Heterodyned likelihood for rapid gravitational wave parameter inference

Neil J. Cornish

eXtreme Gravity Institute, Department of Physics, Montana State University, Bozeman, Montana 59717, USA

(Received 3 September 2021; accepted 5 November 2021; published 22 November 2021)

Inferring the source properties of a gravitational wave signal has traditionally been very computationally intensive and time consuming. In recent years, several techniques have been developed that can significantly reduce the computational cost while delivering rapid and accurate parameter inference. One of the most powerful of these techniques is the heterodyned likelihood, which uses a reference waveform to base band the likelihood calculation. Here, an efficient implementation of the heterodyned likelihood is presented that can be used for a wide range of signal types and for both ground-based and space-based interferometers. The computational savings relative to direct calculation of the likelihood vary between two and four orders of magnitude depending on the system. The savings are greatest for low mass systems such as neutron star binaries. The heterodyning procedure can incorporate marginalization over calibration uncertainties and the noise power spectrum.

DOI: 10.1103/PhysRevD.104.104054

I. INTRODUCTION

Full parameter inference [1,2] for gravitational wave signals can be very computationally intensive, with runs on single systems taking days or weeks using contemporary hardware. This inefficiency has motivated the development of novel approaches to speed up the process. Some of these methods work by speeding up the calculation of waveform templates using singular value decomposition [3,4] or reduced-order modeling [5,6]. Other methods speed up the likelihood evaluation using techniques such as heterodyning [7–10], waveform decomposition and precomputation [8,9,11,12], reduced-order quadrature [13,14], and variable frequency banding [15]. Hardware-based acceleration has also been investigated [16–18], as has machine learning [19–21].

The heterodyning approach, first introduced in 2010 [7], and later rebranded under the unfortunate term "relative binning" [10], has yet to be widely adopted, despite it being widely applicable, easy to implement, and incredibly fast. The goal of this paper is to present an efficient implementation of the heterodyned likelihood using discrete Legendre polynomial expansions on adaptively spaced frequency grids. The speed and accuracy of the heterodyning approach is demonstrated, along with a discussion of how it can be used when marginalizing over calibration and noise models.

II. THE HETERODYNED LIKELIHOOD

The idea behind the heterodyned likelihood is very simple [7]. To match the signal well enough to give a decent likelihood, the phase and amplitude evolution of the waveform template has to closely match that of the signal. Thus, if $\bar{h}(f) = \bar{\mathcal{A}}(f)e^{i\bar{\Phi}(f)}$ is a reference template with high likelihood and $h(f) = \mathcal{A}(f)e^{i\Phi(f)}$ is another template with high likelihood, the ratio $\zeta(f) = h(f)/\bar{h}(f)$ will be a slowly varying function. For signals made up of multiple harmonics, the same reason applies harmonic by harmonic. The terms that appear in the log likelihood,

$$\ln L = (d|h) - \frac{1}{2}(h|h), \tag{1}$$

where d is the data and $(a|b)$ denotes the noise weighted inner product, can be factored into slowly varying and rapidly varying components,

$$\begin{aligned}(d|h) &= 2\int \frac{d(f)h^*(f) + d^*(f)h(f)}{S(f)}\,df \\ &= 2\int (\kappa(f)\zeta^*(f) + \kappa^*(f)\zeta(f))df,\end{aligned} \tag{2}$$

where $\kappa(f) = d(f)\bar{h}^*(f)/S(f)$ is rapidly varying, and $\zeta(f)$ is slowly varying. Similarly,

$$\begin{aligned}(h|h) &= 2\int \frac{h(f)h^*(f) + h^*(f)h(f)}{S(f)}\,df \\ &= 4\int df|\zeta(f)|^2\sigma(f)df,\end{aligned} \tag{3}$$

where $\sigma^2(f) = |\bar{h}(f)|^2/S(f)$ is rapidly varying, and $|\zeta(f)|^2$ is slowly varying [the rapid variation of $\sigma(f)$ is due to spectral lines]. The heterodyning procedure uses a

low-order polynomial expansion of the slow terms to dramatically decrease the computational cost of computing the likelihood. The accuracy of the approximation can be improved by writing the likelihood as

$$\ln L = \ln \bar{L} - (\bar{r}|\Delta h) - \frac{1}{2}(\Delta h|\Delta h), \tag{4}$$

where $\ln \bar{L}$ is the likelihood computed using the reference waveform, $\bar{r} = d - \bar{h}$ is the reference residual, and $\Delta h = \bar{h} - h$. Since the terms involving Δh are generally small, the fractional error introduced by approximating the integrals results in a smaller absolute error than approximating the full integrals $(d|h)$ and $(h|h)$. The slow varying term is now $\Delta\zeta(f) = \Delta h(f)/\bar{h}(f) = 1 - \zeta(f)$.

A. Legendre expansion

In practice, the integrals become sums over frequency that can be efficiently computed using an expansion in discrete Legendre polynomials P_{nk}, where n denotes the polynomial order and $k = 0, 1, \ldots, N$ denotes the frequency bin. The P_{nk} satisfy the orthogonality condition,

$$P_{nk}P_{mk} = \delta_{nm}\alpha_{nN}, \tag{5}$$

where α_{nN} is a normalization factor [22], and we have used the Einstein summation convention. The first few P_{nk} have the forms

$$\begin{aligned} P_{0k} &= 1, \\ P_{1k} &= 1 - \frac{2k}{N}, \\ P_{2k} &= 1 - \frac{6k}{N} + \frac{6k(k-1)}{N(N-1)}. \end{aligned} \tag{6}$$

Note that the shapes of the polynomials depends on the number of frequency bins, $N + 1$. The higher-order polynomials can be generated efficiently using a recursion relation [22].

A function $g_k = g(f_k)$ can be expanded,

$$g_k = \gamma_n P_{nk}, \tag{7}$$

where $n = 0, 1, ..N$, and

$$\gamma_n = \frac{P_{nk}g_k}{\alpha_{nN}}. \tag{8}$$

In this way, a term in the likelihood such as $(d|h)$ can be written as

$$(d|h) = 2\alpha_{nN}(\kappa_n^R\zeta_n^R + \kappa_n^I\zeta_n^I), \tag{9}$$

where, for example, $\{\kappa_n^R, \kappa_n^I\}$ denote the expansion coefficients of the real and imaginary parts of $\kappa(f)$. The expression (9) is exact, it is simply a rewriting of the original sum. As it stands, (9) actually represents an *increase* in the computational cost from $\mathcal{O}(N)$ to $\mathcal{O}(N^2)$ [though there are fast Chebyshev–Legendre transforms [23] that reduce the cost to $\mathcal{O}(N(\log N)^2/\log\log N)$]. The savings come by restricting the sum over n in (9) to a small number of coefficients. This can be done with little loss of accuracy since the high-order terms in the expansion of the slowly varying components diminish very quickly with increasing n.

The sum (9) can be most efficiently approximated by breaking the full sum into smaller segments. The sum can be restricted to frequencies where the reference waveform contributes significantly to the signal to noise. The efficiency is further improved by adaptively determining the width of the segments such that the slow terms can be accurately covered by polynomials of some chosen order J. The coefficients of the fast varying terms have to be computed at the full frequency resolution, but this can be done once and the results stored. Thus, the cost of the heterodyned likelihood comes down to computing the coefficients of the slow components in each frequency band. This cost scales as J^2Q, where Q are the number of bands. The number of waveform evaluations for the slow terms scales as $M = JQ$. While increasing J allows us to use wider bands for a given error tolerance, the increase in bandwidth scales slower than the polynomial order J, making it more efficient to use low polynomial order and more bands.

We denote the slow varying terms such as $\zeta(f)$, $\Delta\zeta(f)$ generically as $s(f)$. Because $s(f)$ is slowly varying, it does not need to be sampled at each of the $N + 1$ frequency bins in a given frequency band. Instead, it can be sampled on a much sparser grid of values, $s(f_j)$, where $j \in [0, J]$ and J is the highest order used in the polynomial expansion. The discrete Legendre polynomials will not be orthogonal on this sparse subgrid, so the usual expression for the expansion coefficients (8) cannot be used. Instead, we can start with the defining relation

$$s(f_j) = s_n P_n(f_j) \equiv s_n Y_{nj}, \tag{10}$$

where the polynomial expansion is restricted to $n \in [0, J]$. The $(J + 1) \times (J + 1)$ matrix $Y_{nj} = P_n(f_j)$ contains the subsampled values of the discrete Legendre polynomials. We can then solve for the expansion coefficients,

$$s_n = Y_{nj}^{-1} s(f_j). \tag{11}$$

The frequency samples f_j do not have to be evenly spaced, but the matrix Y_{nj} can become ill conditioned if the spacing of any two bins exceeds $\sim 2N/J$. To ensure optimal accuracy, it is best to space the subsamples uniformly across each frequency band. The matrix inverses for each frequency band can be computed once and stored for later

use. If greater accuracy is desired, the number of samples in each frequency band used in the fit (10) can be increased while keeping the polynomial order fixed, with the solution for the coefficients found using a singular value decomposition rather than the simple matrix inverse (11).

B. Frequency spacing

The goal is to resolve the slow varying function $\Delta\zeta(f)$ to sufficient accuracy using the smallest number of frequency samples. The first thing to consider is how far from the reference waveform any template is likely to be. For Gaussian posterior distributions, we know that twice the log likelihood is chi-squared distributed with D degrees of freedom, where D is the number of parameters in the model. This scaling holds remarkable well even when the posterior distributions are non-Gaussian. Thus, to account for any waveforms that will contribute to the posterior distribution, we need to cover deviations from the reference likelihood of order a few times the standard deviation $\sqrt{D}$. As a proxy, we can use the chi-squared $\chi^2 = (\Delta h|\Delta h)$ as a measure of the deviation from the reference waveform, and so long as the frequency grid accurately covers departures as large as $\chi^2 \sim 20 \to 50$, the parameter estimation will be reliable.

By perturbing any one source parameter by a suitable amount, it is possible to arrive at the desired chi-squared value, but the perturbation in each parameter will lead to a different behavior for the $\Delta\zeta(f)$. One way to cover all possibilities is to consider perturbations along the eigendirections, $\vec{v}_i$, of the Fisher information matrix $\Gamma_{ij} = (\partial_{\theta^i}\bar{h}|\partial_{\theta^j}\bar{h})$, such that $\Delta\vec{\theta}_i = \alpha_i\vec{v}_i$ (no sum on the i). (Strictly speaking, Γ_{ij} should only be called the Fisher matrix if $\bar{h}$ is computed at the maximum likelihood point. More properly, Γ_{ij} is the signal space metric evaluated at the reference waveform.) In principle, setting $\alpha_i = \beta/\sqrt{\lambda_i}$, where λ_i is the eigenvalue corresponding to the $\vec{v}_i$ eigenvector, should yield $\chi^2 = \beta^2$, but the Fisher matrix is often ill conditioned, and the β's need to be iteratively adjusted to give the desired chi-square value. Figure 1 shows the real and imaginary parts of $\Delta\zeta$ for the black hole binary GW150914 [24] evaluated at the LIGO Hanford detector and using the IMRPhenomD waveform model [25]. The parameters were perturbed along each eigendirection of the Fisher matrix and scaled to give $\chi^2 = 50$. We see that functions vary most rapidly at low frequencies and near merger (at around 200–300 Hz).

The frequency spacing can then be determined by considering a linear fit to the $\Delta\zeta$ for perturbations along each eigenvector direction and in each detector. Starting with a frequency spacing of one frequency bin, the spacing is steadily incremented until the difference between $\Delta\zeta$ and the linear fit across the interval deviates by more than some specified tolerance. Figure 2 shows the frequency spacings for various error tolerances and chi-square values. We see

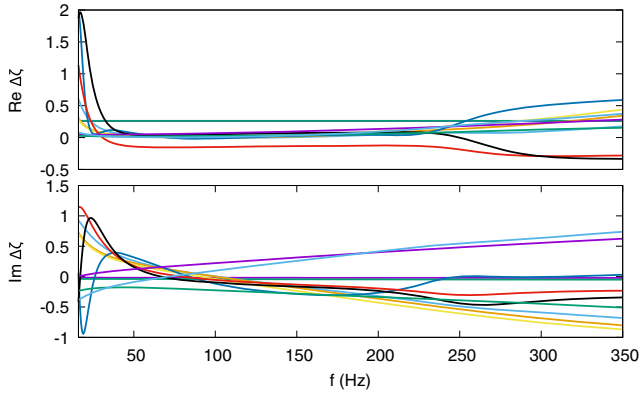

FIG. 1. Real and imaginary parts of the heterodyned waveform differences, $\Delta\zeta = (\bar{h} - h)/\bar{h}$, with h displaced along successive eigendirections of the Fisher information matrix, with the differences scaled such that $\chi^2 = (\Delta h|\Delta h) = 50$.

that the samples are concentrated at low frequencies. To keep the number of samples to a minimum, we can use a relative rather than absolute error tolerance. Since the contribution to the likelihood in each frequency band scales as $\sigma^2(f) = |\bar{h}(f)|^2/S(f)$, we can achieve a desired relative error tolerance by scaling the absolute error tolerance by $[\sigma^2]/\sigma^2(f)$ where the square brackets denotes the average value.

The frequency bandwidths for higher-order polynomial fits are derived from the reference linear fit by finding the smallest frequency spacing in a given region and multiplying that number by the polynomial order J. This results in slightly more samples than for the linear fit as the wider frequency bands are less able to accommodate to the ideal source-adaptive frequency spacing. Note that the enveloping procedure leads to slightly more samples being used even at linear ($J = 1$) order. Figure 3 compares the frequency spacing for GW150914 at linear and quadratic

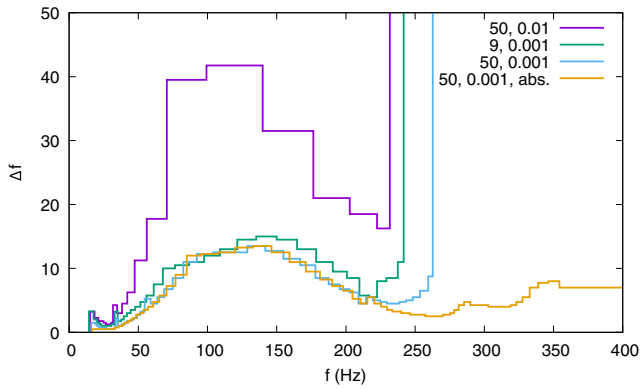

FIG. 2. Source-adaptive frequency spacing of the waveform samples for GW150914 using various choices of linear error tolerance and chi-squared offset. The spacings used a relative error tolerance, save for the one labeled "abs", which used an absolute error tolerance.

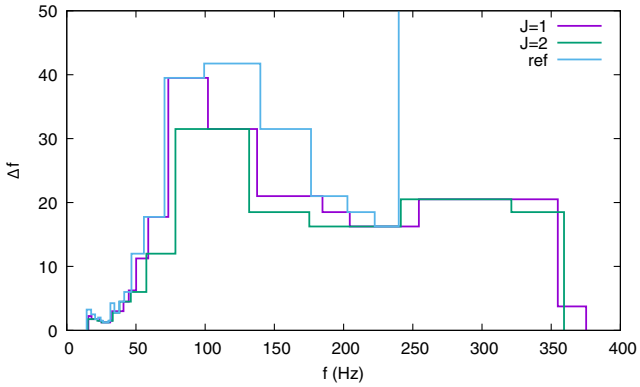

FIG. 3. Source-adaptive frequency spacing of the waveform samples for GW150914 at linear and quadratic order. The reference spacing is also shown.

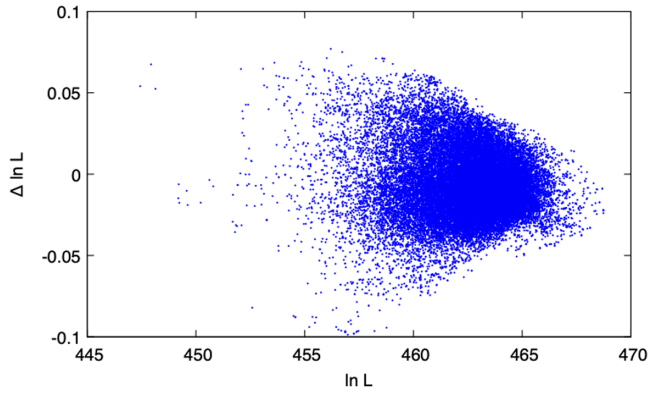

FIG. 5. Difference between the full and heterodyned likelihood for neutron star binary GW170817. In both cases, the heterodyne used second-order Legendre polynomials with a linear fit tolerance of 0.01 and a chi-squared value of $(\Delta h|\Delta h) = 50$.

polynomial order to the reference linear spacing. There are 32 samples at linear order and 34 at quadratic order, as compared to 22 for the reference spacing.

The frequency spacing adjusts to account for the characteristics of each system. Figure 4 compares the frequency grids for three systems of decreasing total mass. In each case, the minimum frequency was set at 8 Hz, and maximum frequency was set at 1024 Hz. The time span analyzed increases with decreasing mass: 4 sec for black hole binary GW150914, 8 sec for black hole binary GW151012, and 128 sec for the binary neutron star binary GW170817 [26]. The number of frequency samples in the heterodyned likelihood were 34 for GW150914, 52 for GW151012, and just 22 for GW170817. The small number of samples for the neutron star binary GW170817 relative to the two black hole binaries is due to the fact that GW170817 entered the band at 22 Hz and exited the band prior to merger. The savings in waveform evaluations using the heterodyned likelihood relative to the direct likelihood for the three systems are a factor of 120 for GW150914, 156 for GW151012, and 5900 for GW170817.

Figure 5 shows the difference, $\Delta \ln L$, between the full and heterodyned likelihood calculation and a function of the likelihood for the samples collected during a Markov chain Monte Carlo (MCMC) run on the GW170817 data. The absolute value of the error never exceeds 0.1, and the average absolute error is just 1.6×10^{-2}.

Smaller linear error tolerances result in more frequency samples in the heterodyne and a reduction in error in the likelihood. Figure 6 shows how the error in the likelihood and the number of samples in the heterodyne scale with the linear error tolerance. The error in the likelihood is measured by the average of the absolute value of the difference between the full and heterodyned likelihood, $\overline{|\Delta \ln L|}$, for accepted samples in a MCMC run. The error decreases as the tolerance is reduced from 0.1 to 0.001, but

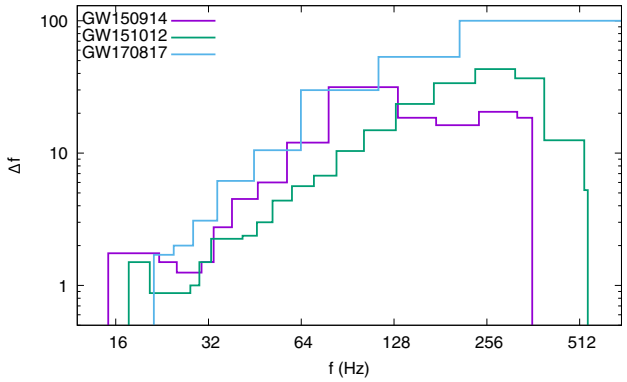

FIG. 4. Comparison of the source-adaptive frequency spacing for the black hole binaries GW150914 and GW151012 and the binary neutron star binary GW170817. The heterodyne used second-order Legendre polynomials with a linear fit tolerance of 0.01 and a chi-squared value of $(\Delta h|\Delta h) = 50$.

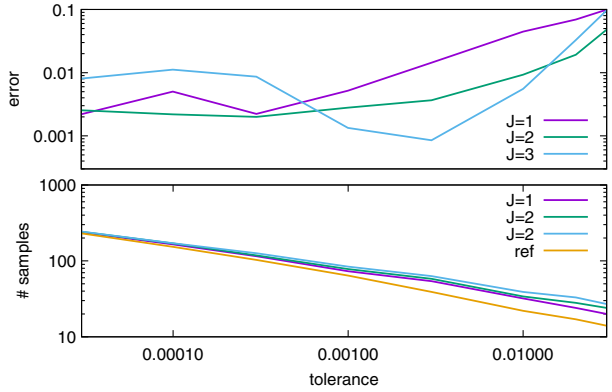

FIG. 6. The upper panel shows the average of the absolute value of the difference between the full likelihood and the heterodyned likelihood as a function of the linear error tolerance for GW150914 at linear, quadratic, and cubic orders. The lower panel shows the number of frequency sample used in the heterodyne as a function of the linear error tolerance.

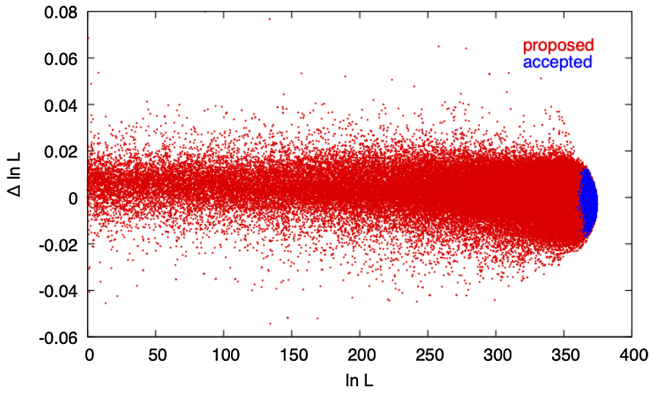

FIG. 7. Difference between the full and heterodyned likelihood for GW150914 for proposed and accepted parameter values. The heterodyne used second-order Legendre polynomials with a linear fit tolerance of 0.001 and a chi-squared value of $(\Delta h|\Delta h) = 50$.

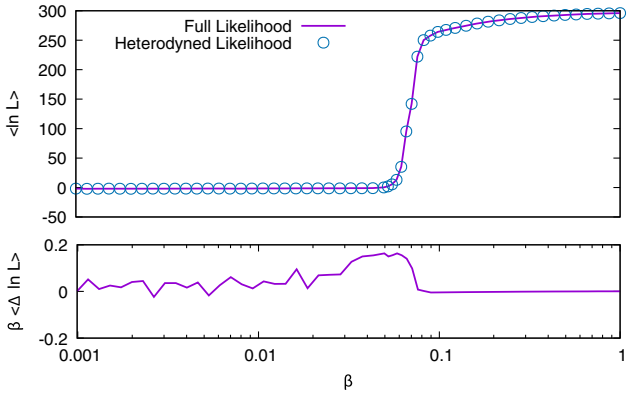

FIG. 8. The upper panel shows the average log likelihood as a function of the inverse temperature of the chains, β, for the full likelihood and the heterodyned likelihood for GW150914 using the same settings as Fig. 7. The lower panel shows the difference in the thermodynamic integrands.

then asymptotes or even increases. The reason for this behavior can be traced to the error introduced in estimating the Legendre expansion of the slow terms using (8). As the error tolerance is decreased, the number of frequency bands grows as the error tolerance to the power of ~ -0.35, leading to an increase in the number of error contributions. The errors in (8) grow with polynomial order, as does the overall cost of computing the heterodyned likelihood. The sweet spot is to use a quadratic ($J = 2$) fit with an error tolerance between 0.01 and 0.001. While illustrated here for just one system, a similar behavior was found to hold across the mass spectrum.

In considering what an acceptable error tolerance for $\Delta \ln L$ might be, recall that one standard deviation in the likelihood is $\sigma_{\ln L} \simeq (D/2)^{1/2}$. For the IMRPhenomD waveform model [25] used here, $D = 11$ and $\sigma_{\ln L} \simeq 2.3$. The finite sampling that occurs in any numerical approach to Bayesian inference will introduce uncertainties of at least a few percent of the posterior distribution, so demanding that $|\Delta \ln L| < 0.1$ should be sufficient for most applications. Interestingly, while the heterodyning procedure was designed to work for waveforms that are "close" to the reference waveform, it continues to work for waveforms that are far from the reference. This is illustrated in Fig. 7, where the error in the likelihood is shown as a function for the likelihood for both accepted and proposed points in a MCMC run.

Somewhat surprisingly, the heterodyne is accurate all the way down to zero relative log likelihood. This means that the heterodyned likelihood can not only be used for parameter estimation but can also be used for computing the model evidence using methods such as thermodynamic integration [27]. Thermodynamic integration requires us to compute the integral

$$I = \int_0^1 \langle \ln L \rangle_\beta d\beta = \int_{-\infty}^0 \beta \langle \ln L \rangle_\beta d \ln \beta. \qquad (12)$$

Here, $\langle \ln L \rangle_\beta$ is the average log likelihood for chains with inverse temperature β. Figure 8 shows the average log likelihood as a function of inverse temperature using the full and heterodyned likelihood. The lower panel of the figure shows the difference in the integrand. The evidence computed using the heterodyned likelihood agrees with the value computed using the full likelihood to within $\Delta I = 0.1$.

The heterodyning approach is widely applicable. For example, it can handle the very high signal-to-noise (SNR) systems that are expected to be detected by the Laser Interferometer Space Antenna (LISA). The high SNR demands a smaller linear error tolerance, but the savings are still very large. For example, typical LISA sources with masses in the $10^5 \to 10^7\ M_\odot$ range, such as those shown in

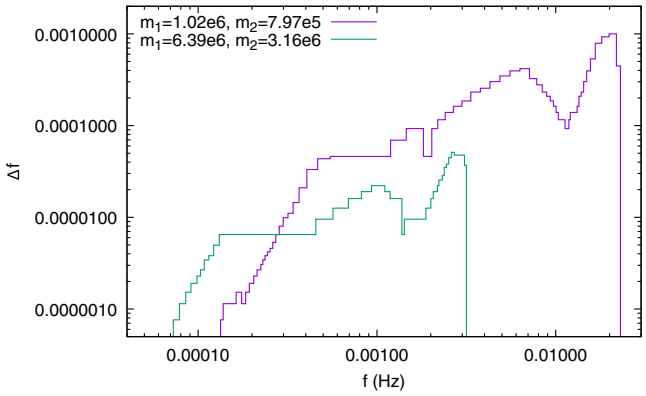

FIG. 9. Frequency spacing for two massive black hole binaries taken from the second LISA Data Challenge training data set. Despite the difference in total mass, both system required roughly 200 frequency samples in the heterodyne. The detector frame masses of the black holes (in units of solar mass) are given in the figure legend. The heterodyne used second-order Legendre polynomials with a linear fit tolerance of 10^{-5} and a chi-squared value of $(\Delta h|\Delta h) = 50$.

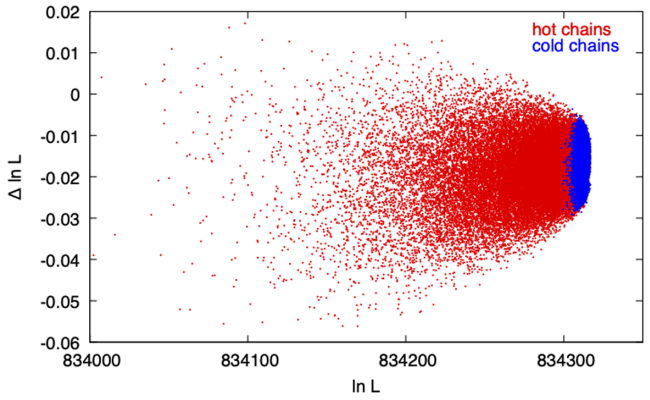

FIG. 10. Difference between the full and heterodyned likelihood for a simulated LISA black hole binary with detector frame masses $m_1 = 1.02 \times 10^6\ M_\odot$, $m_2 = 7.97 \times 10^5\ M_\odot$. A parallel tempered MCMC was used with four cold chains and 12 hot chains, geometrically spaced in temperature by a factor of 1.3. The signal to noise of the system was very high: SNR = 1292. The heterodyne used second-order Legendre polynomials with a linear fit tolerance of 10^{-5} and a chi-squared value of $(\Delta h|\Delta h) = 50$.

Fig. 9, can be accurately covered by frequency grids with just a few hundred points. This is far fewer than the $\sim 10^6$ samples that are required when computing the regular likelihood for these sources.

The accuracy of the heterodyned likelihood, even for LISA sources with signal-to-noise ratios in the thousands, such as the system shown in Fig. 10, is impressive: absolute errors of order 0.02 and fractional errors of order 10^{-8}.

C. Noise marginalization

To produce parameter estimates that are robust against instrument calibration uncertainties and finite sample uncertainties in the on-source noise spectral estimates, it is desirable to marginalize over the calibration model and noise model. This poses a challenge for rapid parameter estimation techniques that rely on precomputing quantities at full cadence. For the heterodyned likelihood, calibration uncertainties pose no problem as they introduce small changes in the amplitude and phase evolution that can be incorporated in the slow terms and thus have minimal impact on the computational cost. In contrast, changes in the noise model are generally not smooth due to the presence of sharp spectral lines. Figure 11 shows the median and 90% credible band for the on-source power spectrum model in the LIGO Hanford detector using 8 sec of data surrounding the black binary hole merger GW151012. Here, the spectral model is a fixed dimension variant of the `BayesLine` model [28] used by the `QuickCBC` [9] parameter estimation pipeline.

The variations in the power spectrum look small when viewed on a logarithmic scale, and it is tempting to try and incorporate these variations by writing

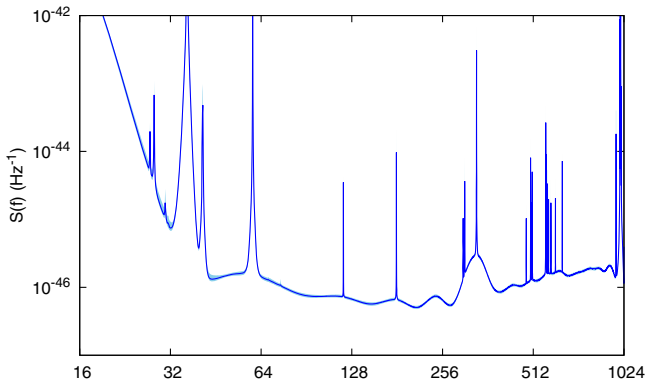

FIG. 11. Power spectral density estimates for the LIGO Hanford detector using 4 sec of data surrounding the binary merger GW151012. The light blue band indicates the 90% credible region, while the solid dark blue lines indicates the median.

$\kappa(f) = d(f)\bar{h}^*(f)/\bar{S}(f)$ for the fast term and $\zeta(f) = \bar{S}(f)h(f)/(S(f)\bar{h}(f))$ for the slow term, where $\bar{S}(f)$ is some reference model for the power spectral density (PSD). Figure 12 compares the ratio of a fair draw from the power spectrum to the median of the spectral model. The spectral lines lead to sharp features in the ratio that prevent it from being incorporated into the slow varying terms in the heterodyned likelihood. One way of handling the lines would be to excise the region around each line and calculate the likelihood directly in those regions and use the heterodyne for the remainder. A simpler approach is to use "blocked Gibbs" sampling, whereby the MCMC sampler alternates between updating the source parameters and the noise model parameters, with each noise model update followed by a recomputation of the Legendre polynomial expansion of the slow terms. This latter approach is relatively inexpensive since the reference waveform does not have to be recomputed, just the Legendre expansion. Typically, the cost of the recomputation is several times less than a standard likelihood

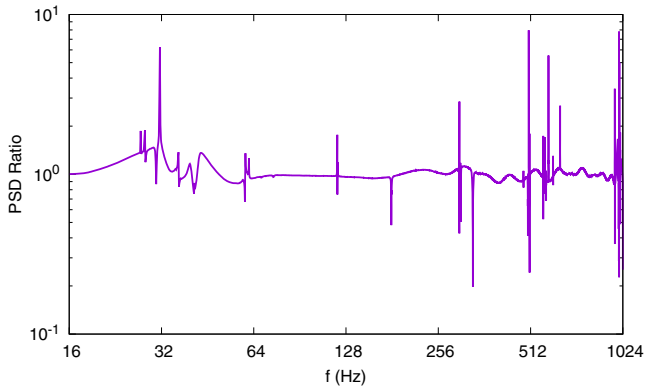

FIG. 12. The ratio of a fair draw from the power spectrum model and the reference power spectrum model for the LIGO Hanford detector using 8 sec of data surrounding GW151012.

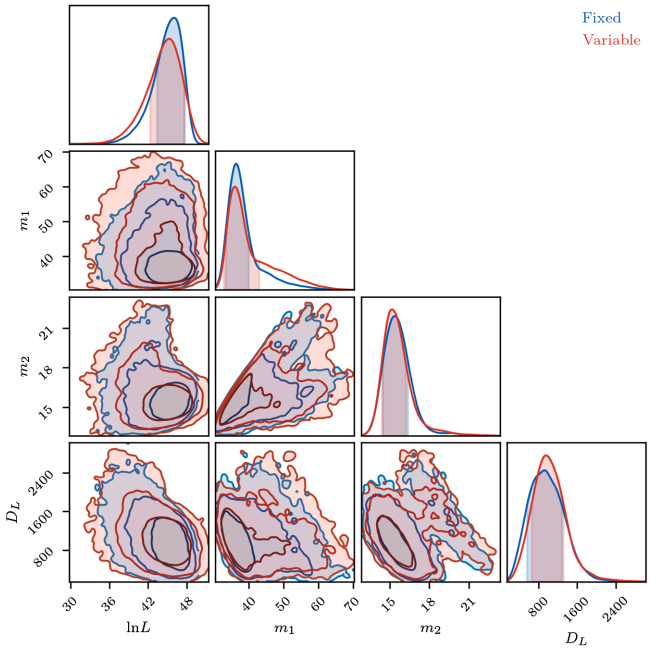

FIG. 13. Posterior distributions for GW151012 with (variable) and without (fixed) marginalization over the power spectral density. Marginalizing over the PSD slightly inflates the spread in the posterior distributions.

evaluation. Moreover, if using multiple chains in a parallel tempered setup, it is sufficient to limit the noise model updates to the cold chain and to share the update with the hot chains. In this way, noise marginalization can be incorporated at little additional cost.

Figure 13 compares the posterior distributions for the masses and distance of binary merger GW151012 with and without noise marginalization. Marginalizing over the noise model slightly inflates the widths of the posterior distributions and slightly shifts the peaks. Similar small shifts were found for other systems in the LIGO-Virgo catalog [29,30], suggesting that noise marginalization is probably less impactful than waveform uncertainties or calibration uncertainties (a similar conclusion was reached in Ref. [31]).

The heterodyne procedure can also be applied to data with nonstationary noise that is locally stationary [32,33]. Locally stationary data can be whitened and made stationary by transforming to the discrete wavelet domain, and dividing each wavelet pixel by the square root of the dynamic, or evolutionary spectrum, $S(f,t)$. This procedure decorrelates the noise in both time and frequency [34]. The rescaled white/stationary data can then be transformed to the frequency domain and used in the heterodyned likelihood. The steps are illustrated in Eq. (13),

$$d(t) \xrightarrow[\text{DFT}]{} d(t,f) \xrightarrow[\text{decorr}]{} \frac{d(t,f)}{\sqrt{S(f,t)}} \xrightarrow[\text{FT}]{} \tilde{d}_w(f). \tag{13}$$

The waveform template $\tilde{h}(f)$ can be dynamically whitened using the time-frequency mapping,

$$t(f) = \frac{1}{2\pi}\frac{d\bar{\Phi}(f)}{df}, \tag{14}$$

and defining $S(f) = S(f, t(f))$. There will be a different time-frequency mapping for each harmonic. Using this procedure, nonstationary gravitational wave data can be analyzed just as quickly as stationary data.

III. SUMMARY

The heterodyned likelihood [7] can be used to dramatically speed up gravitational wave parameter inference without sacrificing accuracy. The heterodyning procedure can be efficiently implemented using discrete Legendre polynomial expansions and a dynamic spacing of the frequency samples. The method can be applied to any waveform model and detector configuration, with the largest savings in computational cost occurring for low mass systems. The savings decrease as the number of harmonics in the waveform model increase, since each harmonic has to be be treated separately. The heterodyning approach can incorporate marginalization over calibration uncertainties and variations in the noise model.

ACKNOWLEDGMENTS

The author thanks Tyson Littenberg and Katerina Chatziioannou for discussions about noise marginalization and Soichiro Morisaki for providing feedback on an earlier draft. This work was supported by NSF Grant No. PHY1912053 and NASA LISA foundation Science Grant No. 80NSSC19K0320 and used data obtained from the LISA Data Challenge [35] and the Gravitational Wave Open Science Center [36], a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

[1] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin *et al.*, Phys. Rev. D **91**, 042003 (2015).
[2] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[3] K. Cannon, C. Hanna, and D. Keppel, Phys. Rev. D **85**, 081504 (2012).
[4] R. J. E. Smith, K. Cannon, C. Hanna, D. Keppel, and I. Mandel, Phys. Rev. D **87**, 122002 (2013).
[5] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Phys. Rev. Lett. **106**, 221102 (2011).
[6] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014).
[7] N. J. Cornish, arXiv:1007.4820.
[8] N. J. Cornish and K. Shuman, Phys. Rev. D **101**, 124008 (2020).
[9] N. J. Cornish, Phys. Rev. D **103**, 104057 (2021).
[10] B. Zackay, L. Dai, and T. Venumadhav, arXiv:1806.08792.
[11] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. D **92**, 023002 (2015).
[12] N. J. Cornish, arXiv:1606.00953.
[13] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Phys. Rev. D **94**, 044031 (2016).
[14] S. Morisaki and V. Raymond, Phys. Rev. D **102**, 104020 (2020).
[15] S. Morisaki, Phys. Rev. D **104**, 044062 (2021).
[16] D. Wysocki, R. O'Shaughnessy, J. Lange, and Y.-L. L. Fang, Phys. Rev. D **99**, 084026 (2019).
[17] R. J. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Mon. Not. R. Astron. Soc. **498**, 4492 (2020).
[18] M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, Phys. Rev. D **104**, 064047 (2021).
[19] D. George and E. Huerta, Phys. Lett. B **778**, 64 (2018).
[20] A. Delaunoy, A. Wehenkel, T. Hinderer, S. Nissanke, C. Weniger, A. R. Williamson, and G. Louppe, arXiv:2010.12931.
[21] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, arXiv:2106.12594.
[22] C. P. Neuman and D. I. Schonbach, Int. J. Numer. Methods Eng. **8**, 743 (1974).
[23] N. Hale and A. Townsend, SIAM J. Sci. Comput. **36**, A148 (2014).
[24] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).
[25] L. Santamaria *et al.*, Phys. Rev. D **82**, 064016 (2010).
[26] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119**, 161101 (2017).
[27] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007).
[28] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015).
[29] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).
[30] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **11**, 021053 (2021).
[31] S. Biscoveanu, C.-J. Haster, S. Vitale, and J. Davies, Phys. Rev. D **102**, 023008 (2020).
[32] R. Dahlhaus, Ann. Stat. **25**, 1 (1997).
[33] M. Vogt, Ann. Stat. **40**, 2601 (2012).
[34] N. J. Cornish, Phys. Rev. D **102**, 124038 (2020).
[35] https://lisa-ldc.lal.in2p3.fr.
[36] https://www.gw-openscience.org.