PHYSICAL REVIEW D **86,** 124032 (2012)

Astrophysical model selection in gravitational wave astronomy

Matthew R. Adams and Neil J. Cornish
Department of Physics, Montana State University, Bozeman, Montana 59717, USA

Tyson B. Littenberg
Maryland Center for Fundamental Physics, Department of Physics, University of Maryland, College Park, Maryland 20742, USA and Gravitational Astrophysics Laboratory, NASA Goddard Spaceflight Center, 8800 Greenbelt Road, Greenbelt, Maryland 20771, USA

(Received 3 October 2012; published 18 December 2012)

Theoretical studies in gravitational wave astronomy have mostly focused on the information that can be extracted from individual detections, such as the mass of a binary system and its location in space. Here we consider how the information from multiple detections can be used to constrain astrophysical population models. This seemingly simple problem is made challenging by the high dimensionality and high degree of correlation in the parameter spaces that describe the signals, and by the complexity of the astrophysical models, which can also depend on a large number of parameters, some of which might not be directly constrained by the observations. We present a method for constraining population models using a hierarchical Bayesian modeling approach which simultaneously infers the source parameters and population model and provides the joint probability distributions for both. We illustrate this approach by considering the constraints that can be placed on population models for galactic white dwarf binaries using a future space-based gravitational wave detector. We find that a mission that is able to resolve ~5000 of the shortest period binaries will be able to constrain the population model parameters, including the chirp mass distribution and a characteristic galaxy disk radius to within a few percent. This compares favorably to existing bounds, where electromagnetic observations of stars in the galaxy constrain disk radii to within 20%.

DOI: 10.1103/PhysRevD.86.124032 PACS numbers: 04.30.−w, 04.80.Nn

I. INTRODUCTION

There is an old joke in astrophysics that with one source you have a discovery, and with two you have a population. With a population of sources it becomes possible to constrain astrophysical models. Until recently, studies of milli-Hertz gravitational wave science have either focused on making predictions about the source populations [1–3], or have looked at detection and parameter estimation for individual source types [4–6]. These types of studies have featured heavily in the science assessment of alternative space-based gravitational wave mission concepts, where metrics such as detection numbers and histograms of the parameter resolution capabilities for fiducial population models were used to rate science performance (see, e.g., Ref. [7]). These are certainly useful metrics, but they only tell part of the story. A more powerful and informative measure of the science capabilities is the ability to discriminate between alternative population models.

Inferring the underlying population model, and the attendant astrophysical processes responsible for the observed source distribution, from the time series of a gravitational wave detector is the central science challenge for a future space mission. It folds together the difficult task of identifying and disentangling the multiple overlapping signals that are in the data, inferring the individual source parameters, and reconstructing the true population distributions from incomplete and imperfect information.

The past few years have seen the first studies of the astrophysical model selection problem in the context of space-based gravitational astronomy. Gair and collaborators [8–11] have looked at how extreme mass ratio inspiral formation scenarios and massive black hole binary assembly scenarios can be constrained by gravitational wave observations using Bayesian model selection with a Poisson likelihood function. Plowman and collaborators [12,13] have performed similar studies of black hole population models using a frequentist approach based on error kernels and the Kolmogorov-Smirnov test. Related work on astrophysical model selection for ground based detectors can be found in Refs. [14,15].

We develop a simple yet comprehensive hierarchical Bayesian modeling approach that uses the full multidimensional and highly correlated parameter uncertainties of a collection of signals to constrain the joint parameter distributions of the underlying astrophysical models. The method is general and can be applied to any number of astrophysical model selection problems [16–18].

A remarkable feature of the hierachial Bayesian method is that in its purest form it is completely free of selection effects such as the Malmquist bias. By ‘‘purest form’’ we mean where the signal model extends over the entire source population, including those with vanishingly small signal-to-noise ratio [19]. In practice it is unclear how to include arbitrarily weak sources in the analysis, and in any

case the computational cost would be prohibitive, so we are forced to make some kind of selection cuts on the signals, and this will introduce a bias if left uncorrected [20].

To illustrate the hierachical Bayesian approach and to investigate where a bias can arise, we look at the problem of determining the population model for white dwarf binaries in the Milky Way. Future space-based missions are expected to detect thousands to tens of thousands of white dwarf binaries [11,21–24]. Here we focus on determining the spatial distribution and the chirp mass distribution, but in future work we plan to extend our study to include a wider class of population characteristics such as those described in Ref. [22]. Determining the galaxy shape using gravitational wave observations of white dwarf binaries will be an independent measure on the shape of the galaxy to complement electromagnetic observations. Additionally, the white dwarf binaries that are not detectable form a very bright stochastic foreground. Accurately modeling the confusion foreground level is crucial for the detection of extragalactic stochastic gravitational wave signals [25].

The paper is organized as follows: The hierarchical Bayesian approach is described in Sec. II and is illustrated using a simple toy model in Sec. III. A more realistic toy model is developed in Sec. IV to explore mismodeling biases that can occur when using Gaussian approximations to the likelihood function. In Sec. V the method is applied to simulated observations of galactic white dwarf binaries, and in Sec. the possibility of using the Fisher information matrix approximation to the likelihood is explored. Concluding thoughts follow in Sec. VII.

II. HIERARCHICAL BAYESIAN MODELING

Hierarchical Bayesian modeling has been around since at least the 1950s [26–29], but it is only now becoming widely known and used. The term ''hierarchical'' arises because the analysis has two levels. At the highest level are the space of models being considered, and at the lower level are the parameters of the models themselves. Hierachical Bayes provides a method to simultaneously perform model selection and parameter estimation. In this work we will consider models of fixed dimension that can be parametrized by smooth functions of one or more hyperparameters. A hyperparameter is a parameter of the prior distribution or the likelihood function. The joint posterior distribution for the model parameters $\vec{\lambda}$ and the hyperparameters $\vec{\alpha}$ given data s follows from Bayes' theorem:

$$p(\vec{\lambda}, \vec{\alpha}|s) = \frac{p(s|\vec{\lambda}, \vec{\alpha})p(\vec{\lambda}|\vec{\alpha})p(\vec{\alpha})}{p(s)}, \quad (1)$$

where $p(s|\vec{\lambda}, \vec{\alpha})$ is the likelihood, $p(\vec{\lambda}|\vec{\alpha})$ is the prior on the model parameters for a model described by hyperparameters, $\vec{\alpha}$, $p(\vec{\alpha})$ is the hyperprior and $p(s)$ is a normalizing factor

$$p(s) = \int p(s, \vec{\alpha})d\vec{\alpha} = \int p(s|\vec{\lambda}, \vec{\alpha})p(\vec{\lambda}|\vec{\alpha})p(\vec{\alpha})d\vec{\lambda}d\vec{\alpha}. \quad (2)$$

The quantity $p(s, \vec{\alpha})$ can be interpreted as the ''density of evidence'' for a model with hyperparameters $\vec{\alpha}$.

The integral marginalizing over the hyperparameters is often only tractable numerically, and this can be computationally expensive. Empirical Bayes is a collection of methods that seek to estimate the hyperparameters in various ways from the data [30,31]. Markov chain Monte Carlo (MCMC) techniques allow us to implement hierachical Bayesian modeling without approximation by producing samples from the joint posterior distributions, which simultaneously informs us about the model parameters $\vec{\lambda}$ and the hyperparameters $\vec{\alpha}$. This approach helps reduce systematic errors due to mismodeling, as the data help select the appropriate model. An example of this is the use of hyperparameters in the instrument noise model, such that the noise spectral density is treated as an unknown to be determined from the data [25,32,33].

Hierarchical Bayesian modeling can be extended to discrete and even disjoint model spaces using the reverse jump Markov chain Monte Carlo [34] algorithm. Each discrete model can be assigned its own set of continuous hyperparameters.

III. TOY MODEL I

As a simple illustration of hierarchical Bayesian modeling, consider some population of N signals, each described by a single parameter x_i that is drawn from a normal distribution with standard deviation α_0. The measured values of these parameters are affected by instrument noise that is drawn from a normal distribution with standard deviation β. The maximum likelihood value for the parameters is then $\bar{x}_i = \alpha_0\delta_1 + \beta\delta_2$ where the δ's are independent and identically distributed unit standard deviates. Now suppose that we employ a population model where the parameters are distributed according to a normal distribution with standard deviation α. Each choice of α corresponds to a particular model with posterior distribution

$$p(x_i|s, \alpha) = \frac{1}{p(s, \alpha)}\prod_{i=1}^{N}\frac{1}{(2\pi\alpha\beta)}e^{-(\bar{x}_i - x_i)^2/2\beta^2}e^{-x_i^2/2\alpha^2} \quad (3)$$

and model evidence

$$p(s, \alpha) = \frac{1}{(\sqrt{2\pi}\sqrt{\alpha^2 + \beta^2})^N}\prod_i e^{-\bar{x}_i^2/2(\alpha^2+\beta^2)}. \quad (4)$$

To arrive at a hierarchical Bayesian model we elevate α to a hyperparameter and introduce a hyperprior $p(\alpha)$ which yields the joint posterior distribution

$$p(x_i, \alpha|s) = \frac{p(x_i|s, \alpha)p(\alpha)}{p(s)}. \quad (5)$$

Rather than selecting a single "best fit" model, hierarchical Bayesian methods reveal the range of models that are consistent with the data. In the more familiar, nonhierarchical approach we would maximize the model evidence (4) to find the model that best describes the data, which is here given by

$$\alpha^2_{\rm ME} = \frac{1}{N}\sum_{i=1}^{N} \bar{x}_i^2 - \beta^2. \quad (6)$$

Since $\mathrm{Var}(\bar{x}_i) = \alpha_0^2 + \beta^2$, we have

$$\alpha^2_{\rm ME} = \alpha_0^2 \pm \mathcal{O}(\sqrt{2}(\alpha_0^2 + \beta^2)/\sqrt{N}). \quad (7)$$

The error estimate comes from the sample variance of the variance estimate. In the limit that the experimental errors β are small compared to the width of the prior α_0, the error in α scales uniformly as $1/\sqrt{N}$. The scaling is more complicated when we have a collection of observations with a range of measurement errors. Suppose that the measurement errors are large compared to the width of the prior, and that we have N_1 observations with standard error β_1, N_2 observations with standard error β_2, etc., then the error in the estimate for α is

$$\Delta\alpha^2 = \left(\sum_i \frac{N_i}{\beta_i^4}\right)^{-1/2}. \quad (8)$$

Recalling that $1/\beta_i$ scales with the signal-to-noise ratio of the observation, we see that a few high signal-to-noise ratio (SNR) observations constrain α far more effectively than a large number of low SNR observations.

The above calculation shows that the maximum evidence criteria provides an unbiased estimator for the model parameter α_0, but only if the measurement noise is consistently included in both the likelihood and the simulation of the $\bar{x}_i$. Using the likelihood from Eq. (3) but failing to include the noise in the simulations leads to the biased estimate $\alpha^2_{\rm ME} = \alpha_0^2 - \beta^2$. Conversely, including noise in the simulation and failing to account for it in the likelihood leads to the biased estimate $\alpha^2_{\rm ME} = \alpha_0^2 + \beta^2$. These same conclusions apply to the hierarchical Bayesian approach, as we shall see shortly.

A. Numerical simulation

The joint posterior distribution (5) can be explored using MCMC techniques. To do this we produced simulated data with $N = 1000$, $\alpha_0 = 2$ and $\beta = 0.4$ and adopted a flat hyperprior for α. The posterior distribution function for α, marginalized over the x_i, is shown in Fig. 1. The distribution includes the injected value and has a spread consistent with the error estimate of Eq. (7). The maximum-*a-posteriori* (MAP) estimate for α has been displaced from the injected value of $\alpha_0 = 2$ by the simulated noise.

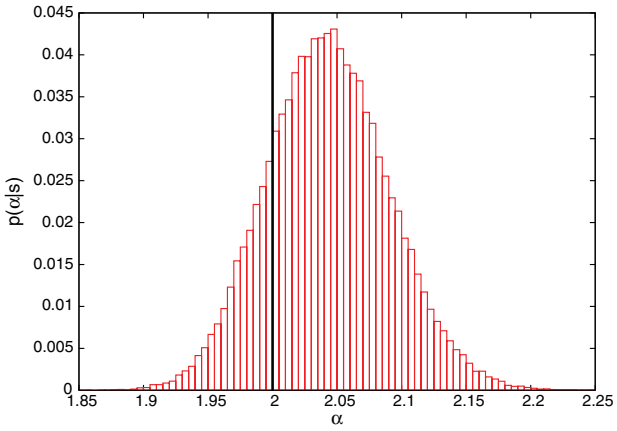

FIG. 1 (color online). The marginalized posterior distribution function for α. The injected value is indicated by the vertical black line.

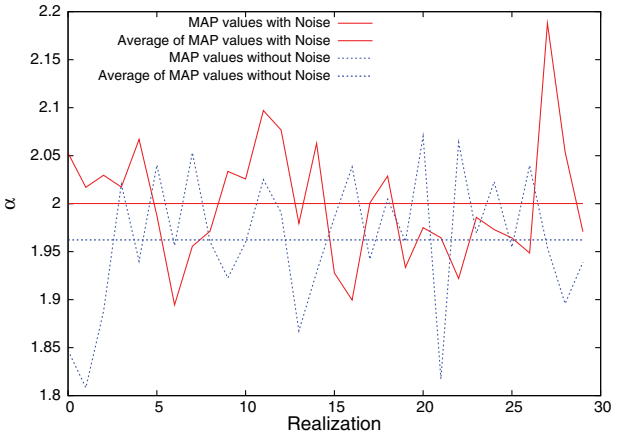

FIG. 2 (color online). MAP values for 30 different simulations of the toy model. The red curve includes noise in the simulated signal and converges to α_0 as expected. The blue curves does not include noise in the simulation and converges to $\alpha_0^2 - \beta^2$.

To test that there is no bias in the recovery of the model hyperparameter α, we produced 30 different realizations of the data and computed the average MAP value. Figure 2 shows the MAP value for each of these realizations and the corresponding average. We see that as we average over multiple realizations α does indeed converge to the injected value. The blue line in Fig. 2 shows a biased recovery for α when noise is not included in the data. We instead recover $\alpha = \sqrt{\alpha_0^2 - \beta^2} \approx 1.96$.

IV. TOY MODEL II

The hierarchical Bayesian approach produces unbiased estimates for the model parameters if the signal and the noise (and hence the likelihood) are correctly modeled. However, in some situations the cost of computing the likelihood can be prohibitive, and it becomes desirable to use approximations to the likelihood, such as the Fisher

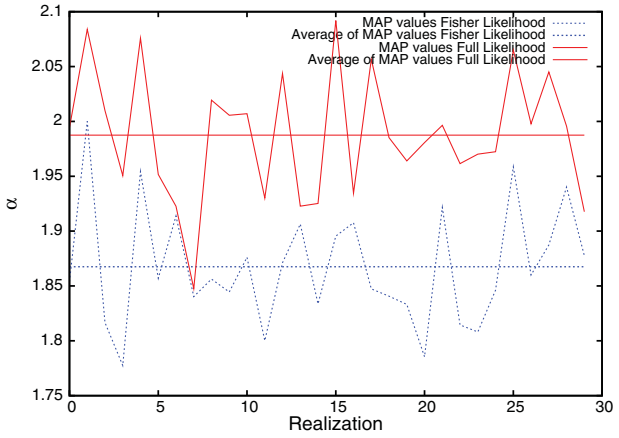

FIG. 3 (color online). MAP values for 30 different realizations of the toy model II. Using the full likelihood (red, solid) the MAP values converge to the injected value, but with the Fisher matrix approximation to the likelihood (blue, dotted) there is a bias.

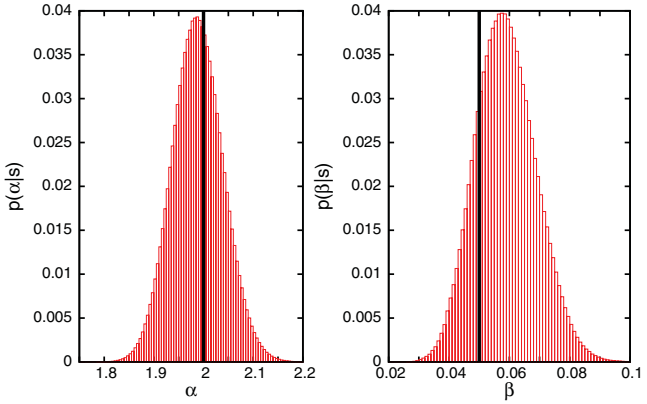

FIG. 4 (color online). PDFs for the prior hyperparameter α and the noise level β for toy model II. Both are individually constrained in this model. The injected values are shown by the bold vertical lines.

information matrix. For example, to investigate how the design of a detector influences its ability to discriminate between different astrophysical models, it is necessary to Monte Carlo the analysis over many realizations of the source population for many different instrument designs, which can be very costly using the full likelihood.

To explore these issues we introduce a new toy model that more closely resembles the likelihood functions encountered in gravitational wave data analysis. Consider a wave form h_0 that represents a single data point (e.g., the amplitude of a wavelet or a Fourier component), which can be parametrized in terms of the distance to the source d_0. The instrument noise n is assumed to be Gaussian with variance β^2. Here we will treat the noise level β as a hyperparameter to be determined from the observations. Adopting a fiducial noise level β_0 allows us to define a reference signal-to-noise ratio $\mathrm{SNR}_0^2 = h_0^2/\beta_0^2$. The likelihood of observing data $s = h_0 + n$ for a source at distance d with noise level β is then

$$p(s|d, \beta) = \frac{1}{\sqrt{2\pi}\beta} e^{-(s-h)^2/(2\beta^2)}, \tag{9}$$

where $h = (d_0/d)h_0$. The likelihood is normally distributed in the inverse distance $1/d$, with a maximum that depends on the particular noise realization n:

$$\frac{1}{d_{\mathrm{ML}}} = \frac{1 + n/(\beta_0 \mathrm{SNR}_0)}{d_0}. \tag{10}$$

Now suppose that the distances follow a one-sided normal distribution $p(d \geq 0) = \frac{2}{\sqrt{2\pi}\beta} \exp(-d^2/2\alpha_0^2)$, and that we adopt a corresponding model for the distance distribution with hyperparameter α and a flat hyperprior.

We simulate the data from $N = 1000$ sources with $\alpha_0 = 2$ and $\beta = 0.05$. The values of α_0 and β were chosen to give a fiducial SNR $= 5$ for $d = 2\alpha_0$. In the first of our simulations the value of β was assumed to be known, and we computed the MAP estimates of α for 30 different simulated data sets. As shown in Fig. 3, the average MAP estimate for α converges to the injected value.

In contrast to the first toy model where only the combination $\alpha^2 + \beta^2$ is constrained by the data, in this more realistic toy model both the noise level β and the model hyperparameter α are separately constrained. Figure 4 shows the marginalized posterior distribution function (PDFs) for both β and α. Tests using multiple realizations of the data show that the MAP values of α and β are unbiased estimators of the injected parameter values.

A. Approximating the likelihood

For stationary and Gaussian instrument noise the log likelihood for a signal described by parameters $\vec{\lambda}$ is given by

$$L(\vec{\lambda}) = -\frac{1}{2}(s - h(\vec{\lambda})|s - h(\vec{\lambda})), \tag{11}$$

where $(a|b)$ denotes the standard noise-weighted inner product, and we have suppressed terms that depend on the noise hyperparameters. We can expand the wave form $h(\vec{\lambda})$ about the injected source parameters $\vec{\lambda}_0$:

$$h(\vec{\lambda}) = h(\vec{\lambda}_0) + \Delta\lambda^i \bar{h}_{,i} + \Delta\lambda^i \Delta\lambda^j \bar{h}_{,ij} + \mathcal{O}(\Delta\lambda^3), \tag{12}$$

where $\Delta\vec{\lambda} = \vec{\lambda} - \vec{\lambda}_0$, and it is understood that the derivatives are evaluated at $\vec{\lambda}_0$. Expanding the log likelihood we find

$$L(\Delta\vec{\lambda}) = -\frac{1}{2}(n|n) + \Delta\lambda^i(n|h_{,i}) - \frac{1}{2}\Delta\lambda^i\Delta\lambda^j(h_{,i}|h_{,j}) + \mathcal{O}(\Delta\lambda^3). \tag{13}$$

The maximum likelihood solution is found from $\partial L/\partial\Delta\lambda^i = 0$, which yields $\Delta\lambda^i_{\mathrm{ML}} = (n|h_j)\Gamma^{ij}$, where Γ^{ij} is the inverse of the Fisher information matrix $\Gamma_{ij} = (h_{,i}|h_{,j})$.

Using this solution to eliminate $(n|h_{,i})$ from Eq. (13) yields the quadratic, Fisher information matrix approximation to the likelihood

$$L(\vec{\lambda}) = \text{const} - \frac{1}{2}(\lambda^i - \lambda^i_{\rm ML})(\lambda^j - \lambda^j_{\rm ML})\Gamma_{ij}. \quad (14)$$

This form of the likelihood can be used in simulations by drawing the $\Delta\lambda^i_{\rm ML}$ from a multivariate normal distribution with covariance matrix Γ^{ij}.

In our toy model $\Gamma_{dd} = \text{SNR}_0^2\beta_0^2/(\beta^2 d_0^2)$, and $L(d) = -\text{SNR}_0^2\beta_0^2(d - d_{\rm ML})^2/(2\beta^2 d_0^2)$. The approximate likelihood follows a normal distribution in d while the full likelihood follows a normal distribution in $1/d$. For signals with large SNR this makes little difference, but at low SNR the difference becomes significant and results in a bias in the recovery of the model hyperparameters, as shown in Fig. 3. In this instance there is a simple remedy: using $u = 1/d$ in place of d in the quadratic approximation to the likelihood exactly reproduces the full likelihood in this simple toy model. However, it is not always so easy to correct the deficiencies of the quadratic, Fisher information matrix approximation to the likelihood.

V. WHITE DWARF BINARIES IN THE MILKY WAY

To illustrate how the hierarchical Bayesian approach can be applied to an astrophysically relevant problem, we investigate how population models for the distribution of white dwarf binaries in the Milky Way galaxy can be constrained by data from a space-based gravitational wave detector. Several studies have looked at parameter estimation for individual white dwarf binaries in the Milky Way [35–37]. We extend these studies to consider how the individual observations can be combined to infer the spatial and mass distributions of white dwarf binaries in the Galaxy.

We use the Laser Interferometer Space Antenna (LISA) [38] as our reference mission. We focus this analysis on short-period galactic binaries, with gravitational wave frequencies above 4 mHz. Our conclusions would be little changed if we considered the recently proposed eLISA [11] mission instead, as both are able to detect roughly the same number of galactic binaries in the frequency bands considered here. For the population model we are considering, which was also used in the mock-LISA data challenges [4], there are total of $\sim 40,000$ detectable binaries for the original LISA mission assuming a 4 year lifetime, and $\sim 4,500$ for the eLISA mission assuming a 2 year lifetime. However, for the frequencies above 4 mHz that we focus on here, the detections numbers are comparable: $\sim$5000 for LISA and $\sim$2500 for eLISA.

The 4 mHz lower limit is chosen to simplify the analysis in two ways. First, it avoids the signal overlap and source confusion problems that become significant at lower frequencies [21], and second, it circumvents the issue of sample completeness and Malmquist selection bias since

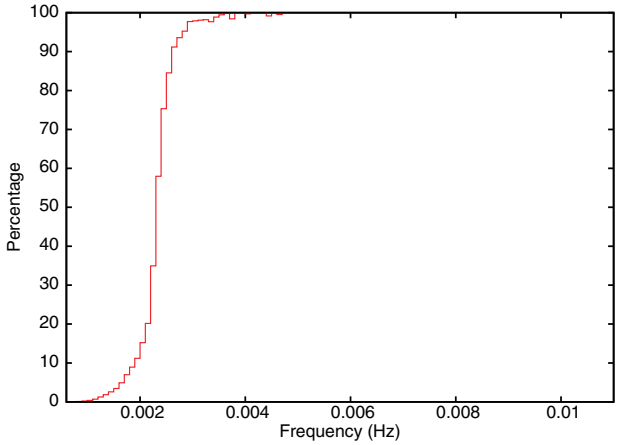

FIG. 5 (color online). The percentage of sources which are detectable as a function of frequency. Virtually 100% of the white dwarf binaries in the Milky Way above 4 mHz would be detected by LISA.

LISA's coverage of the galaxy is complete at high frequencies. This claim is substantiated in Fig. 5 showing the cumulative percentage of binaries detected as a function of frequency for a 4 year LISA mission. A given frequency bin represents the percentage of binaries of that frequency and higher that are detected. All binaries above $\sim$4 mHz are detectable by LISA, of which there are $\sim$5000.

It is possible to extend our analysis to include all detectable white dwarf binaries if we were to properly account for the undetectable sources. One way to do this is to convolve the astrophysical model priors by a function that accounts for the selection effects [20] so that we are working with the predicted observed distribution rather than the theoretical distribution. Another approach is to marginalize over the undetectable signals [19].

The high frequency signals are not only the simplest to analyze, but they also tend to have the highest signal-to-noise ratios, the best sky localization and the best mass and distance determination due to their more pronounced evolution in frequency. When simulating the population of detectable sources we will assume that binaries of all frequencies above 4 mHz are homogeneously distributed throughout the Galaxy and share the same chirp mass distribution. In reality the population is likely to be more heterogenous, and more complicated population models will have to be used.

A. Likelihood

The likelihood for a single source is given by

$$p(s|\vec{\lambda}) = Ce^{-(s-h(\vec{\lambda})|s-h(\vec{\lambda}))/2}. \quad (15)$$

Here $p(s|\vec{\lambda})$ is the likelihood that the residual $s - h(\vec{\lambda})$ is drawn from Gaussian noise, where s is the data, and $h(\vec{\lambda})$ is the signal produced in the detector by a source described by parameters $\vec{\lambda}$. The simulated data $s = h(\vec{\lambda}_0) + n$

includes a wave form $h(\vec{\lambda}_0)$ and a realization of the LISA instrument noise n. The normalization constant C depends on the instrument noise levels, but is independent of the wave form parameters.

The wave form for a white dwarf binary is well approximated by

$$h_+(t) = \frac{1}{d}\frac{4G\mathcal{M}\Omega^2}{c^4}\left(\frac{1+\cos^2\iota}{2}\right)\cos(\Omega t)$$
$$h_\times(t) = \frac{1}{d}\frac{4G\mathcal{M}\Omega^2}{c^4}\cos\iota\sin(\Omega t), \quad (16)$$

where $\Omega = 2\pi f$. We have 8 parameters that describe a white dwarf binary signal, the frequency f, the distance to the source d, the chirp mass $\mathcal{M}$, the inclination angle ι, a polarization angle ψ, a phase angle φ_0, and sky location parameters θ and ϕ. To leading order, the frequency evolves as

$$\dot{f} = \frac{96\pi}{5}(\pi\mathcal{M})^{5/3}f^{11/3}. \quad (17)$$

Sources with $\dot{f}T^2\mathrm{SNR} \sim 1$, where T is the observation time, provided useful measurements of the chirp mass $\mathcal{M}$ and the distance d [39,40]. The strong f dependence in Eq. (17) is the reason why high frequency binaries are the best candidates for placing strong constraints on the distance and chirp mass.

B. Model for the galaxy

We adopt a bulge plus disk model for the galaxy shape [41–44]. Choosing the x-y plane as the plane of the galaxy, the density of stars in the galaxy is given by

$$\rho(x,y,z) = \rho_0(Ae^{-r^2/R_b^2} + (1-A)e^{-u/R_d}\mathrm{sech}^2(z/Z_d)). \quad (18)$$

Here, $r^2 = x^2 + y^2 + z^2$, $u^2 = x^2 + y^2$, R_b is the characteristic radius for the bulge, and R_d and Z_d are a characteristic radius and height for the disk, respectively. The quantity ρ_0 is a reference density of stars, and the coefficient A, which ranges between 0 and 1, weights the number of stars in the bulge versus the number in the disk. We produced synthetic galaxies using the catalog of binaries provided by Gijs Nelemans for the mock-LISA data challenges [4,45].

With appropriate normalization, the spatial density ρ becomes our prior distribution for the spatial distribution of galactic binaries. The parameters of the density distribution A, R_b, R_d and Z_d become hyperparameters in the hierarchical Bayesian analysis. Each set of values for the four parameters corresponds to a distinct model for the shape of the galaxy. For our simulations, we chose a galaxy with $A = 0.25$, $R_b = 500$ pc, $R_d = 2500$ pc and $Z_d = 200$ pc.

C. Chirp mass prior

The distribution of sources as a function of frequency and chirp mass is not chosen by hand, but rather results from a complicated population synthesis model that takes into account many physical processes that affect binary formation and evolution [41–44]. We use the chirp mass distribution from a Monte Carlo realization of the population synthesis model to develop a parametrized analytic fit to the distribution, which we then use as a parametrized prior in our analysis.

Figure 6 shows the chirp mass distribution for binaries in our simulated galaxy. We use this distribution to construct a hyperprior on the chirp mass, approximated by the following distribution:

$$p(\mathcal{M}) = \frac{C}{(\frac{\mathcal{M}}{\mathcal{M}_0})^{-a} + \frac{a}{b}(\frac{\mathcal{M}}{\mathcal{M}_0})^{b}}, \quad (19)$$

where $\mathcal{M}_0$, a and b are hyperparameters in our model. C is the normalization constant which can be calculated analytically and is given by

$$C = \mathcal{M}_0\pi\frac{b^{\frac{a+1}{a+b}}a^{-\frac{a+1}{a+b}}}{(a+b)\sin\frac{\pi(b-1)}{a+b}}. \quad (20)$$

$\mathcal{M}_0$ is the mode of the distribution. The hyperparameters a and b determine the width of the distribution, which can be seen by calculating the full width at half maximum (FWHM). It is given by

$$\mathrm{FWHM} \simeq \mathcal{M}_0([2(b/a+1)]^{1/b} - [2(a/b+1)]^{-1/a}). \quad (21)$$

We further assume that the orbital evolution is due only to the emission of gravitational waves, and is thus adequately described by Eq. (17). In principle, one would want to be more careful and consider tidal effects and mass transfer [46] as possible contributions to $\dot{f}$. However, it is expected

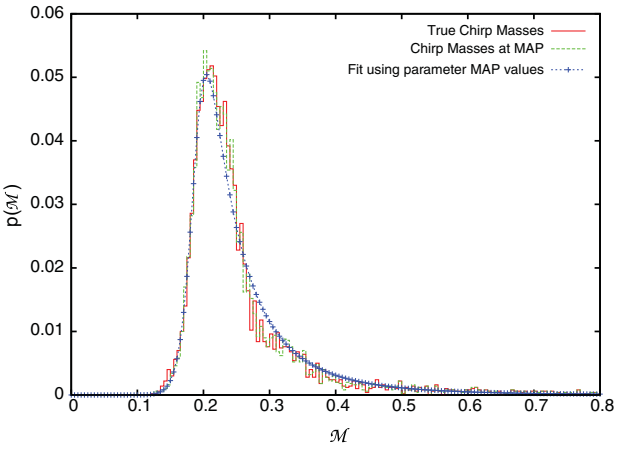

FIG. 6 (color online). The chirp mass distribution of the 5000 binaries used in our simulations is shown in red (solid). The green (dashed) distribution shows the MAP values of the recovered chirp mass for each binary, and the blue (dashed with crosses) shows the model (19) using the MAP values for the chirp mass prior hyperparameters. The brightest binaries accurately capture the chirp mass distribution, which serves as a useful prior for sources whose chirp masses are not so well determined.

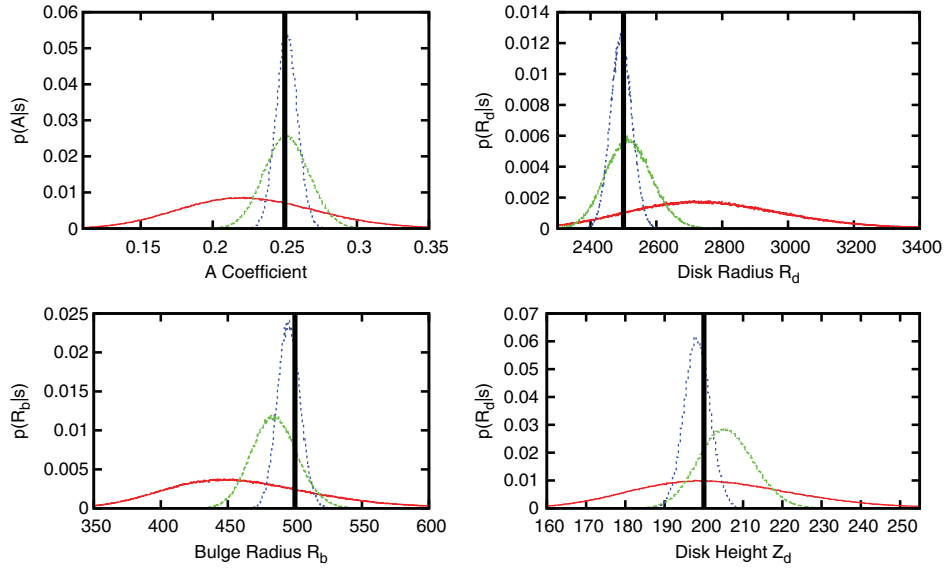

FIG. 7 (color online). PDFs for the four galaxy model hyperparameters. The red (solid) is for a simulation using 100 binaries, the green (dotted) 1000 binaries and the blue (dashed) 5000 binaries. The black lines (bold vertical lines) show the true values of the distribution from which the binaries were drawn.

that the high frequency sources we are focusing on will be mostly detached white dwarf binaries where tidal or mass transfer effects are unlikely to be significant [47].

Our ability to measure the hyperparameters of the spatial distribution depends on how well we measure the sky location and distance for each binary. For many sources, the distance is poorly determined because it is highly correlated with the chirp mass. For slowly evolving systems, the distance determination is significantly improved by having a well determined chirp mass prior. If effect, systems with sufficiently high frequency, chirp mass and/or SNR provide tight constraints on the chip mass distribution, which in turn improves the distance determination for lower SNR, less massive or lower frequency sources.

VI. RESULTS

We are able to efficiently calculate the full likelihood for each source [Eq. (15)] using the fast wave form generator developed by Cornish and Littenberg [32]. The following results are all derived from simulations using the full likelihood. Using the same MCMC approach from our toy models, we sample the posterior and get PDFs for source

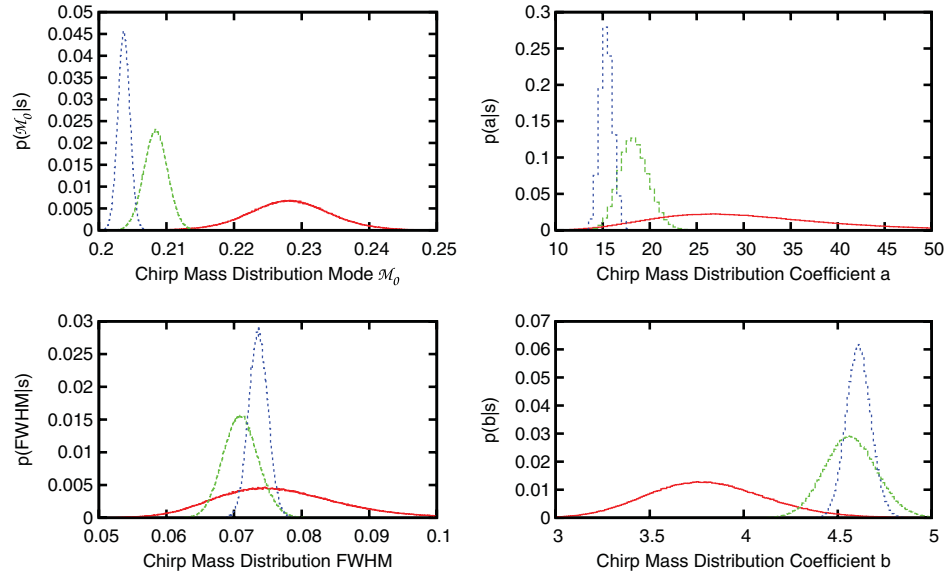

FIG. 8 (color online). PDFs for the three chirp mass model hyperparameters and the FWHM of the distribution. The red (solid) is for a simulation using 100 binaries, the green (dotted) 1000 binaries and the blue (dashed) 5000 binaries.

TABLE I. MAP values and variances for the galaxy hyperparameters when using 100, 1000 and 5000 galactic binaries in the analysis. The simulated values were $A = 0.25$, $R_b = 500$ pc, $R_d = 2500$ pc and $Z_d = 200$ pc.

	100		1000		5000	
Parameter	MAP	σ	MAP	σ	MAP	σ
A	0.262	0.047	0.226	0.0157	0.249	0.0074
Rb (pc)	440	58.9	490	17.1	480	8.38
Rd (pc)	2465	237.5	2584	70.2	2461	32.4
Zd (pc)	193	20.8	201	7.02	195	3.25
$\mathcal{M}_0$	0.226	0.0063	0.208	0.0018	0.205	0.00088
FWHM	0.07	0.0094	0.071	0.0026	0.076	0.0014

and model parameters simultaneously. We check for convergence by starting the chains at different locations in the prior volume and find that regardless of starting location, the chains converge to the same PDFs.

Our procedure successfully recovers the correct chirp mass distribution, as shown in Fig. 6, and is able to meaningfully constrain the parameters of the galaxy distribution and chirp mass distribution models, with PDFs shown in Figs. 7 and 8, respectively.

We ran simulations with 100, 1000 and 5000 binaries to show how the constraints on the galaxy hyperparameters improved as we include more sources (for comparison, eLISA is expected to detect between 3500–4100 white dwarf binaries during a 2 year mission lifetime [11]). The chains ran for 1 million, 500 k and 100 k iterations, respectively. Even for a relatively modest number of detections we begin to get meaningful measurements on the population model of white dwarf binary systems. The more binaries we use in our analysis the tighter our constraints on the hyperparameters.

Table I lists the recovered MAP values and the variance of the marginalized posterior distribution function for each hyperparameter. Gravitational wave observations would be very competitive with existing electromagnetic observations in constraining the shape of the galaxy [48,49]. Making direct comparisons between our results to those in the literature is complicated, as the actual values of the bulge and disk radii are very model dependent. For example, Juric uses a model where the galaxy is comprised of both a thin and thick disk. With gravitational wave data in hand, this comparison could easily be made by trivially substituting the density profile used here.

What matters for this proof-of-principle study is how well the parameters can be constrained. In the models of Juric *et al.* constraints for the disk radii are around 20%. We find similar accuracy when using a pessimistic population of 100 systems. Adopting a source catalog that is more consistent with theoretical predictions, we find constraints for the disk parameters as low as 1.5%—a substantial improvement over the current state-of-the-art electromagnetic constraints.

A. Approximating the likelihood

While we happen to have a very efficient method for computing the full likelihood for galactic binaries, this is not always the case. For other signal types, such as extreme mass ratio inspirals [50], the full likelihood can be very expensive to compute, posing problems if we wish to do extensive studies of many astrophysical models or detector configurations. For such exploratory studies it is preferable to use the Fisher information matrix, which provides a good approximation to the likelihood (14) for signals with high SNR [51] However, as we saw with the toy model in Sec. IV,

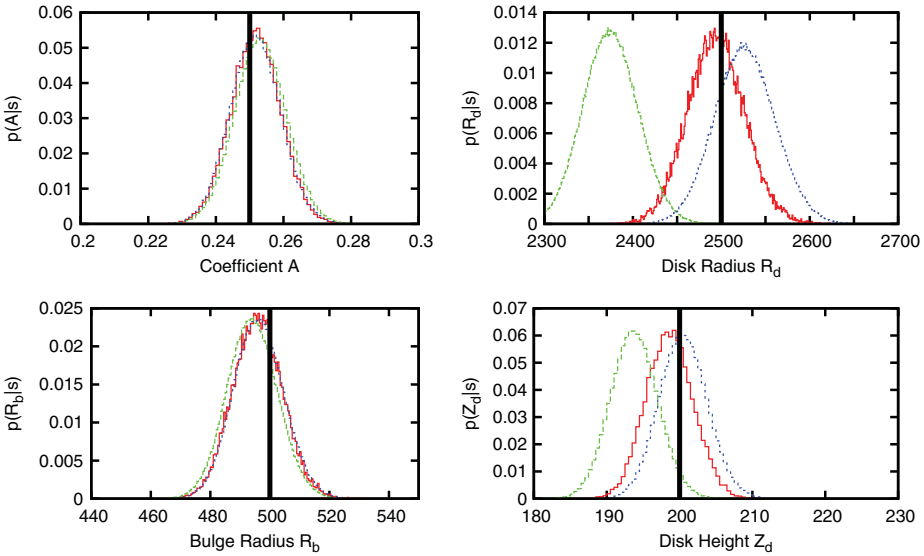

FIG. 9 (color online). PDFs from a simulation using 5000 binaries for the four galaxy model hyperparameters for the full likelihood (red, solid), a Fisher approximation in d (green, dotted) and a Fisher approximation in $1/d$ (blue, dashed).

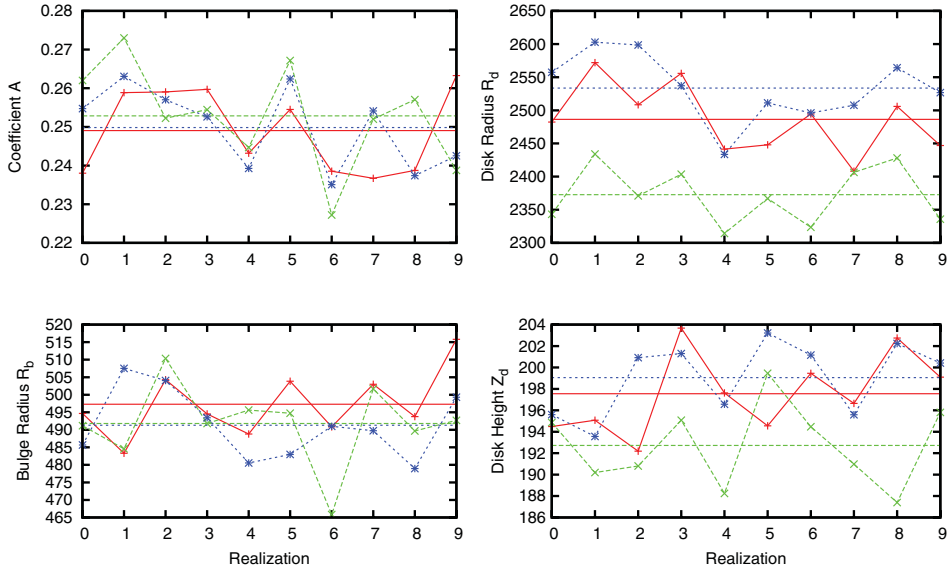

FIG. 10 (color online). MAP values and corresponding averages from a simulations using 5000 binaries for the four galaxy model hyperparameters for the full likelihood (red, solid with plus symbols), a Fisher Matrix approximation parametrized with d (green, dashed with times symbols) and a Fisher Matrix approximation using $1/d$ (blue, dashed with star symbols).

this can lead to biases in the recovered parameters. The Fisher matrix Γ_{ij} is not a coordinate invariant quantity, and we can at least partially correct the bias by reparametrizing our likelihood. Just as in Sec. IV, instead of using the distance d as a variable, we can instead use $1/d$, which provides a much better approximation to the full likelihood. We test these shortcuts by redoing the analysis of the galactic population using the Fisher matrix approximation to the likelihood (both with d and $1/d$ as parameters) and compare to the results from the previous analysis using the full likelihood. Figure 9 shows PDFs for the galaxy hyperparameters using the three different methods for computing $p(d|\vec{\lambda})$ with the full sample of 5000 binaries.

We find that the approximation using $1/d$ matches the full likelihood better than the likelihood parametrized with d; however, there are additional discrepancies due to nonquadratic terms in the sky location $\{\theta, \phi\}$ that we have not accounted for. The dependence of the wave form on $\{\theta, \phi\}$ is more complicated than the distance, and is not so easily corrected by a simple reparametrization. The approximation could be improved by carrying the expansion of the likelihood beyond second order; however, this is computationally expensive and can be numerically unstable (though not always [52]).

If we analyze several realizations of the galaxy using the three different likelihood functions and average the results, we find the biases are persistent for the approximate methods. Figure 10 shows the MAP values and the average of the MAP values for 10 realizations of our fiducial galaxy model. The biases in the recovered disk radius and disk height are particularly pronounced when using the Fisher matrix approximation to the likelihood parametrized with d.

VII. CONCLUSION

We have demonstrated a general hierarchical Bayesian method capable of constraining the model parameters for a population of sources. In the particular case of white dwarf binaries in the Milky Way, we can constrain the spatial distribution of the galaxy to levels better than current electromagnetic observations using the anticipated number of systems detectable by space-based gravitational wave detectors. Even if the currently held event rates for white dwarf binaries turn out to be optimistic by more than an order of magnitude, the constraints possible with a gravitational wave detector are comparable to our current estimates of the Milky Way's shape.

When the data from a space-borne detector has been collected, the resolvable white dwarf binaries will be regressed from the data, leaving behind a confusion-limited foreground which will significantly contribute to the overall power in the data around $\sim$1 mHz. Measuring the overall shape of the galaxy as demonstrated here will provide additional means to characterize the level of the confusion noise. As we will show in an upcoming paper, we can then use the detailed understanding of the foreground signal to detect a stochastic gravitational wave background at levels well below the confusion noise.

Analyzing simulated data with the full likelihood is computationally taxing and, when performing a large suite of such studies, could prove to be prohibitive. To mitigate the cost of such analyses, we test a much faster approach (approximately 50 times faster), using the Fisher matrix approximation to the likelihood. We find the results are significantly less biased by the Fisher approximation when using $1/d$ as the parameter that encodes the distance to the

source. This simple adjustment gives adequately reliable results in significantly less time than the brute-force calculation, and will provide an additional, useful, metric to gauge the relative merits of proposed space-based gravitational wave missions.

ACKNOWLEDGMENTS

N. J. C. and M. A. were supported by NASA Grant No. NNX07AJ61G. T. B. L. was supported by NASA Grant No. 08-ATFP08-0126.

[1] D. Hils, P. Bender, and R. Webbink, Astrophys. J. **360**, 75 (1990).
[2] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, Astrophys. J. **623**, 23 (2005).
[3] J. R. Gair, L. Barack, T. Creighton, C. Cutler, S. L. Larson, E. S. Phinney, and M. Vallisneri, Classical Quantum Gravity **21**, S1595 (2004).
[4] S. Babak *et al.* (Mock LISA Data Challenge Task Force), Classical Quantum Gravity **27**, 084009 (2010).
[5] L. Barack and C. Cutler, Phys. Rev. D **69**, 082005 (2004).
[6] N. J. Cornish and J. Crowder, Phys. Rev. D **72**, 043005 (2005).
[7] R. T. Stebbins, Classical Quantum Gravity **26**, 094014 (2009).
[8] J. R. Gair, C. Tang, and M. Volonteri, Phys. Rev. D **81**, 104014 (2010).
[9] J. R. Gair, A. Sesana, E. Berti, and M. Volonteri, Classical Quantum Gravity **28**, 094018 (2011).
[10] A. Sesana, J. Gair, E. Berti, and M. Volonteri, Phys. Rev. D **83**, 044036 (2011).
[11] P. Amaro-Seoane, S. Aoudia, S. Babak, P. Binetruy, E. Berti *et al.*, Classical Quantum Gravity **29**, 124016 (2012).
[12] J. E. Plowman, D. C. Jacobs, R. W. Hellings, S. L. Larson, and S. Tsuruta, arXiv:0903.2059.
[13] J. E. Plowman, R. W. Hellings, and S. Tsuruta, arXiv:1009.0765.
[14] I. Mandel, Phys. Rev. D **81**, 084029 (2010).
[15] R. O'Shaughnessy, arXiv:1204.3117.
[16] K. S. Mandel, W. M. Wood-Vasey, A. S. Friedman, and R. P. Kirshner, Astrophys. J. **704**, 629 (2009).
[17] K. Soiaporn, D. Chernoff, T. Loredo, D. Ruppert, and I. Wasserman, arXiv:1206.3540.
[18] T. J. Loredo, arXiv:1208.3036.
[19] C. Messenger and J. Veitch, arXiv:1206.3461.
[20] B. F. Schutz, Classical Quantum Gravity **28**, 125023 (2011).
[21] J. Crowder and N. Cornish, Phys. Rev. D **75**, 043008 (2007).
[22] S. Nissanke, M. Vallisneri, G. Nelemans, and T. A. Prince, Astrophys. J. **758**, 131 (2012).
[23] S. E. Timpano, L. J. Rubbo, and N. J. Cornish, Phys. Rev. D **73**, 122001 (2006).
[24] T. B. Littenberg, Phys. Rev. D **84**, 063009 (2011).
[25] M. R. Adams and N. J. Cornish, Phys. Rev. D **82**, 022002 (2010).
[26] I. J. Good, *The Estimation of Probabilities: An Essay on Modern Bayesian Methods* (MIT Press, Cambridge, Massachussetts, 1965).
[27] D. V. Lindley and A. F. M. Smith, J. R. Stat. Soc. Ser. B **34**, 1 (1972).
[28] C. N. Morris and S. L. Normand, *Bayesian Statistics 4*, edited by A. P. D. J. M. Bernardo, J. O. Berger, and A. F. M. Smith (Oxford University Press, Oxford, 1992), p. 321.
[29] D. J. C. MacKay, Neural Comput. **11**, 1035 (1999).
[30] G. Casella, The American Statistician **39**, 83 (1985).
[31] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis* (Chapman and Hall, London, 2000), 2nd ed.
[32] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007).
[33] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **82**, 103007 (2010).
[34] P. J. Green, N. L. Hjort, and S. Richardson, *Highly Structured Stochastic Systems* (Oxford University Press, Oxford, 2003).
[35] N. J. Cornish and S. L. Larson, Phys. Rev. D **67**, 103001 (2003).
[36] A. Vecchio and E. D. Wickham, Phys. Rev. D **70**, 082002 (2004).
[37] A. Błaut, S. Babak, and A. Królak, Phys. Rev. D **81**, 063008 (2010).
[38] R. Stebbins *et al.*, Laser interferometer space antenna (lisa) a response to the astro2010 rfi for the particle astrophysics and gravitation panel, White paper, NASA (2009), available online, http://lisa.nasa.gov/documentation.html.
[39] B. F. Schutz, Nature (London) **323**, 310 (1986).
[40] R. Takahashi and N. Seto, Astrophys. J. **575**, 1030 (2002).
[41] G. Nelemans, L. Yungelson, and S. P. Zwart, Mon. Not. R. Astron. Soc. **349**, 181 (2004).
[42] G. Nelemans, L. R. Yungelson, S. F. P. Zwart, and F. Verbunt, Astron. Astrophys. **365**, 491 (2001).
[43] G. Nelemans, S. F. Portegies Zwart, F. Verbunt, and L. Yungelson, Astron. Astrophys. **368**, 939 (2001).
[44] G. Nelemans, L. Yungelson, and S. F. P. Zwart, Astron. Astrophys. **375**, 890 (2001).
[45] K. Arnaud, S. Babak, J. Baker, M. Benacquista, N. Cornish *et al.*, Classical Quantum Gravity **24**, S551 (2007).
[46] A. Stroeer and G. Nelemans, Mon. Not. R. Astron. Soc. Lett. **400**, L24 (2009).
[47] B. Willems, C. Deloye, and V. Kalogera, Astrophys. J. **713**, 239 (2010).
[48] P. J. McMillan and J. J. Binney, arXiv:0907.4685.
[49] M. Juric *et al.* (SDSS Collaboration), Astrophys. J. **673**, 864 (2008).
[50] N. Yunes, A. Buonanno, S. A. Hughes, Y. Pan, E. Barausse, M. Miller, and W. Throwe, Phys. Rev. D **83**, 044044 (2011).
[51] M. Vallisneri, Phys. Rev. Lett. **107**, 191104 (2011).
[52] S. Vitale and M. Zanolin, Phys. Rev. D **82**, 124065 (2010).