f, \theta, \mu)}{d\phi} &= -iX(f, \theta, \mu) \\ \frac{dX(f, \theta, \mu)}{da_n} &= (\log_{10}(f))^n X(f, \theta, \mu) \\ \frac{dX(f, \theta, \mu)}{dp_n} &= 2\pi i (\log_{10}(f))^n X(f, \theta, \mu) \end{aligned} \quad (\text{B6})$$

where we used the calibration error function defined in equation (B1) and we evaluate the Fisher matrix for zero calibration error, to ensure no bias, as discussed in appendix A.

As the source is a galactic binary, we can assume the waveform frequency is equal to the central frequency of the GB ($f \rightarrow f_0$). Using this transformation, the definition of the inner product integral is simplified :

$$\langle a|b \rangle \equiv 4\text{Re} \left[\frac{\tilde{a}^*(f_0) \tilde{b}(f_0)}{S_n(f_0)} \right] \Delta f \quad (\text{B7})$$

readily invert them to obtain the relationship between the cubic spline parameters and the error parameters :

where Δf is the Fourier transform resolution of the \tilde{a} and \tilde{b} quantities.

The non zero Fisher matrix elements for the waveform amplitude, waveform initial phase and calibration error parameters can thus be analytically calculated :

$$\begin{aligned} \Gamma^{\mathcal{A}, \mathcal{A}} &= \left(\frac{\rho}{\mathcal{A}} \right)^2, \quad \Gamma^{\phi, a_n} = \rho^2 \frac{\log_{10}(f)^n}{\mathcal{A}}, \quad \Gamma^{\phi, \phi} = -\rho^2, \\ \Gamma^{\phi, p_n} &= 2\pi \rho^2 (-\log_{10}(f))^n, \quad \Gamma^{a_n, a_m} = \rho^2 (\log_{10}(f))^{n+m}, \\ \Gamma^{p_n, p_m} &= 4\pi^2 \rho^2 (\log_{10}(f))^{n+m} \end{aligned} \quad (\text{B8})$$

where we have defined the signal-to-noise ratio via:

$$\rho = 4X(f_0)^2 / S_n(f_0) \Delta f. \quad (\text{B9})$$

From equation (16), we can extract the ‘‘calibrated’’ waveform amplitude and initial phase uncertainties (index c) compared to the nominal uncertainties in the absence of calibration uncertainty (index 0) :

$$\begin{aligned} \sigma_{\mathcal{A}_c}^2 &= \sigma_{\mathcal{A}_0}^2 + \mathcal{A}_0^2 \sum_n (\log_{10}(f_0))^{2n} \sigma_{a_n}^2 \\ \sigma_{\phi_c}^2 &= \sigma_{\phi_0}^2 + 4\pi^2 \sum_n (\log_{10}(f_0))^{2n} \sigma_{p_n}^2 \end{aligned} \quad (\text{B10})$$

where σ_{a_n} and σ_{p_n} are the calibration error spline-function parameters defined in equation (B1). These analytical predictions are shown along with the numerical results in Figure 14 and we see they are in full agreement.

3. Galactic binary MCMC confirmation

The Fisher matrix is an approximation to the shape of the posterior, valid in the limit of high signal-to-noise ratio. To verify that it is valid for the calculations being performed here, we will now cross-check against results based on the numerical evaluation of the full Bayesian posterior distribution computed via Markov Chain Monte Carlo (MCMC) sampling. We proceed as follows. For each value of the calibration error uncertainty, σ_C , we generate a large number (2048 in this case) of realisations of a data set containing the same GB source plus calibration error with parameters drawn randomly from $\mathcal{N}(0, \sigma_C^2)$. For each realisation, we find the peak of the posterior numerically and hence evaluate the bias in the recovered amplitude and phase of the

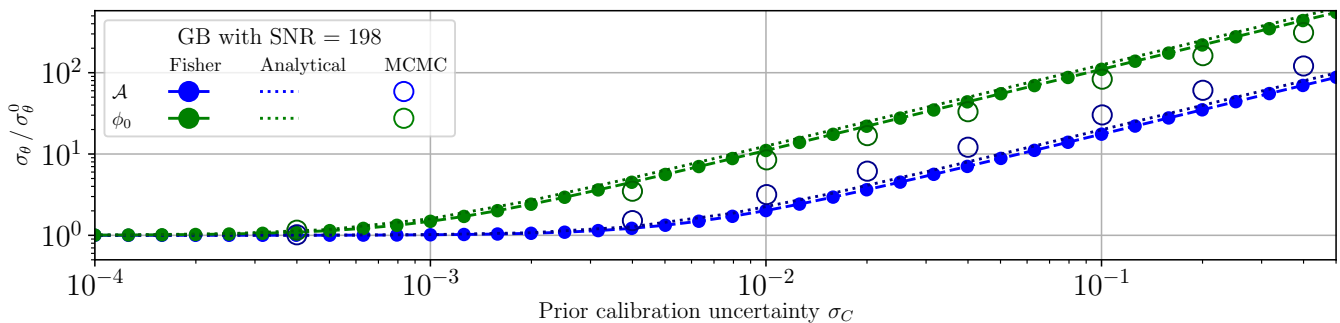


FIG. 14: Ratio between the uncertainty of a waveform parameter with (σ_θ) and without (σ_{θ_0}) calibration error uncertainty as a function of the calibration error uncertainty σ_C . The dotted lines represent the numerical Fisher matrix results, which are in full agreement with the analytical predictions shown with solid lines.

GW. The distribution of biases over the realisations is Gaussian, with zero mean (because the data generating process and likelihood are consistent), and standard deviation σ_A for the GW amplitude and σ_{ϕ_0} for the GW phase. If the Fisher matrix is valid, these standard deviations should coincide with the Fisher matrix prediction.

The results from this procedure as shown as open circles in Figure 14. While the agreement is not perfect, it is within a factor of 2 for all choices of σ_C , indicating the Fisher matrix is providing a very reliable guide to the impact of calibration uncertainty, at a much smaller computational cost than MCMC.

4. Stellar-origin black hole binary semi-analytical confirmation

In this section we now verify the Fisher matrix for SBHBs. This cannot be done fully analytically, but it is possible to do it without computing numerical derivatives. For an SBHB, the noiseless TDI combinations (here X) used to describe the observed gravitational wave can be represented as:

$$\tilde{d}(f) \equiv X(f, \theta, \mu) = C(f, \mu) R_X(f, \theta) \mathcal{A}(f, \theta) e^{i\Phi(t)(f, \theta)} \quad (\text{B11})$$

where θ are the waveform parameters, μ the calibration error parameters, C is the calibration error function, R_X is the LISA transfer function for the TDI X combination, \mathcal{A} is the waveform amplitude and $\Phi(t)$ is the waveform phase defined in equation (17). For this source type the LISA transfer function R_X depends on all the waveform parameters.

As shown in the main section of the paper, the SNR is very low for this kind of source. In this appendix, we will artificially increase the SNR (by decreasing the noise power spectral density) in order to reveal the impact of the calibration error. As seen in Figure 15, the initial phase and the distance are mostly affected by the cali-

bration error. Unlike GBs, SBHBs have a broad spectral signature that does not allow the integral over frequency to be approximated by evaluating the integrand at the central frequency, so we still have to evaluate the elements via a numerical integration. However, we can directly write down the derivatives that appear in those integrals. Since the effect on the initial phase is equivalent to that in the case of GBs (subsection B2), we will focus on the chirp mass \mathcal{M}_c and the symmetric mass ratio η in this subsection. The generic derivative of the TDI combination X with respect to a parameter θ is given by :

$$\frac{dX(f, \theta, \mu)}{d\theta} = \left(\frac{1}{R_x(f, \theta)} \frac{\partial R_x(f, \theta)}{\partial \theta} + \frac{1}{\mathcal{A}(f, \theta)} \frac{\partial \mathcal{A}(f, \theta)}{\partial \theta} + i \frac{\partial \Phi(f, \theta)}{\partial \theta} \right) X(f, \theta, \mu) \quad (\text{B12})$$

For the chirp mass \mathcal{M}_c and symmetric mass ratio η , the transfer function R_x and amplitude \mathcal{A} derivatives can be neglected compared to the phase Φ derivative :

$$\begin{aligned} \frac{dX(f, \theta, \mu)}{d\mathcal{M}_c} &\simeq i \frac{\partial \Phi(f, \theta)}{\partial \mathcal{M}_c} X(f, \theta, \mu), \\ \frac{dX(f, \theta, \mu)}{d\eta} &\simeq i \frac{\partial \Phi(f, \theta)}{\partial \eta} X(f, \theta, \mu) \end{aligned} \quad (\text{B13})$$

where the partial derivatives of the phase are analytical functions that can be derived from the IMRPhenomD formulae. The derivatives with respect to the calibration error parameters are the same as for GBs (equation B6). Evaluating the integrals of products of these various derivative expressions over frequency we can obtain an alternative semi-analytic computation of the Fisher matrix. This is compared to the fully numerical results in Figure 15. We can see that the two sets of results are in perfect agreement.

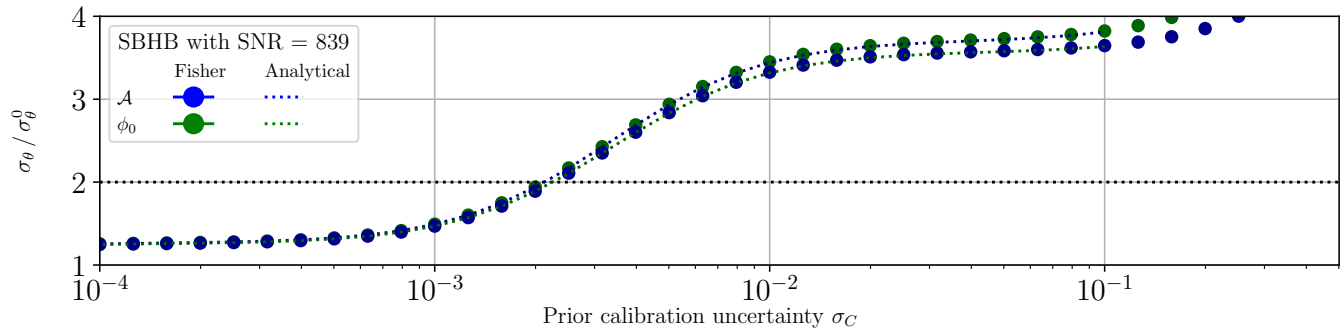


FIG. 15: As Figure 14, but now for the SBHB source. The dotted line shows the result of the semi-analytic Fisher matrix evaluation described in this appendix, while the open circles show the results from full numerical evaluation of the Fisher matrix.

-
- [1] Massimo Tinto and Sanjeev V. Dhurandhar. Time-delay interferometry. *Living Rev. Rel.*, 24(1):1, 2021.
- [2] B. P. Abbott et al. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. *Phys. Rev. D*, 95(6):062003, 2017.
- [3] Curt Cutler and Michele Vallisneri. LISA detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms. *Phys. Rev. D*, 76(10):104018, November 2007.
- [4] Salvatore Vitale, Walter Del Pozzo, Tjonnie G. F. Li, Chris Van Den Broeck, Ilya Mandel, Ben Aylott, and John Veitch. Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the Advanced Detectors era. *Phys. Rev. D*, 85:064034, 2012.
- [5] Salvatore Vitale, Carl-Johan Haster, Ling Sun, Ben Farr, Evan Goetz, Jeff Kissel, and Craig Cahillane. Physical approach to the marginalization of ligo calibration uncertainties. *Phys. Rev. D*, 103:063016, Mar 2021.
- [6] Pau Amaro Seoane et al. The Gravitational Universe. 5 2013.
- [7] Benjamin J. Owen and B. S. Sathyaprakash. Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement. *Phys. Rev. D*, 60:022002, Jun 1999.
- [8] Jeff Crowder and Neil Cornish. A Solution to the Galactic Foreground Problem for LISA. *Phys. Rev. D*, 75:043008, 2007.
- [9] Arkadiusz Blaut, Stanislav Babak, and Andrzej Krolak. Mock LISA Data Challenge for the galactic white dwarf binaries. *Phys. Rev. D*, 81:063008, 2010.
- [10] Tyson B. Littenberg. A detection pipeline for galactic binaries in LISA data. *Phys. Rev. D*, 84:063009, 2011.
- [11] Tyson Littenberg, Neil Cornish, Kristen Lackeos, and Travis Robson. Global Analysis of the Gravitational Wave Signal from Galactic Binaries. *Phys. Rev. D*, 101(12):123021, 2020.
- [12] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, 2016.
- [13] Leor Barack and Curt Cutler. LISA Capture Sources: Approximate Waveforms, Signal-to-Noise Ratios, and Parameter Estimation Accuracy. *Phys. Rev. D*, 69(8):082005, April 2004.
- [14] K. G. Arun, Stas Babak, Emanuele Berti, Neil Cornish, Curt Cutler, Jonathan Gair, Scott A. Hughes, Bala R. Iyer, Ryan N. Lang, Ilya Mandel, Edward K. Porter, Bangalore S. Sathyaprakash, Siddhartha Sinha, Alicia M. Sintes, Miquel Trias, Chris Van Den Broeck, and Marta Volonteri. Massive Black Hole Binary Inspirals: Results from the LISA Parameter Estimation Taskforce. *Class. Quantum Grav.*, 26(9):094027, May 2009.
- [15] Enrico Barausse, Vitor Cardoso, and Paolo Pani. Can environmental effects spoil precision gravitational-wave astrophysics? *Phys. Rev. D*, 89(10):104059, May 2014.
- [16] Jonathan R. Gair, Alberto Sesana, Emanuele Berti, and Marta Volonteri. Constraining properties of the black hole population using LISA. *Class. Quantum Grav.*, 28(9):094018, May 2011.
- [17] Jonathan R. Gair, Christopher Tang, and Marta Volonteri. LISA extreme-mass-ratio inspiral events as probes of the black hole mass function. *Phys. Rev. D*, 81(10):104014, May 2010.
- [18] Pau Amaro-Seoane, Jonathan R. Gair, Marc Freitag, M. Coleman Miller, Ilya Mandel, Curt J. Cutler, and Stanislav Babak. Intermediate and Extreme Mass-Ratio Inspirals – Astrophysics, Science Applications and Detection using LISA. *Class. Quantum Grav.*, 24(17):R113–R169, September 2007.
- [19] Jonathan R. Gair, Michele Vallisneri, Shane L. Larson, and John G. Baker. Testing General Relativity with Low-Frequency, Space-Based Gravitational-Wave Detectors. *Living Rev. Relativ.*, 16(1):7, December 2013.
- [20] Leor Barack and Curt Cutler. Using LISA EMRI sources to test off-Kerr deviations in the geometry of massive black holes. *Phys. Rev. D*, 75(4):042003, February 2007.
- [21] Chelsea L. MacLeod and Craig J. Hogan. Precision of Hubble constant derived using black hole binary absolute distances and statistical redshift information. *Phys. Rev. D*, 77(4):043512, February 2008.
- [22] Danny Laghi, Nicola Tamanini, Walter Del Pozzo, Al-

- berto Sesana, Jonathan Gair, and Stanislav Babak. Gravitational wave cosmology with extreme mass-ratio inspirals, 2021.
- [23] Jonathan R. Gair, Leor Barack, Teviet Creighton, Curt Cutler, Shane L. Larson, E. Sterl Phinney, and Michele Vallisneri. Event rate estimates for LISA extreme mass ratio capture sources. *Class. Quantum Grav.*, 21(20):S1595–S1606, October 2004.
- [24] Jonathan R. Gair. Probing black holes at low redshift using LISA EMRI observations. *Class. Quantum Grav.*, 26(9):094034, April 2009.
- [25] Stanislav Babak, Jonathan Gair, Alberto Sesana, Enrico Barausse, Carlos F. Sopuerta, Christopher P. L. Berry, Emanuele Berti, Pau Amaro-Seoane, Antoine Petiteau, and Antoine Klein. Science with the space-based interferometer LISA. V: Extreme mass-ratio inspirals. *Phys. Rev. D*, 95(10):103012, May 2017.
- [26] Zhen Pan and Huan Yang. Formation rate of extreme mass ratio inspirals in active galactic nucleus, 2021.
- [27] Pau Amaro-Seoane. Relativistic Dynamics and Extreme Mass Ratio Inspirals. *Living Rev Relativ*, 21(1):4, December 2018.
- [28] Michael L. Katz, Alvin J. K. Chua, Lorenzo Speri, Niels Warburton, and Scott A. Hughes. FastEMRIWaveforms: New tools for millihertz gravitational-wave data analysis. 4 2021.
- [29] Michael L. Katz, Alvin J. K. Chua, Niels Warburton, and Scott A. Hughes. BlackHolePerturbation-Toolkit/FastEMRIWaveforms: Official Release, August 2020.
- [30] Alvin J.K. Chua and Jonathan R. Gair. Improved analytic extreme-mass-ratio inspiral model for scoping out eLISA data analysis. *Class. Quant. Grav.*, 32:232002, 2015.
- [31] Alvin J.K. Chua, Christopher J. Moore, and Jonathan R. Gair. Augmented kludge waveforms for detecting extreme-mass-ratio inspirals. *Phys. Rev. D*, 96(4):044005, 2017.
- [32] Ryuichi Fujita and Masaru Shibata. Extreme mass ratio inspirals on the equatorial plane in the adiabatic order. *Phys. Rev. D*, 102(6):064005, 2020.
- [33] Leo C. Stein and Niels Warburton. Location of the last stable orbit in Kerr spacetime. *Phys. Rev. D*, 101(6):064007, 2020.
- [34] Curt Cutler. Angular Resolution of the LISA Gravitational Wave Detector. *Phys. Rev. D*, 57(12):7089–7102, June 1998.
- [35] Leor Barack and Curt Cutler. Using lisa extreme-mass-ratio inspiral sources to test off-kerr deviations in the geometry of massive black holes. *Physical Review D*, 75(4), Feb 2007.
- [36] T Kupfer, V Korol, S Shah, G Nelemans, T R Marsh, G Ramsay, P J Groot, D T H Steeghs, and E M Rossi. LISA verification binaries with updated distances from Gaia Data Release 2. *Monthly Notices of the Royal Astronomical Society*, 480(1):302–309, 06 2018.