

## Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources

Lalit Pathak<sup>1,\*</sup>, Amit Reza<sup>2,3,†</sup> and Anand S. Sengupta<sup>1,‡</sup>

<sup>1</sup>*Indian Institute of Technology Gandhinagar, Gujarat 382055, India*

<sup>2</sup>*Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*

<sup>3</sup>*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands*



(Received 2 November 2022; revised 6 July 2023; accepted 23 August 2023; published 28 September 2023)

Several rapid parameter estimation methods have recently been advanced to deal with the computational challenges of the problem of Bayesian inference of the properties of compact binary sources detected in the upcoming science runs of the terrestrial network of gravitational wave detectors. Some of these methods are well-optimized to reconstruct gravitational wave signals in nearly real-time necessary for multimessenger astronomy. In this context, this work presents a new, computationally efficient algorithm for fast evaluation of the likelihood function using a combination of numerical linear algebra and meshfree interpolation methods. The proposed method can rapidly evaluate the likelihood function at any arbitrary point of the sample space at a negligible loss of accuracy and is an alternative to the grid-based parameter estimation schemes. We obtain posterior samples over model parameters for a canonical binary neutron star system by interfacing our fast likelihood evaluation method with the nested sampling algorithm. The marginalized posterior distributions obtained from these samples are statistically identical to those obtained by brute force calculations. We find that such Bayesian posteriors can be determined within a few minutes of detecting such transient compact binary sources, thereby improving the chances of their prompt follow-up observations with telescopes at different wavelengths. It may be possible to apply the blueprint of the meshfree technique presented in this study to Bayesian inference problems in other domains.

DOI: [10.1103/PhysRevD.108.064055](https://doi.org/10.1103/PhysRevD.108.064055)

### I. INTRODUCTION

The detection of gravitational waves (GW) from the GW170817 [1] binary neutron star (BNS) system, followed by the prompt multiwavelength (gamma rays to radio) observation of its electromagnetic (EM) counterpart, has led to several fundamental discoveries; and is hailed as a significant breakthrough in astronomy. These discoveries include the validation of long-held hypotheses that BNS mergers are ideal sites for r-process nucleosynthesis and produce short gamma ray bursts, the first GW-based constraints on the equation of state of nuclear matter in such stars, and the measurement of Hubble constant independent of the cosmic distance ladder.

The inevitable improvement of the detectors' sensitivity in future observation runs is likely to have a twofold impact on the prospects of multimessenger observations: *Firstly*, the increased bandwidth of the detectors (especially improved sensitivity at low frequencies) will result in a tremendous increase in the computational cost of Bayesian

inference of source parameters, including sky localization essential for prompt observation of EM counterparts. Although the BAYESTAR [2] algorithm could be used to produce rapid sky maps, it has been recently shown [3] that coherent parameter estimation (PE) can localize the sources better by an average reduction of  $14 \text{ deg}^2$  in the uncertainty, underlining the importance of developing fast PE algorithms. *Second*, the reach of the terrestrial network of GW detectors will extend out to several Gpc to the effect that one would have far too many detections of BNS/NSBH signals to contend with whilst generating prompt sky-location maps [4]; so much so that one may have to prioritize the GW sources for EM follow-up based on prospects of new science from a rapid estimation of their mass and spin components as shown by Margalit and Metzger [5], thereby helping EM observatories to use resources optimally. Several fast PE algorithms have been developed recently, such as the coherent multidetector extension of the relative binning/heterodyne method by Finstaaad and Brown (2020) [3], which produces the posterior within twenty minutes for BNS systems with 32 CPU cores. Well-trained machine learning PE methods [6–8] can significantly reduce the runtimes and produce the posteriors in nearly real-time. In the past, algorithms for

\*lalit.pathak@iitgn.ac.in

†areza@nikhef.nl

‡asengupta@iitgn.ac.in

accelerated parameter estimation have mainly focussed on speeding up the overlap integral. These include reduced-order models (ROMs) [9–11], machine-learning aided ROMs [12], Gaussian process regression based interpolation [13] and relative binning [3,14,15] algorithms. Our approach takes inspiration from the grid-based likelihood interpolation method [16] based on orthonormal Chebyshev polynomials. The grid-based techniques have a drawback in that the number of interpolation nodes grows exponentially with the dimensionality of the parameter space.

In this work, we propose a new and alternative approach to grid-based likelihood interpolation method [16], a computationally efficient method for evaluating the likelihood function (a key ingredient in Bayesian inference) using meshfree interpolation methods with dimension reduction techniques. We directly interpolate the likelihood function over the parameter space, bypassing the generation of templates and brute-force computation of the overlap integral altogether. Our scheme can quickly approximate the log-likelihood function with high accuracy and produce statistically indistinguishable posteriors over source parameters. Further, both the `GstLAL` search framework [17] and the meshfree method use the idea of dimension reduction using singular value decomposition (SVD) [18], it may be prudent to incorporate this method with the low-latency `GstLAL` search pipeline for rapid, automated follow-ups of the detected events.

## II. BAYESIAN INFERENCE

Given data  $\mathbf{d} = \mathbf{h}(\vec{\Lambda}_{\text{true}}) + \mathbf{n}$  recorded at a detector containing an astrophysical GW signal  $\mathbf{h}(\vec{\Lambda}_{\text{true}})$  embedded in additive Gaussian noise  $\mathbf{n}$ , one is interested in solving the inverse problem to estimate the source parameters. Bayesian inference is a stochastic inversion method where the posterior probability density  $p(\vec{\Lambda}|\mathbf{d})$  over the source parameters is related to the likelihood function  $\mathcal{L}(\mathbf{d}|\vec{\Lambda})$  of observing the data through the Bayes' theorem,

$$p(\vec{\Lambda}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\vec{\Lambda})p(\vec{\Lambda})}{p(\mathbf{d})}, \quad (1)$$

where  $p(\vec{\Lambda})$  is the prior distribution over the model parameters  $\vec{\Lambda} \equiv \{\vec{\lambda}^{\text{ext}}, \vec{\lambda}\}$ . In our notation,  $\vec{\lambda}$  denotes the intrinsic parameters such as component masses and spins. The set of extrinsic parameters is denoted by  $\vec{\lambda}^{\text{ext}}$ . We are particularly interested in estimating the extrinsic parameter  $t_c$  denoting the fiducial time of coalescence of the two masses.  $t_c$  will be mentioned explicitly wherever required, as it is treated in a special way in our analysis.

The forward generative frequency-domain restricted waveform model for nonprecessing compact binaries can be expressed as  $\mathbf{h}(\vec{\Lambda}) = \mathcal{A}h_+(f_k; \vec{\lambda})$ , where the complex

amplitude  $\mathcal{A}$  depends only on the extrinsic parameters and  $h_+(f_k; \vec{\lambda})$  is the '+' polarization of the signal that depends only on the intrinsic parameters [19]. Here  $\{f_k\}_{k=0}^{N_s/2}$  defines positive Fourier frequencies, and  $N_s$ <sup>1</sup> is the number of sample points. A GW signal, observed by an interferometric detector, can be considered as a linear combination of the two polarizations weighted by the antenna pattern function. The  $h_+(f_k; \vec{\lambda})$  polarization is related to the  $h_\times(f_k; \vec{\lambda})$  polarization for nonprecessing GW signal as  $h_+(f_k; \vec{\lambda}) \propto ih_\times(f_k; \vec{\lambda})$  [20]. This allows us to write the detector response in terms of any one of the polarizations alone [we have chosen the  $h_+(f_k; \vec{\lambda})$  polarization]. Using this model, the posterior  $p(\vec{\Lambda}|\mathbf{d})$  can be directly evaluated at every point in  $\vec{\Lambda}$  using Eq. (1). However, in view of the high dimensionality of  $\vec{\Lambda}$ , it is more efficient to sample the posterior using stochastic sampling algorithms such as nested sampling [21], or Markov Chain Monte Carlo (MCMC) [19]. From Eq. (1), it is evident that for a quick estimation of the posterior distribution, it is imperative to rapidly evaluate the likelihood function.

We work with the phase-marginalized log-likelihood function [22],

$$\ln \mathcal{L}(\vec{\Lambda}, t_c) = \ln I_0[|\mathcal{A}|z(\vec{\lambda}, t_c)] - \frac{1}{2} \|\mathbf{h}(\vec{\Lambda})\|_2^2, \quad (2)$$

where  $I_0(\cdot)$  is the zeroth-order modified Bessel function of the first kind, and  $z(\vec{\lambda}, t_c)$  is the frequency-domain overlap integral,

$$z(\vec{\lambda}, t_c) = 4\Delta f \left| \sum_{k=0}^{N_s/2} \frac{d^*(f_k)h_+(f_k, \vec{\lambda})}{S_h(f_k)} e^{-2\pi i f_k t_c} \right|, \quad (3)$$

inversely by  $S_h(f_k)$ , the detector's one-sided noise power spectral density (PSD). The data and template vectors are sampled at discrete frequencies  $\{f_k\}_{k=0}^{N_s/2}$ .

The complexity of evaluating the overlap integral scales directly with the number of data samples, which in turn, scales with the seismic cutoff frequency (approximately) as  $N_s \sim f_{\text{low}}^{-8/3}$ . As we progress from the O4 observational run ( $f_{\text{low}} = 20$  Hz) to O5 at design sensitivity ( $f_{\text{low}} = 10$  Hz), evaluating  $p(\vec{\Lambda}|\mathbf{d})$  is likely to take at least  $\times 6.3$  longer. In addition, additional costs will be incurred in constructing longer templates at the proposal points. Therefore, the likelihood calculation can be expensive. However, our method is immune to this issue as our scheme directly approximates the likelihoods at different sample points. In this work, we have used nonorthonormal radial basis functions (RBFs) (Gaussian kernels) centered at interpolation nodes that can be randomly scattered over the volume

<sup>1</sup> $N_s$  = signal duration  $\times$  sampling frequency.

of the intrinsic parameter space. In this manner, we have effective control of their number in higher-dimensional parameter spaces.

### III. MESHFREE LIKELIHOOD INTERPOLATION

The computational cost of Bayesian inference comprises of two parts: the first is incurred in waveform generation followed by likelihood evaluation at a point proposed by the sampler. Typically, a sampler proposes a large number ( $\sim 10^6$ – $10^7$ ) of points to adequately capture the posterior distribution which makes this part computationally expensive. The other part of the total computational cost can be attributed to the overheads of the sampling method itself. Since the latter cost depends on the efficiency of the sampling algorithm used (and its software implementation) and is significantly less in comparison to the overall cost of PE, we shall ignore it in our discussions.

We assume that the parameter estimation is “seeded” by the most significant trigger  $\vec{\Lambda}^*$  from an upstream detection pipeline [23,24]. The sampling algorithm draws new proposals from a sample space which is taken to be a moderate-sized “hyper-rectangle” in  $\vec{\Lambda}$ , centered around the most significant search trigger.

From Eqs. (2) and (3), it is evident that we need to interpolate two pieces,  $z(\vec{\lambda}, t_c)$  and  $\|h_+(\vec{\lambda})\|_2^2$ , and combine them with the amplitude  $\mathcal{A}$  to calculate the log-likelihood ratio at a given ‘query’ point  $\vec{\lambda}^q$  and a particularly given value of  $t_c$ . For this purpose, a set of  $n$  unique interpolation nodes  $\vec{\lambda}^\alpha, \alpha = 1, \dots, n$  are randomly chosen from a uniform distribution over the sample space.

The template norm  $\|h_+(\vec{\lambda})\|_2^2$  is a smoothly varying scalar field over  $\vec{\lambda}$ . We can interpolate its value at a query point  $\vec{\lambda}^q$  by first evaluating the values explicitly at the interpolation nodes  $\vec{\lambda}^\alpha$ , and then expressing  $\|h_+(\vec{\lambda}^q)\|_2^2$  at an arbitrary point as a linear combination of Gaussian RBF kernels centered at these nodes. The unknown coefficients of this linear combination can be uniquely found by enforcing the interpolation criteria as explained in the next section.

On the other hand, as the overlap integral has to be evaluated at an arbitrary point  $(\vec{\lambda}^q, t_c)$ , it will turn out to be more convenient to interpolate it as a vector. In this case, a set of overlap-integral vectors  $\vec{z}_\alpha$  are first constructed at the interpolation nodes, sampled on a uniform grid over  $t_c$ . The vectors  $\vec{z}_\alpha$  have elements  $z_\alpha[k] \equiv z(\vec{\lambda}^\alpha, k\Delta t)$ , where  $\Delta t$  is the sampling interval,  $k$  is an integer  $\in \text{int}([t_c^* \pm \tau]/\Delta t)$ ,  $t_c^*$  represents the ‘reference’ coalescence time as triggered by the search pipeline and  $2\tau$  is the dimension of the sample-space (hyper-rectangle in  $\vec{\Lambda}$ ) along the  $t_c$  direction.

Since the bulk of the support for the posterior distribution comes from near the peak of these time series, we choose its samples that are centered around the triggered value. From Eq. (3) it is clear that  $\vec{z}_\alpha$ ’s can be constructed

efficiently using FFT correlations. Once the set of vectors  $\{\vec{z}_\alpha\}$  is available, they can be projected over a suitable set of basis vectors  $\{\vec{u}_\mu\}$  with linear coefficients  $\{C_\mu(\vec{\lambda}^\alpha)\}$ . As each of the coefficients is smoothly varying scalar fields over  $\vec{\lambda}$  (sampled at the interpolation nodes), we can use meshfree methods to interpolate their values at an arbitrary query point. The meshfree scheme can be divided into two stages: (i) a preparatory, startup stage where we explicitly determine the RBF interpolating functions (interpolants) from the precomputed likelihood values at the interpolation nodes; and (ii) an online stage where these interpolants are evaluated on the fly to ‘predict’ the likelihood values at arbitrary query points in the sample space.

#### A. Start-up stage

In this stage, the nodes are randomly sprayed over the sample space, and the meshfree interpolants are constructed;

- (1) *SVD basis*: We are interested in finding a suitable set of basis vectors that span the space of  $n$  input overlap vectors  $\{\vec{z}_\alpha\}$ . This is conveniently performed by stacking these vectors row-wise and performing a SVD of the resultant matrix,

$$\vec{z}_\alpha = \sum_{\mu=1}^n C_\mu^\alpha \vec{u}_\mu, \quad (4)$$

where  $C_\mu^\alpha$  are the coefficients for the set of orthonormal basis vectors  $\vec{u}_\mu$  in decreasing order of their relative importance as determined from the spectrum of singular values. A strong correlation between the  $\vec{z}_\alpha$ ’s implies that the overlap vectors lie in the span of the top- $\ell$  basis vectors ( $\ell \ll n$ ). A vector  $\vec{z}_q$  at a query point in the sample space can also be spanned by the same set of basis vectors.

Note that for a fixed index  $\mu$ , the coefficients  $C_\mu^\alpha$  represent a surface whose values are known only at the input nodes. Along with Eq. (4), this implies that the interpolation of the inner-product vector at an arbitrary query point essentially boils down to interpolating the value of the coefficients  $C_\mu^q \equiv C_\mu(\vec{\lambda}^q)$ .

- (2) *Creating meshfree interpolants*: In this step, we create and explicitly determine the “meshfree” interpolants for each of the coefficient surfaces [that appear in Eq. (4)] independently. The interpolant for the coefficient corresponding to the  $\mu$ th basis vector can be expressed [25] as a linear combination of RBF kernels centered on the scattered, distinct nodes  $\vec{\lambda}^\alpha$ , augmented by monomials ranging up to a specific order,

$$C_\mu^q = \sum_{\alpha=1}^n r_\alpha \phi(\|\vec{\lambda}^q - \vec{\lambda}^\alpha\|_2) + \sum_{j=1}^M b_j f_j(\vec{\lambda}^q), \quad (5)$$

where  $\phi$  is the Gaussian kernel centered on  $\vec{\lambda}^\alpha \in \mathbb{R}^d$ , and  $\{f_j\}$ 's are monomials that span the space of polynomials of some preset target degree  $\nu$  in  $d$  variables. Also,  $\mathbf{r} = [r_1, r_2, \dots, r_n]^T$  and  $\mathbf{b} = [b_1, b_2, \dots, b_M]^T$  are the set of  $(n + M)$  coefficients that need to be uniquely determined to determine the interpolant.

Since  $C_\mu^q$  are known at the interpolation nodes, it allows us to enforce  $n$  interpolation conditions.  $M$  additional conditions  $\sum_{k=1}^n r_k f_j(\vec{\lambda}^k) = 0, j = 1, \dots, M$  are added to ensure a unique solution. Together, these lead to a system of equations:

$$\begin{bmatrix} \mathbf{K} & \mathbf{F} \\ \mathbf{F}^T & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_\mu^\alpha \\ \mathbf{0} \end{bmatrix}, \quad (6)$$

where the matrices  $\mathbf{K}$  and  $\mathbf{F}$  have components  $K_{ij} = \phi(\|\vec{\lambda}^i - \vec{\lambda}^j\|_2)$  and  $F_{ij} = f_j(\vec{\lambda}^i)$  respectively;  $\mathbf{O}_{M \times M}$  is a zero matrix and  $\mathbf{0}_{M \times 1}$  is a zero vector. Equation (6) can be solved uniquely for the unknown coefficients  $\mathbf{r}$  and  $\mathbf{b}$ , thus completely determining the meshfree interpolant in Eq. (5). The solution for  $\mathbf{r}$  and  $\mathbf{b}$  can be shown to be unique if  $\mathbf{F}$  has full column rank.

A minimum set of  $n = (\nu + d)$  interpolation nodes are required to be uniformly distributed over the  $d$ -dimensional intrinsic parameter space to determine the coefficients uniquely. Euclidean distances between pairs of points seem to work well, possibly due to the small volume of the sample space in a typical PE analysis. Parameter-space metric-based distances could also be used.

A similar procedure is followed to create a separate meshfree interpolant for  $\|h_+(\vec{\lambda}^q)\|_2^2$ .

Note that by construction, the meshfree interpolation eliminates both the factors that contribute to the high computational cost, namely, (a) generating the forward signal model and (b) explicitly calculating the overlap integral at every query point proposed by the sampling algorithm.

### B. Online stage

In this stage, the  $(\ell + 1)$  interpolants (prepared in the offline stage earlier) are evaluated on the fly at arbitrary query points  $(\vec{\lambda}^q, t_c)$  proposed by the sampling algorithm.

The interpolant for the square of the template norm can be directly evaluated to get the interpolated value  $\|h_+(\vec{\lambda}^q)\|_2^2$ . Similarly, the  $\ell$  interpolants for SVD coefficients [Eq. (5)] are evaluated to get a set of interpolated coefficients, which are then combined with the corresponding top- $\ell$  basis vectors  $\vec{u}_\mu$  [see Eq. (4)] to obtain the interpolated overlap integral  $\vec{z}_q$ .

By construction,  $\vec{z}_q$  is uniformly sampled over the interval  $[t_c^* \pm \tau]$ . As such, it is possible that the query  $t_c$  does not coincide with the discrete time samples of  $\vec{z}_q$ .

In such a case, we use a one-dimensional cubic-spline interpolation to evaluate  $z(\vec{\lambda}^q, t_c)$  using a few ‘nearby’ grid samples of  $\vec{z}_q$  as input. This implies that  $\vec{z}_q$  has to be reconstituted only at a few ( $\sim 10$ ) consecutive sample points, which considerably accelerates the matrix-vector multiplication in Eq. (4).

Combining the interpolated values  $z(\vec{\lambda}^q, t_c)$  and  $\|h_+(\vec{\lambda}^q)\|_2^2$  with the extrinsic parameter-dependent complex amplitude  $\mathcal{A}$  [see Eq. (2)], we finally obtain the log-likelihood ratio  $\ln \mathcal{L}$  at the arbitrary point  $(\vec{\lambda}^q, t_c)$ .

## IV. NUMERICAL EXPERIMENTS

To demonstrate the accuracy and speed of the meshfree method in reconstructing the source parameters, we prepared synthetic 360 s long data  $\mathbf{d}$ , sampled at 4096 Hz. For this, a simulated GW signal  $\mathbf{h}$  from a canonical BNS system with component masses  $m_{1,2} = 1.4M_\odot$  and mass-weighted effective dimensionless spin  $\chi_{\text{eff}} = 0.05$  was injected (using the IMRPhenomD [26] signal model) in colored Gaussian noise using the noise power spectral density model [27] of aLIGO detectors. The distance to the source was adjusted for a moderate matched-filtering signal-to-noise ratio (SNR) of 10. The seismic cutoff frequency was chosen to be 20 Hz to mimic data from the upcoming O4 science run.

We performed Bayesian inference on this simulated data using both (a) the direct likelihood calculation used in PyCBC inference, and (b) by using the proposed meshfree likelihood interpolation scheme outlined in earlier sections. We used publicly available software [28] for radial basis functions and the DYNesty [29,30] nested sampling package for carrying out the Bayesian inference analysis. We varied four intrinsic parameters (component masses and aligned-spin magnitudes) and two extrinsic parameters (luminosity distance and coalescence time), keeping other parameters fixed.

For this exercise, we used  $n = 800$  random input nodes over  $\vec{\lambda}$ . The top  $\ell = 120$  basis vectors and a polynomial order  $\nu = 6$  with a corresponding nominal median relative error  $\sim 10^{-5}$  across the sample space in approximating the log-likelihood function. The accuracy trade-offs of likelihood reconstruction by varying basis size is shown in the Appendix A.

The meshfree parameter reconstruction was completed in 5.3 min in comparison to 31.7 h taken by the direct calculation. The likelihood function was evaluated 676 times faster using the meshfree method.

Some of the estimated parameters have been compared in Table I, which show identical values obtained by both methods. The marginalized probability density function (PDF) over three parameters  $\mathcal{M}$ ,  $\eta$ , and  $\Delta t_c$  are shown in the corner plot Fig. 1. The figure also contains cumulative distribution function (CDF) profiles of these distributions obtained from both PyCBC inference and the meshfree

TABLE I. Reconstruction of a canonical BNS event.

	$M/M_{\odot}$	$\eta$	$\chi_{\text{eff}}$	SNR
Injection	1.2187	0.25	0.05	10.00
Standard	$1.2187^{1.2188}_{1.2185}$	$0.2487^{0.2498}_{0.2456}$	$0.050^{0.051}_{0.049}$	9.67
Meshfree	$1.2187^{1.2188}_{1.2185}$	$0.2487^{0.2498}_{0.2456}$	$0.050^{0.051}_{0.049}$	9.67

method plotted together. While both the PDF and CDF profiles look virtually indistinguishable, we also calculate statistical measures of similarity [31] using the PDFs obtained from the two methods for further validation: both the Kolmogorov-Smirnov statistic (0.0130) and the Bhattacharyya distance (0.0006) between the chirp-mass PDF profiles support the fact that the two distributions are nearly identical. We get similar results for posterior distributions of other parameters.

This numerical example shows that the meshfree method can generate a statistically indistinguishable replica of the posterior distributions in a GW Bayesian inference problem at a small fraction of the total computational cost.

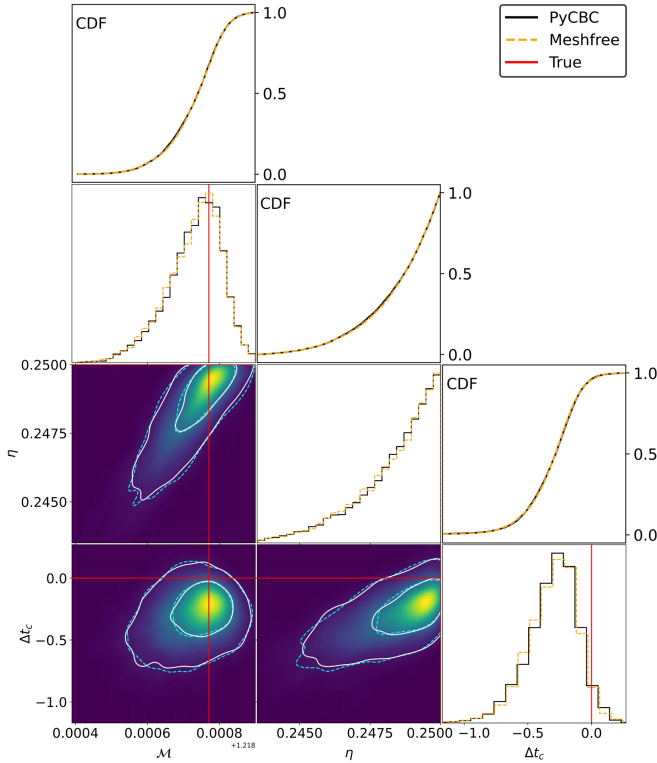


FIG. 1. Marginalized PDF for chirp mass  $\mathcal{M}$ , symmetric mass ratio  $\eta$ , and  $\Delta t_c$  parameters of a simulated BNS event at a seismic cutoff of 20 Hz. The injection parameters are shown as red lines. The 50% and 90% contours for meshfree method (dashed cyan trace) and PyCBC (solid white trace) are also shown. The plot-overlaid marginalized PDF obtained from the proposed meshfree method (dashed orange trace) and standard PyCBC inference (black line) are virtually indistinguishable.

## V. SPEED-UP ANALYSIS

We calculated the ratio of (average) time taken to compute the log-likelihood by these techniques at a fixed accuracy of reconstruction. Several simulated data sets were used in this study, generated by injecting signals having different parameters in colored Gaussian noise using the aLIGO noise model at a fixed matched-filtering SNR of 10. Seismic cutoff frequencies at 20 Hz (10 Hz) were considered to mimic data from upcoming O4 (O5) science runs.

There is an obvious trade-off between accuracy and speed-up of the meshfree method, which are determined by the choice of  $(n, \ell, \nu)$  parameters. Larger values can lead to more accurate likelihood estimates, albeit at a higher computational cost and vice-versa. We used a heuristic combination to guarantee median (relative) errors  $\lesssim 9 \times 10^{-5}$  in the estimated log-likelihood values across the entire sample space.

The log-likelihood function was evaluated and timed for a large number of random points uniformly distributed over the  $(\vec{\lambda}, t_c)$  space. Table II summarizes the speed-ups corresponding to the two different seismic cutoff frequencies for three compact binary systems with equal component masses. It is further elucidated in Fig. 2, where the speed-up comparison is drawn between a larger number of equal-mass binary systems covering a wider range of parameters. From Table II and Fig. 2, it is clear that the likelihood computation can be sped up  $\sim 4000$  times faster for canonical BNS systems at a nominal error of  $\sim 10^{-5}$  using the meshfree method against standard likelihood implementation in PyCBC inference. However, this would not reflect the true benefit of our proposed method, as PyCBC inference is not optimized to calculate the fast posterior distribution for LIGO. Our method computes the posterior within a few minutes for

TABLE II. The first and second columns define the seismic cutoff frequency and total mass of the injected GW signal, respectively. The third and fourth columns show the median time (in ms) taken for a single evaluation of the log-likelihood function using the standard PyCBC method and our approach at a nominal relative error of  $\mathcal{O}(10^{-5})$ . The last column provides the relative speed-up of our method in comparison with the standard likelihood calculation. Median time (in ms) taken for a single evaluation of the log-likelihood function using standard PyCBC method and the meshfree approach at a nominal relative error of  $\sim 10^{-5}$ .

Seismic cutoff	$M/M_{\odot}$	$t_{\text{mf}}/\text{ms}$	$t_{\text{pycbc}}/\text{ms}$	Speed-up
$f_{\text{low}} = 10$ Hz	2.8	0.90	3604.76	4005.2
	4.0	1.06	1405.84	1326.2
	20.0	0.88	29.27	33.2
$f_{\text{low}} = 20$ Hz	2.8	0.67	452.93	676.0
	4.0	0.81	207.35	256.0
	20.0	0.63	3.32	5.3

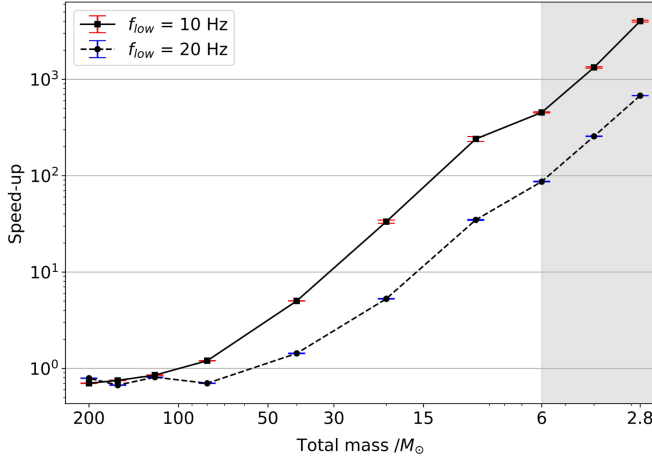


FIG. 2. Speed-up factors (averaged over several thousand evaluations) in calculating the log-likelihood function with the meshfree method for different equal-mass ( $q = 1$ ) systems with (dimensionless) spin magnitude of 0.05(0.2) for BNS (BBH) systems. The shaded region denotes BNS systems. The combination of  $(n, \ell, \nu)$  parameters were tuned for a median error of  $\sim 10^{-5}$ .

BNS system. Thus, our proposed scheme has the potential to perform rapid reconstruction of source parameters in upcoming observation runs for aLIGO detectors similar to the other existing optimized methods.

The low-mass systems with a large number of in-band cycles would benefit most from the meshfree method. In contrast to the standard method, the time taken by the meshfree method is relatively unaffected by the chirp time of the signals. All the tests were performed on a single-core AMD EPYC 7542 CPU@2.90GHz CPU.

## VI. CONCLUSION AND OUTLOOK

We have presented an alternative approach to the grid-based method [16], which is computationally efficient for accurately evaluating the log-likelihood function. Our meshfree method can be easily integrated into well-known sampling algorithms (e.g., MCMC, nested sampling) to substantially accelerate the Bayesian inference of source parameters of coalescing compact binary sources, triggering prompt observation of their EM counterparts in the future. Using synthetically generated data of a GW signal from a merging BNS system, we have demonstrated that the posterior distributions are statistically identical to those obtained using the standard PyCBC inference [32]. For BNS systems, the likelihood function can be calculated  $\sim 4000$  times faster at any point proposed by the sampling algorithm than the direct likelihood calculation implemented in PyCBC inference. At this point, we want to remind the readers that we made the comparison of our scheme against PyCBC inference only to verify the robustness of our scheme. PyCBC inference is not used for the fast PE run by LVK. Therefore, for a fair speed-up comparison,

we must compare our methods against those schemes [3,6,7,9–11,13–15] computing the real-time posterior distribution. However, the comparison with the optimized schemes is beyond the scope of the current work. In the follow-up works, we will compare our method against those schemes in detail.

Numerical experiments with a coherent, multidetector implementation of the meshfree algorithm suggest that we can solve the coherent BNS PE problem (over a ten-dimensional parameter space) in  $\sim 164$  seconds using 64 CPU cores. Further optimizations and runs over the full parameter space are underway. These details will be available in a follow-up paper [33]. It may be prudent to incorporate this method with the low-latency GstLAL [23] search pipeline for rapid, automated follow-ups of the detected events since both use the idea of dimension reduction using SVD. In the current implementation of this algorithm, we need to create meshfree RBF interpolants from scratch for each new event triggered by the search pipelines. However, this task is embarrassingly parallel and can benefit from multiple CPU cores to expedite the preparatory stage. Future refinements could involve using the Fisher matrix [34,35] as a guide to identifying the sample space volume and using sophisticated interpolation node distribution algorithms [36]. Finally, the techniques of dimension reduction and meshfree approximation could be applied to situations where the likelihood function varies smoothly over the sample space. It is thus possible to adapt this idea to Bayesian inference in fields as diverse as cosmology, biochemical kinetic processes, and systems biology.

## ACKNOWLEDGMENTS

We thank M. K. Haris for various suggestions and for his help at the conceptual stages of this work. We thank M. Vallisneri and T. Dent for carefully reading the manuscript and for offering several comments and suggestions to improve the presentation and content of the paper. We also thank other LSC colleagues S. Mitra, A. Ganguly, and S. Kapadia, for their useful comments. L. P. is supported by the Research Scholarship Program of Tata Consultancy Services (TCS). A. R. is supported by the research program of the Netherlands Organisation for Scientific Research (NWO). A. S. gratefully acknowledges the generous grant provided by the Department of Science and Technology, India, through the DST-ICPS cluster project funding. We would like to thank the HPC support staff at IIT Gandhinagar for their help and cooperation. The authors are grateful for the computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This material is based upon work supported by NSF's LIGO Laboratory, which is a major facility fully funded by the National Science Foundation.

### APPENDIX A: ACCURACY TRADE-OFFS: EFFECT OF VARYING BASIS SIZE

We present the results of a simulation to demonstrate the accuracy of the mesh-free method on the basis size. We inject a GW signal from a compact binary system with component masses ( $1.4M_{\odot}$ ,  $1.4M_{\odot}$ ) in simulated noise generated using the advanced LIGO design sensitivity curve with a lower seismic cut-off frequency of 10 Hz. On one hand, we compute the log-likelihood-ratio (LLR) using the standard functions available in the PyCBC package (considered to be the ‘ground truth’ for this simulation) and compare it with the approximate value using the mesh-free interpolation method as outlined in our manuscript. The meshfree method uses  $10^3$  randomly chosen initial nodes over the sample space. We vary the number of the basis vectors for reconstructing the likelihood and estimate the difference between interpolated likelihood and the ‘ground-truth’ likelihood value. For

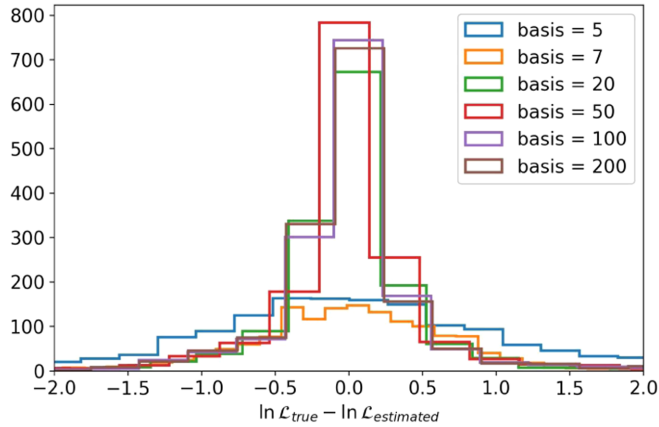


FIG. 3. The likelihood reconstruction error has been shown by varying the number of basis vectors.

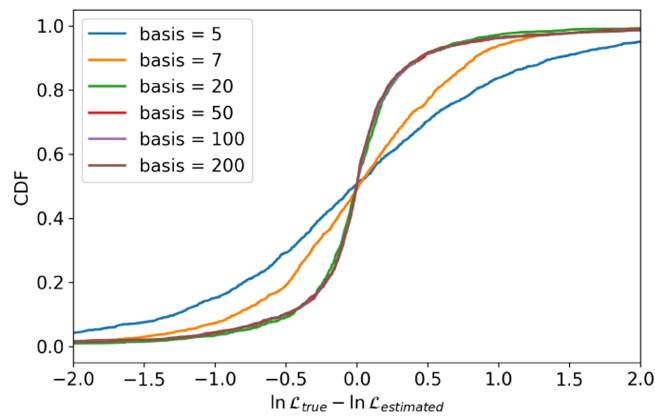


FIG. 4. The CDFs with different number of basis vectors has been plotted for the same likelihood reconstruction error shown in Fig. 3.

these simulations, we observe that the top 20 basis vectors are sufficient to approximate the likelihood with sufficient accuracy. However, we demonstrated the absolute error between the true and approximated value for 5, 7, 20, 50, 100, and 200 basis vectors, respectively. Fig. 3 shows the probability distribution function (PDF) of the difference between true and estimated likelihood, and the Fig. 4 shows the corresponding cumulative distribution function (CDF). It is clear from these figures that as we increase the number of basis vectors, the error distribution becomes more concentrated around zero as compared to the relatively flattened distribution for smaller basis sizes.

### APPENDIX B: SCALING OF THE SPEED-UP FACTOR WITH WAVEFORM DURATION

The meshfree likelihood evaluation has no dependence on the length of the waveform. On the other hand, in the standard method, a significant time is spent on the waveform generation, followed by the evaluation of the likelihood integral. The latter depends on the length of the waveform. As shown in Table II, the speed-up ratios decrease with higher masses (shorter waveforms).

To verify this, we calculated the likelihood evaluation time using both PyCBC ( $t_{\text{pycbc}}$ ) and meshfree method ( $t_{\text{rbf}}$ ) and also evaluated the waveform generation time ( $t_{\text{wfeval}}$ ) for the four compact binary sources at  $10^4$  points each. As shown in the plot below (Fig. 5), the waveform generation part is the dominant cost in traditional calculations. For a BNS system, it takes about  $\sim 7$  times longer to generate the waveform as compared to the time for evaluating the likelihood integral. As we go towards heavy CBC systems, the speed-up ratio decreases as expected.

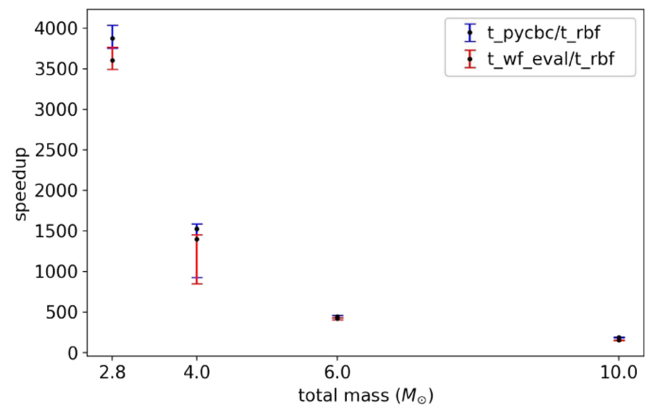


FIG. 5. The relative speed-up between the likelihood evaluation using PyCBC and the proposed meshfree method is shown (blue). Also, The ratio of overall time for generating waveforms (total mass:  $2.8M_{\odot}$ – $10M_{\odot}$ ) using PyCBC and the likelihood evaluation via meshfree scheme has been compared (red). The dominant cost in the traditional likelihood calculation arises from the generation of long duration waveforms.

- [1] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi Gamma-ray Burst Monitor and INTEGRAL Collaborations), Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, *Astrophys. J. Lett.* **848**, L13 (2017).
- [2] L. P. Singer and L. R. Price, Rapid Bayesian position reconstruction for gravitational-wave transients, *Phys. Rev. D* **93**, 024013 (2016).
- [3] D. Finstad and D. A. Brown, Fast parameter estimation of binary mergers for multimessenger follow-up, *Astrophys. J. Lett.* **905**, L9 (2020).
- [4] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **23**, 1 (2020).
- [5] B. Margalit and B. D. Metzger, The multi-messenger matrix: The future of neutron star merger constraints on the nuclear equation of state, *Astrophys. J. Lett.* **880**, L15 (2019).
- [6] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-Time Gravitational Wave Science with Neural Posterior Estimation, *Phys. Rev. Lett.* **127**, 241103 (2021).
- [7] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, *Nat. Phys.* **18**, 112 (2021).
- [8] M. Dax, S. R. Green, J. Gair, M. Pürer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, *Phys. Rev. Lett.* **130**, 171403 (2023).
- [9] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, Accelerated Gravitational Wave Parameter Estimation with Reduced Order Modeling, *Phys. Rev. Lett.* **114**, 071104 (2015).
- [10] H. Qi and V. Raymond, Python-based reduced order quadrature building code for fast gravitational wave inference, *Phys. Rev. D* **104**, 063031 (2021).
- [11] S. Morisaki and V. Raymond, Rapid parameter estimation of gravitational waves from binary neutron star coalescence using focused reduced order quadrature, *Phys. Rev. D* **102**, 104020 (2020).
- [12] A. J. Chua and M. Vallisneri, Learning Bayesian Posteriors with Neural Networks for Gravitational-Wave Inference, *Phys. Rev. Lett.* **124**, 041102 (2020).
- [13] J. Lange, R. O’Shaughnessy, and M. Rizzo, Rapid and accurate parameter inference for coalescing, precessing compact binaries, [arXiv:1805.10457](https://arxiv.org/abs/1805.10457).
- [14] N. J. Cornish, Heterodyned likelihood for rapid gravitational wave parameter inference, *Phys. Rev. D* **104**, 104054 (2021).
- [15] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, [arXiv:1806.08792](https://arxiv.org/abs/1806.08792).
- [16] R. Smith, C. Hanna, I. Mandel, and A. Vecchio, Rapidly evaluating the compact-binary likelihood function via interpolation, *Phys. Rev. D* **90**, 044074 (2014).
- [17] K. Cannon, A. Chapman, C. Hanna, D. Keppel, A. C. Searle, and A. J. Weinstein, Singular value decomposition applied to compact binary coalescence gravitational-wave signals, *Phys. Rev. D* **82**, 044025 (2010).
- [18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (The Johns Hopkins University Press, Baltimore, 2013).
- [19] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The MCMC Hammer, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [20] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* **85**, 122006 (2012).
- [21] J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Anal.* **1**, 833 (2006).
- [22] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, *Pub. Astron. Soc. Aust.* **36**, e010 (2019).
- [23] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cariou, S. Caudill, S. J. Chamberlin, J. D. Creighton, R. Everett *et al.*, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, *Phys. Rev. D* **95**, 042001 (2017).
- [24] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, *Classical Quantum Gravity* **33**, 215004 (2016).
- [25] G. E. Fasshauer, *Meshfree Approximation Methods with Matlab* (World Scientific, Singapore, 2007), [10.1142/6437](https://doi.org/10.1142/6437).
- [26] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, *Phys. Rev. D* **93**, 044007 (2016).
- [27] L. Barsotti, S. Gras, M. Evans, and P. Fritschel, The updated Advanced LIGO design curve, Technical Report No. LIGO-T1800044-v5, LIGO Scientific Collaboration, 2018.
- [28] T. Hines, Python package containing the tools necessary for radial basis function (RBF) applications (2015), <https://github.com/treverhines/RBF.git>.
- [29] J. S. Speagle, DYNesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [30] S. Koposov, J. Speagle, K. Barbary, G. Ashton, E. Bennett, J. Buchner, C. Scheffler, B. Cook, C. Talbot, J. Guillochon *et al.*, [joshspeagle/dynesty: v2.1.0, 10.5281/zenodo.7600689](https://doi.org/10.5281/zenodo.7600689) (2023).
- [31] C. P. Ayanendranath Basu and Hiroyuki Shioya, *Statistical Inference: The Minimum Distance Approach*, 1st ed. (Chapman and Hall/CRC, London, 2011).
- [32] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals, *Publ. Astron. Soc. Pac.* **131**, 024503 (2019).
- [33] L. Pathak, S. Munishwar, A. Reza, and A. S. Sengupta, Prompt sky localization of compact binary sources using meshfree approximation (to be published).



- [34] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, *Phys. Rev. D* **77**, 042001 (2008).
- [35] C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy, Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences, *Phys. Rev. D* **92**, 023002 (2015).
- [36] G. M. Manca and M. Vallisneri, Cover art: Issues in the metric-guided and metric-less placement of random and stochastic template banks, *Phys. Rev. D* **81**, 024004 (2010).