

# Accelerating global parameter estimation of gravitational waves from Galactic binaries using a genetic algorithm and GPUs

Stefan H. Strub<sup>✉,\*</sup>, Luigi Ferraioli<sup>✉</sup>, Cédric Schmelzbach<sup>✉</sup>, Simon C. Stähler<sup>✉</sup>, and Domenico Giardini<sup>✉</sup>

*Institute of Geophysics, ETH Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland*



(Received 7 July 2023; accepted 5 October 2023; published 13 November 2023)

The Laser Interferometer Space Antenna (LISA) is a planned space-based gravitational wave telescope with the goal of measuring gravitational waves in the millihertz frequency band, which is dominated by millions of Galactic binaries. While some of these binaries produce signals that are loud enough to stand out and be extracted, most of them blur into a confusion foreground. Current methods for analyzing the full frequency band recorded by LISA to extract as many Galactic binaries as possible and to obtain Bayesian posterior distributions for each of the signals are computationally expensive. We introduce a new approach to accelerate the extraction of the best fitting solutions for Galactic binaries across the entire frequency band from data with multiple overlapping signals. Furthermore, we use these best fitting solutions to omit the burn-in stage of a Markov chain Monte Carlo method and to take full advantage of graphics processing unit (GPU)-accelerated signal simulation, allowing us to compute posterior distributions in 2 s per signal on a laptop-grade GPU.

DOI: [10.1103/PhysRevD.108.103018](https://doi.org/10.1103/PhysRevD.108.103018)

## I. INTRODUCTION

The detection of gravitational waves (GWs) by the LIGO detector in 2015 marked a significant breakthrough in astrophysics [1]. This achievement spurred the development of the Laser Interferometer Space Antenna (LISA), a space-based interferometric system capable of detecting low-frequency GWs in the [0.1, 100] mHz range, free from terrestrial seismic and anthropogenic noise sources [2]. LISA is an L-class mission of the European Space Agency and currently set for launch in the 2030s.

The primary sources in the LISA frequency band are tens of millions of Galactic binaries (GBs) emitting quasimonochromatic gravitational waves. These sources are far from merging, allowing for their gravitational waves to be continuously measured during LISA's nominal 4 year operational time [2]. It is estimated that tens of thousands of these overlapping signals are resolvable by an experiment of LISA's arm length, resolution and measurement duration, while the rest blurs into a galactic foreground noise. Accurately estimating the parameters of GBs provides valuable information for studying the dynamical evolution of binaries [3–7].

Several methods have been proposed for extracting GB signals, including maximum likelihood estimate (MLE) [8–10] and Bayesian approaches. MLE methods are used to find the best matching simulated signal to the data, while Bayesian methods provide a posterior distribution that describes the uncertainty of the source parameters. The

most successful Bayesian approaches are Markov chain Monte Carlo (MCMC)-based methods, such as blocked annealed Metropolis-Hastings [11–13], an MCMC algorithm with simulated annealing, or the reversible jump Markov chain Monte Carlo [14,15] method, which allows for varying parameter dimensions and thus variable numbers of GBs to construct the posterior distribution.

In our previous work [16], we demonstrated that signal extraction can be divided into two parts for both isolated and overlapping signals in the frequency domain. The first part involves optimizing the GB parameters in order to achieve the best fit between the simulated signal and the available data. In the second part, Gaussian process regression [17] is used to model the log-likelihood function, which allows for the computation of the posterior distribution without the need to simulate the GW signal for each sample. In this paper, we extend the work to analyze the full Galactic signal population from a simulated LISA data stream.

Furthermore, with recent advances in simulating a GB signal using graphics processing units (GPUs), we swapped the Gaussian process regression modeling with directly computing the log-likelihood function using a GPU [18]. For sampling, we use a Metropolis-Hastings algorithm with a proposal distribution independent of the current state of the chain. Therefore, we can make full use of calculating the log-likelihood for 10 000 signals in parallel and build the Markov chain in the next step. That way we are able to compute the posterior distribution of a single signal within only 1.8 s on a late 2018-released Quadro RTX 4000 Mobile GPU built inside a laptop. We demonstrate the

\*stefan.strub@erdw.ethz.ch

benefit of such a speedup by solving for the GB of the LISA data challenge (LDC)1–4, part of LDC1, which is also called Radler [19]. This challenge encompasses a dataset containing instrument noise as well as 26 million GB signals. Additionally, the pipeline has also been tested on LDC2a, called Sangria, where the injected Massive Black Hole Binaries (MBHBs) are subtracted, resulting in a dataset comprising 30 million GBs along with instrument noise [19].

In Sec. II we introduce Bayesian parameter estimation, and Sec. III provides a detailed description of the new pipeline. In Sec. IV the performance of the pipeline is showcased through its successful handling of the LISA data challenges LDC1–4 and LDC2a. Lastly, Sec. V discusses the performance of the pipeline and the potential for further pipeline development.

## II. BAYESIAN FORMULATION FOR SIGNAL EXTRACTION

Gravitational waves are ripples in the fabric of spacetime caused by the acceleration of massive objects, such as merging black holes, neutron stars and white dwarfs. LISA is a planned space-based mission designed to detect these elusive signals with unprecedented precision. However, the expected LISA data, denoted as  $d(t)$ , will be contaminated by instrument noise and unresolved signals, making the extraction of the underlying gravitational wave signal, denoted as  $s(t, \theta)$ , a challenging task. To tackle this, Bayesian inference and data analysis techniques provide a powerful framework. For convenience, we will omit the notation for dependence on  $t$  for the data  $d$  and the signals  $s(\theta)$  in the following.

In Bayesian inference, we aim to infer the probability distribution of the parameters  $\theta$  describing the gravitational wave signal  $s(\theta)$  given the observed data  $d$ . This is done using Bayes' theorem, which relates the posterior distribution  $p(\theta|d)$ , the prior distribution  $p(\theta)$ , the likelihood  $p(d|\theta)$ , and the model evidence  $p(d)$  as follows:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}. \quad (1)$$

The posterior distribution  $p(\theta|d)$  represents the updated probability distribution of the parameters  $\theta$  after taking into account the measured data  $d$ . The prior distribution  $p(\theta)$  incorporates any prior knowledge or assumptions about the parameters. The model evidence  $p(d)$  is a normalization factor that ensures the posterior distribution integrates to unity, and it is independent of  $\theta$  and, hence, does not affect the relative probabilities.

In GW data analysis, the likelihood  $p(d|\theta)$  quantifies the probability of measuring the data stream  $d$  given the parameters  $\theta$  of the gravitational wave signal. The log-likelihood is commonly used due to its mathematical convenience and is defined as

$$\log p(d|\theta) = -\frac{1}{2} \langle d - s(\theta) | d - s(\theta) \rangle, \quad (2)$$

where  $\langle x(t)|y(t) \rangle$  is the scalar product between two time-domain signals  $x(t)$  and  $y(t)$ , and it is defined as

$$\langle x(t)|y(t) \rangle = 4\mathcal{R} \left( \int_0^\infty \frac{\tilde{x}(f)\tilde{y}^*(f)}{S(f)} df \right). \quad (3)$$

Here,  $\tilde{x}(f)$  marks the Fourier transform of  $x(t)$ , and  $S(f)$  is the one-sided power spectral density of the noise, which characterizes the noise properties of the LISA detector. The noise is estimated and constantly updated during the search. The noise estimate for GB analysis is discussed in Secs. III C and III F.

To eliminate the laser noise in the LISA arms' laser measurements, time-delay interferometry (TDI) will be employed, which combines the measurements into three observables:  $X$ ,  $Y$ , and  $Z$  [20–24]. Consequently, the data  $d$  and the signal  $s(\theta)$  consist of TDI responses with multiple channels, and we write the inner product as the following sum:

$$\langle d - s(\theta) | d - s(\theta) \rangle = \sum_{\alpha \in \mathcal{M}} \langle d_\alpha - s_\alpha(\theta) | d_\alpha - s_\alpha(\theta) \rangle. \quad (4)$$

Here,  $\mathcal{M} = X, Y, Z$  represents the default TDI setting, or  $\mathcal{M} = A, E, T$  where

$$\begin{aligned} A &= \frac{1}{\sqrt{2}}(Z - X), \\ E &= \frac{1}{\sqrt{6}}(X - 2Y + Z), \\ T &= \frac{1}{\sqrt{3}}(X + Y + Z) \end{aligned} \quad (5)$$

are uncorrelated with respect to instrument noise [25]. In this work we utilize  $A$ ,  $E$ , and  $T$ . However, to save computational time, we consider only  $A$  and  $E$  for signals with frequencies  $f < f_*/2 = 1/(4\pi L) \approx 9.55$  mHz, as the contribution of the gravitational wave response for  $T$  is suppressed [14]. By setting the threshold at half the transfer frequency  $f_*$ , we adopt a more conservative approach than including the  $T$  channel only for signals with  $f > f_*$ .

## III. EXTRACTING GALACTIC BINARY SIGNALS IN THE FULL LISA FREQUENCY BAND

The simulation of a GW from a GB system involves eight parameters denoted as  $\theta = \{\mathcal{A}, \lambda, \beta, f, \dot{f}, i, \phi_0, \psi\}$  [26]. These parameters are utilized to model the GW signal, where  $\mathcal{A}$  represents the amplitude and  $\lambda$  and  $\beta$  correspond to the sky coordinates in terms of ecliptic longitude and ecliptic latitude, respectively. The parameter  $f$  represents the frequency of the GW,  $\dot{f}$  denotes the

first-order frequency derivative,  $\iota$  represents the inclination angle,  $\phi_0$  represents the initial phase, and  $\psi$  corresponds to the polarization angle. In this study, we consider only the first-order frequency derivative and neglect higher-order frequency derivatives.

To obtain the MLE we can maximize the signal-to-noise ratio (S/N) defined as

$$\rho = \frac{\langle d|s(\theta') \rangle}{\sqrt{\langle s(\theta')|s(\theta') \rangle}} = \frac{\langle d|s(\theta) \rangle}{\sqrt{\langle s(\theta)|s(\theta) \rangle}}, \quad (6)$$

which is independent of  $\mathcal{A}$  with  $\theta' = \theta \setminus \{\mathcal{A}\}$ , and obtain

$$\mathcal{A}_{\max} = \frac{\langle d|s(\theta') \rangle}{\langle s(\theta')|s(\theta') \rangle} \quad (7)$$

analytically [16].

### A. Frequency segments

Because fitting  $>10000$  signals globally is a currently untractable problem, we split the data into small segments in the frequency domain. In order to have a few signals in one segment while keeping the number of segments small for stability and efficiency, we determine the segment size to be double the size of the broadest signal expected for each frequency segment  $B_{\text{segment}}(f) = 2B_{\text{max}}(f)$ . The width of a signal in the frequency domain is influenced by various factors contributing to signal broadening. These factors include the frequency change of the source itself, LISA's orbital motion around the Sun, and LISA's cartwheel motion.

To obtain the widest expected broadening we multiply the highest-frequency derivative with the observation time  $B_F = \dot{f}_{\text{max}} T_{\text{obs}}$ , where  $\dot{f}_{\text{max}}$  is determined by [14]

$$\dot{f} = \frac{96}{5} \pi^{8/3} \mathcal{M}_c^{5/3} f^{11/3}, \quad (8)$$

where  $\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$  is the chirp mass and  $f$  the frequency. For  $\dot{f}_{\text{max}}$  the masses of the binary are set to the Chandrasekhar limit  $m_1 = m_2 = 1.4M_{\odot}$  [27].

LISA's orbit around the Sun and cartwheel motion smear the signal by  $B_O = 10^{-4}f$  and  $B_C = 4 \cdot \frac{1}{\text{yr}^3}$ , respectively, due to Doppler shift [28]. Since the smearing can increase or decrease the frequency, the resulting bandwidth is  $2B_O$  and  $2B_C$ , respectively. As a result, the broadest signal expected has a width of  $B_{\text{max}} = B_F + 2B_O + 2B_C$  which is shown in Fig. 1.

In Algorithm 1, we outline the procedure for generating the list of frequency segments  $B_{\text{search}}$  for a given global frequency interval. The lower bound of the frequency range,  $f_{\text{min}} = 0.3$  mHz, is chosen based on the absence of expected detectable GBs at frequencies lower than 0.3 mHz. The upper bound  $f_{\text{max}} = f_{\text{Nyquist}}$  is determined

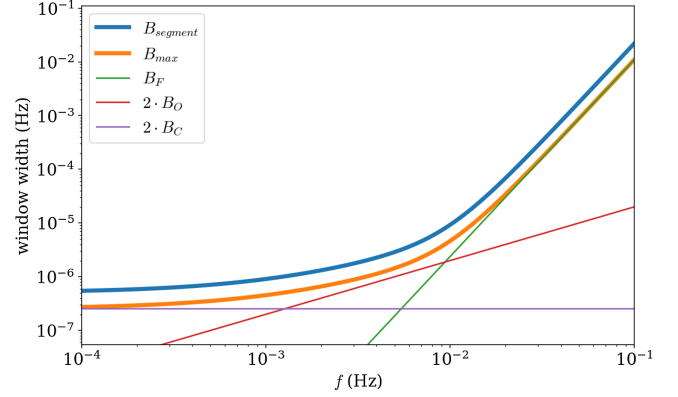


FIG. 1. Frequency segment widths to analyze GBs for  $T_{\text{obs}} = 2$  yr.

by the sampling frequency where the Nyquist criterion states that the sampling frequency should be at least twice the maximum frequency of interest in order to accurately capture the signal [29].

Algorithm 1. Generating the list of frequency segments  $B_{\text{search}}$  for searching GBs within the given ranges.

---

**Function** *segmenting*( $f_{\text{min}}, f_{\text{max}}$ )  
 $B_{\text{search}} \leftarrow \{ \}$   
 $f \leftarrow f_{\text{min}}$   
**while**  $f < f_{\text{max}}$  **do**  
     $f_{\text{next}} \leftarrow f + 2B_{\text{max}}(f)$   
     $B_{\text{search}} \text{ append } [f, f_{\text{next}}]$   
     $f \leftarrow f_{\text{next}}$   
**return**  $B_{\text{search}}$

---

### B. Prior

In Table I we list the prior distribution  $\Theta$  for all parameters. The frequency boundary is the padded frequency segment of interest  $f_{\text{segment}} \in B_{\text{search}}$ . The padding is half of the broadest signal expected  $f_{\text{padding}} = (\max(f_{\text{segment}}) - \min(f_{\text{segment}}))/4$  in case a signal is at the boundary of two neighboring segments like for example the yellow and gray signals at 4.226 mHz in Fig. 2. For the upper bound of  $f$  we use  $\dot{f}_{\text{max}}$  determined by (8). Since we

TABLE I. Boundaries of the prior distribution  $\Theta$ .

Parameter	Lower bound	Upper bound
$\sin \beta$	-1	1
$\lambda$	$-\pi$	$\pi$
$f$	$\min(f_{\text{segment}}) - f_{\text{padding}}$	$\max(f_{\text{segment}}) + f_{\text{padding}}$
$\dot{f}$	$-5 \times 10^{-6} f^{13/3}$	$1.02 \times 10^{-6} f^{11/3}$
$\log \mathcal{A}$	$\log \mathcal{A}(\rho = 7)$	$\log \mathcal{A}(\rho = 1000)$
$\cos \iota$	-1	1
$\phi_0$	0	$2\pi$
$\psi$	0	$\pi$

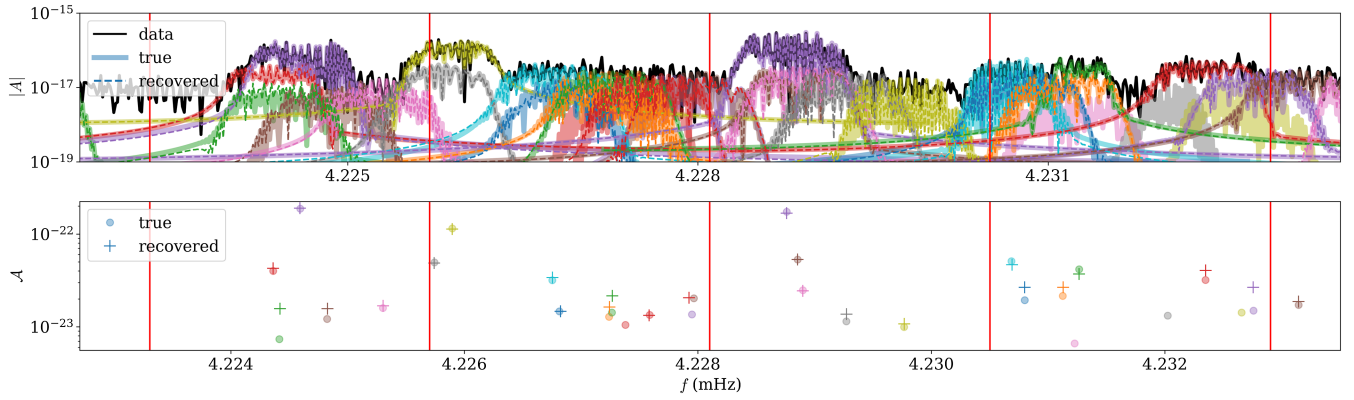


FIG. 2. Displayed are the data, injected signals, and recovered signals of the Radler data challenge with  $T_{\text{obs}} = 2$  yr. The red lines are the boundaries of four adjacent frequency segments. The first plot illustrates the absolute value of the  $A$  TDI channel, while the second plot depicts the amplitude  $\mathcal{A}$  across the frequency spectrum. The red lines mark the boundaries of the frequency segments. The plot is extended to the left and right by the padding of the segments at the borders.

search for detached and interacting binaries the lower bound  $\dot{f}$  is negative and is the same as in [14]. The amplitude boundary is determined by a lower and upper bound S/N and is related to the amplitude by [14]

$$\mathcal{A}(\rho) = 2\rho \left( \frac{S(f)}{T_{\text{obs}} \sin^2(f/f_*)} \right)^{1/2}. \quad (9)$$

### C. Noise estimate within a frequency segment

For estimating the maximum likelihood of the GBs within a frequency segment the noise is estimated individually for each segment by calculating the periodogram [30,31]

$$S_A(f) = \frac{2|A(f)|^2}{Nf_{\text{sample}}} \quad (10)$$

for each frequency window including the padding as determined for the prior listed in Table I.  $N$  marks the number of bins within the padded window and  $f_{\text{sample}}$  represents the sampling frequency of the data  $d$ . In order to reduce the influence of loud signals within the window itself, the median of  $S_A(f)$  is taken as the constant estimate for the full padded frequency segment. This brings a dynamic noise estimate during the search of signals which is updated after each found signal is subtracted from the data. The estimate for other TDI variables  $E$ ,  $T$  is analog to the estimate of  $A$ .

### D. Galactic binary search algorithm within a frequency segment

In Algorithm 2 we present the GB search algorithm for given data  $d_{\text{analyze}}$  to analyze on a given frequency segment  $f_{\text{segment}} \in f_{\text{search}}$ , which outputs a list  $\tilde{\theta}_{\text{in}} = \{\theta_{\text{MLE},1}, \theta_{\text{MLE},2}, \dots\}$  of GB parameters within the unpadded

$f_{\text{segment}}$ . Furthermore,  $n_{\text{signals}}$  is the maximum number of signals per segment.

Algorithm 2. The GB search algorithm within a frequency segment  $f_{\text{segment}}$ .

---



---

```

Function local_GB_search( $f_{\text{segment}}, n_{\text{signals}}, d_{\text{analyze}}$ )
 $\tilde{\theta}_{\text{found}} \leftarrow \{\}$ 
 $\tilde{\theta}_{\text{in}} \leftarrow \{\}$ 
 $\tilde{\theta}_{\text{out}} \leftarrow \{\}$ 
 $d_{\text{residual}} \leftarrow d_{\text{analyze}}$ 
for  $i$  in  $\{1, 2, \dots, n_{\text{signals}}\}$  do
     $\tilde{\theta}'_{\text{MLEs}} \leftarrow \{\}$ 
    for  $j$  in  $\{1, 2, \dots, n_{\text{searches}}\}$  do
         $\theta'_{\text{init}}$  randomly drawn from prior
         $\theta'_{\text{MLE}} \leftarrow \text{argmax}_{\theta'} \rho(\theta', d_{\text{residual}})$  using DE with  $\theta'_{\text{init}}$ 
        if  $\rho(\theta'_{\text{MLE}}, d_{\text{analyze}}) \leq \rho_{\text{threshold}} - 2$  and  $j = 1$  do
            return  $\tilde{\theta}_{\text{in}}$ 
         $\tilde{\theta}'_{\text{MLEs}} \leftarrow \tilde{\theta}'_{\text{MLEs}} \cup \{\theta'_{\text{MLE}}\}$ 
    end for
     $\theta'_{\text{MLE}} \leftarrow \text{argmax}_{\theta' \in \tilde{\theta}'_{\text{MLEs}}} \rho(\theta', d_{\text{residual}})$ 
    if  $\rho(\theta'_{\text{MLE}}, d_{\text{analyze}}) \leq \rho_{\text{threshold}}$  do
        return  $\tilde{\theta}_{\text{in}}$ 
    Compute  $\mathcal{A}_{\text{max}}$  according to (7) with  $\theta'_{\text{MLE}}$ 
     $\theta_{\text{MLE}} \leftarrow \theta'_{\text{MLE}} \cup \{\mathcal{A}_{\text{max}}\}$ 
    if  $\theta'_{\text{MLE}}$  in unpadded  $f_{\text{segment}}$  do
         $\tilde{\theta}_{\text{in}} \leftarrow \tilde{\theta}_{\text{in}} \cup \{\theta_{\text{MLE}}\}$ 
    else do
         $\tilde{\theta}_{\text{out}} \leftarrow \tilde{\theta}_{\text{out}} \cup \{\theta_{\text{MLE}}\}$ 
         $d_{\text{residual}} \leftarrow d_{\text{analyze}} - \sum_{\theta \in \tilde{\theta}_{\text{out}}} s(\theta)$ 
         $\tilde{\theta}_{\text{in}} \leftarrow \text{argmax}_{\tilde{\theta}} \rho(\tilde{\theta}, r)$  using SLSP with  $\tilde{\theta}_{\text{in}}$  as start
         $\tilde{\theta}_{\text{found}} \leftarrow \tilde{\theta}_{\text{in}} \cup \tilde{\theta}_{\text{out}}$ 
         $d_{\text{residual}} \leftarrow d_{\text{analyze}} - \sum_{\theta \in \tilde{\theta}_{\text{found}}} s(\theta)$ 
    end for
return  $\tilde{\theta}_{\text{in}}$ 
    
```

---



---

To save computational time, we limit the integral of the scalar product (3) to the padded frequency segment. To obtain the MLE we use the differential evolution (DE) [32]

algorithm and for the global optimization of all found signals within the unpadded region  $\tilde{\theta}_{\text{in}}$  we use the sequential least squares programming (SLSP) method [33]. Both methods are part of the SciPy library [34]. The pipeline is set to search  $n_{\text{searches}} = 3$  times for the same signal with varying initial parameters  $\theta'_{\text{init}}$  in case the search algorithm gets stuck at a local optimum.

Global optimization within a segment occasionally encounters a situation where two signals exhibit a negative correlation, leading to destructive interference. The correlation between these signals is quantified by the following equation:

$$\mathcal{O}(s(\theta_1), s(\theta_2)) = \frac{\langle s(\theta_1), s(\theta_2) \rangle}{\sqrt{\langle s(\theta_1), s(\theta_1) \rangle \langle s(\theta_2), s(\theta_2) \rangle}}. \quad (11)$$

When such interfering signals arise, they may appear to fit the data, but they are unlikely to represent the correct solution. If the correlation between two recovered signals falls below  $-0.7$ , the segment is subjected to reanalysis using a different seed.

Furthermore, we generalize the S/N to multiple signals  $\tilde{\theta} = \{\theta_1, \theta_2, \dots\}$ :

$$\rho = \frac{\langle d | \sum_{\theta \in \tilde{\theta}} s(\theta) \rangle}{\sqrt{\langle \sum_{\theta \in \tilde{\theta}} s(\theta) | \sum_{\theta \in \tilde{\theta}} s(\theta) \rangle}}. \quad (12)$$

### E. Global GB search pipeline

Segments that are not direct neighbors to each other can be analyzed in parallel as presented in Algorithm 3. Therefore we enumerate all segments  $B_{\text{search}}$  ordered by frequency and divide them into two groups of even and odd segments:

$$B_{\text{even}} := \{B_{\text{search},i} : i \in \{1, 2, \dots, N_{\text{segments}}\} \text{ and } i \text{ is even}\},$$

$$B_{\text{odd}} := \{B_{\text{search},i} : i \in \{1, 2, \dots, N_{\text{segments}}\} \text{ and } i \text{ is odd}\},$$

where  $N_{\text{segments}}$  is the number of segments in the list  $B$ . In Table II we list the parameters: the list of frequency segments  $f_{\text{search}}$ , max number of signals per segment  $n_{\text{signals}}$  and the data to analyze  $d_{\text{analyze}}$ . The analysis is conducted in three sequential runs, starting from the top line of the table, in order

TABLE II. Inputs and outputs of the search pipeline *global\_GB\_search* across all frequency segments  $B$  for given data  $d$ .

Run	$f_{\text{search}}$	$n_{\text{signals}}$	$d_{\text{analyze}}$	Output
1	$B_{\text{even}}$	3	$d$	$\tilde{\theta}_{\text{even}}$
2	$B_{\text{odd}}$	10	$d - \sum_{\theta \in \tilde{\theta}_{\text{even}}} s(\theta)$	$\tilde{\theta}_{\text{odd}}$
3	$B_{\text{even}}$	10	$d - \sum_{\theta \in \tilde{\theta}_{\text{odd}}} s(\theta)$	$\tilde{\theta}_{\text{even}}$

to cover all frequency segments within  $B_{\text{search}}$ . In the first run, we analyze all even segments  $B_{\text{even}}$ , allowing for a maximum of  $n_{\text{signals}} = 3$  signals per segment. We assume that there are not more than three strong signals spilling into the padded regions for each segment.

Algorithm 3. The search algorithm for multiple frequency segments  $f_{\text{search}}$ .

---

```

Function global_GB_search( $f_{\text{search}}, n_{\text{signals}}, d_{\text{analyze}}$ )
 $\tilde{\theta} \leftarrow \{\}$ 
for all  $f_{\text{segment}}$  in  $f_{\text{search}}$  do in parallel
     $\tilde{\theta} \leftarrow \tilde{\theta} \cup \text{local\_GB\_search}(f_{\text{segment}}, n_{\text{signals}}, d_{\text{analyze}})$ 
end for
return  $\tilde{\theta}$ 

```

---

Next, we proceed to analyze the odd segments in a similar manner. The found signals in the odd segments, denoted as  $\tilde{\theta}_{\text{odd}}$ , are subtracted from the original data  $d$ . Finally, we repeat the analysis of the even segments, now free from the influence of neighboring signals located in the neighboring odd segments. By subtracting the signals found in the odd segments and reanalyzing the even segments, we ensure that each segment of  $B_{\text{search}}$  is analyzed independently without being affected by signals of neighboring signals.

The even segments where no signals in neighboring segments were detected and less than three signals were found are not analyzed a second time, because the subtraction of the signals in odd windows did not influence these even segments and there is no need to repeat the search. For these segments, the found signals of the first even segments analysis are directly used for the catalog.

The LISA data will be a time-evolving dataset with new data being constantly added. Therefore the found signals of previous runs can be used to speed up the analysis where  $\tilde{\theta}_{\text{initial}}$  if  $j = 1$  in Algorithm 2 is set to a signal within that frequency segments found in the previous run. Especially for signals  $f > 10$  mHz and  $T_{\text{obs}} > 1$  yr the success rate of *local\_GB\_search* becomes small if  $\tilde{\theta}_{\text{initial}}$  is randomly drawn from the prior. It is advantageous to use the found signals of a previous shorter dataset analysis, for example,  $T_{\text{obs}} = 6$  months, as the initial value of the search algorithm.

The global solution is then  $\tilde{\theta}_{\text{recovered}} = \tilde{\theta}_{\text{even}} \cup \tilde{\theta}_{\text{odd}}$ , where  $\tilde{\theta}_{\text{even}}$  is the solution of the third run. In Fig. 2 we show the solution  $\tilde{\theta}_{\text{recovered}}$  for four neighboring segments at a region with multiple detectable and overlapping signals. We demonstrate with the pipeline a successful recovery rate of 25 out of 30 injected GBs. Among the 25 recovered signals, 24 of them correspond to individual injected signals, indicating a high level of accuracy in the recovery process. In addition, it is worth noting that the recovered signals at  $f = 4.22$  mHz is a composite of two injected signals. However, the

remaining unrecovered signal is characterized by low amplitudes  $\mathcal{A}$ .

### F. Global noise estimate

For the global noise estimate, we subtract each recovered signal  $\hat{\theta}_{\text{recovered}}$  from the data:

$$d_{\text{residual}} = d - \sum_{\theta \in \hat{\theta}_{\text{recovered}}} s(\theta), \quad (13)$$

where  $s(\theta)$  represents the signal corresponding to each MLE  $\theta$ . Furthermore, we proceed to estimate a smooth noise curve, denoted as  $S_{A,\text{welch}}(f)$ , across the entire frequency domain. This estimation is performed by applying Welch's method, utilizing 500 windows and a Hann window function [35]. Next, we address the remaining outlier peaks, mainly of unresolved signals, by implementing a smoothing procedure. We define a frequency window of 30 bins and adjust any values above the window's median to be twice the median value. This process is repeated by shifting the window by 15 frequency bins until the entire power spectral density (PSD) is smoothed. The result is denoted as  $S_{A,\text{median}}(f)$ .

To further enhance the smoothing effect, we utilize the Savitzky-Golay filter [36]. The filter is configured with an order of 1, and we apply two different window lengths depending on the frequency range. For observations with  $T_{\text{obs}}$  equal to either 1 or 2 yr, frequencies below 0.8 mHz are smoothed using a window length of 10, while frequencies above 0.8 mHz are smoothed using a window length of 70. In the case of  $T_{\text{obs}} = 0.5$  yr, frequencies below

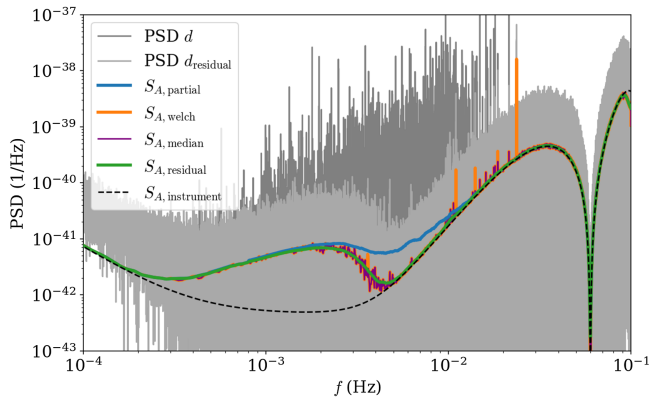


FIG. 3. Noise estimates and PSD of the TDI A channel of the 1 yr Sangria dataset.  $S_{A,\text{instrument}}$  is the noise PSD used for creating the data. The difference between  $S_{A,\text{residual}}$  and the true PSD between 0.2 and 5 mHz is due to the unresolved GBs which can be seen as red crosses in Fig. 7. It is expected that most GBs in that frequency range are unresolvable and therefore merge into the galactic foreground noise.

0.8 mHz employ a window length of 10, and frequencies above 0.8 mHz are smoothed using a window length of 50.

Finally, to obtain a PSD estimate for each desired frequency bin, we spline interpolate the smoothed PSD, resulting in our estimate of the residual noise curve denoted as  $S_{A,\text{residual}}(f)$ .

The noise estimates, depicted in Fig. 3, exhibit a strong agreement with the instrument noise  $S_{A,\text{instrument}}(f)$ , except for the frequency range between 0.2 and 5 mHz. In this range, the unresolved background signals (GBs) merge into the galactic foreground noise, leading to deviations in the noise estimate. The noise of the other TDI channels  $E$  and  $T$  are computed the same way. For determining the posterior distribution we additionally compute  $S_{A,\text{partial}}(f)$  which is described in the next subsection.

### G. GPU-accelerated posterior distribution derivation

In order to derive the posterior distribution, we employ the Metropolis-Hastings Monte Carlo algorithm [37,38]. This algorithm suggests new parameters  $\theta_p$  based on a proposal distribution  $g(\theta_p|\theta_c)$ , which generally depends on the current state of the chain  $\theta_c$ . The proposed parameters are then accepted with probability

$$P(\theta_p, \theta_c) = \min \left( 1, \left[ \frac{p(d|\theta_p) g(\theta_c|\theta_p)}{p(d|\theta_c) g(\theta_p|\theta_c)} \right]^{\frac{1}{T}} \right), \quad (14)$$

where  $T$  is the temperature for simulated annealing.

Previously, [16] demonstrated that the MLE can be effectively utilized to accelerate the computation of the posterior distribution. The posterior distribution tends to be concentrated within a relatively compact region of the parameter space. As a result, it becomes unnecessary to sample beyond specific parameter space boundaries when employing MCMC methods to estimate the posterior. By identifying the reduced parameter space  $\Theta_{\text{reduced}}$  where the posterior is concentrated, we can skip the burn-in phase typically required in MCMC sampling. Moreover, this approach allows for a proposal distribution  $g(\theta_p) = g(\theta_p|\theta_c)$  that is independent of the current state of the chain  $\theta_c$ . This is achieved by randomly drawing samples within  $\Theta_{\text{reduced}}$ . The independence from the chain's state enables the parallel computation of the log-likelihood for all samples in the first step, followed by the construction of the chain during the second step, where if the proposed sample  $\theta_p$  is rejected the chain stays at the current sample  $\theta_c$  as described in Algorithm 4. This approach leverages the computational power of a GPU to rapidly compute the log-likelihood of 10 000 samples in parallel, facilitating a more efficient and rapid estimation of the posterior distribution.

Algorithm 4. The GPU-accelerated posterior distribution algorithm.

---

```

Function posterior( $\theta_{\text{MLE}}, d_{\text{posterior}}, \Theta_{\text{reduced}}, \tilde{\mathcal{T}}, n_{\text{samples}}$ )
 $\Theta_{\text{sample}} \leftarrow \Theta_{\text{reduced}}$ 
for  $\mathcal{T}$  in  $\tilde{\mathcal{T}}$  do
   $\tilde{\theta}_{\text{posterior}} \leftarrow \{\}$ 
   $\tilde{L} \leftarrow \{\}$ 
   $\tilde{\theta}_{\text{samples}} \leftarrow n_{\text{samples}}$  randomly drawn from  $\Theta_{\text{sample}}$ 
  for all  $\theta$  in  $\tilde{\theta}_{\text{samples}}$  do in parallel on GPU
     $\tilde{L} \leftarrow \tilde{L} \cup \{p(d_{\text{posterior}}|\theta)\}$ 
  end for
   $\theta_c \leftarrow \tilde{\theta}_1$ 
   $L_{\text{current}} \leftarrow \tilde{L}_1$ 
  for  $i$  in  $\{2, 3, \dots, n_{\text{samples}}\}$  do
     $\alpha \leftarrow \min(1, \frac{L_i - g(\theta_c)}{L_{\text{current}} - g(\theta_c)})^{\frac{1}{\mathcal{T}}}$ 
    with probability  $\alpha$  do
       $\theta_c = \tilde{\theta}_i$ 
       $L_{\text{current}} = \tilde{L}_i$ 
       $\tilde{\theta}_{\text{posterior}} \leftarrow \tilde{\theta}_{\text{posterior}} \cup \{\theta_c\}$ 
    end for
   $\Theta_{\text{sample}} \leftarrow \text{KDE}(\tilde{\theta}_{\text{posterior}})$ 
end for
return  $\tilde{\theta}_{\text{posterior}}$ 

```

---

To establish the reduced parameter space  $\Sigma_{\text{reduced}}$  we use the inverse of the Fisher information matrix (FIM)

$$F_{ij} = \langle \partial_i p(d|\theta_{\text{MLE}}) | \partial_j p(d|\theta_{\text{MLE}}) \rangle, \quad (15)$$

where  $\partial_i$  denotes the partial derivative with respect to the  $i$ th component of the parameter vector  $\theta$ . To compute the derivatives of the FIM, the second-order forward finite difference method is employed with a step size of  $10^{-9}$  times the search space determined by the prior distribution  $\Theta$ . The estimated uncertainty vector is  $\sigma = \sqrt{\text{diag}(F^{-1})}$ .

In our investigation, we set the volume of the parameter space to  $\Theta_{\text{reduced}} = [\theta_{\text{MLE}} - 4\sigma, \theta_{\text{MLE}} + 4\sigma]$ . It is sufficient to set the boundary for the frequency parameter to  $\Theta_{\text{reduced}}^f = [\theta_{\text{MLE}}^f - \sigma_f, \theta_{\text{MLE}}^f + \sigma_f]$ , where  $\sigma_f$  denotes the estimated uncertainty of the frequency. The frequency derivative parameter space is not reduced and spans the full prior  $\Theta_{\text{reduced}}^f = \Theta^f$ . Given the degeneracy of the polarization and initial phase, the posterior distribution is not concentrated in a single mode but instead spreads across multiple modes of substantial extent. Consequently, we simplify the treatment of the polarization and initial phase distributions by narrowing their search spaces. Hence, we set  $\Theta^\psi = [\theta_{\text{MLE}}^\psi - \frac{\pi}{1000}, \theta_{\text{MLE}}^\psi + \frac{\pi}{1000}]$  and  $\Theta^{\phi_0} = [\theta_{\text{MLE}}^{\phi_0} - \frac{2\pi}{1000}, \theta_{\text{MLE}}^{\phi_0} + \frac{2\pi}{1000}]$ .

Simulated annealing is useful to further speed up the computation of the posterior. As a start, we use uniform sampling in the reduced parameter space  $\Theta_{\text{reduced}}$  as the proposal distribution with a high temperature. Next, we

utilize the obtained posterior distribution as the new proposal distribution by employing multivariate kernel density estimation (KDE) techniques [39,40]. To address the challenge of high-dimensional KDE computations, we group parameters into two-dimensional parameter pairs. Specifically, we group  $\mathcal{A} - \iota, \lambda - \beta$ , and  $f - \dot{f}$  together and perform KDE on each pair. This allows us to overcome the computational limitations associated with kernel density estimation involving four or more parameters. By gradually lowering the temperature  $\mathcal{T}$  during the simulated annealing process, we can achieve more refined and accurate estimations of the posterior distribution while maintaining computational efficiency.

The results presented in the next Sec. IV are created with six different temperatures  $\mathcal{T} = \{15, 10, 5, 3, 2, 1\}$  and constant  $n_{\text{samples}} = 10000$ . The number of samples  $n_{\text{samples}}$  could be changed for each temperature.

The algorithm to compute the posterior distribution for a single signal  $\theta_{\text{MLE}} \in \tilde{\theta}_{\text{recovered}}$  is presented in Algorithm 4. The computation of the likelihood  $p(d_{\text{posterior}}|\theta)$  on the GPU [41–43] and can be found at [44]. The data for the input is

$$d_{\text{posterior}} = d - \sum_{\theta \in \tilde{\theta}_{\text{recovered}}} s(\theta) + s(\theta_{\text{MLE}}). \quad (16)$$

The reduced parameter space  $\Theta_{\text{reduced}}$  is determined with  $\theta_{\text{MLE}}$  and  $d_{\text{posterior}}$  as described above.

The resulting posterior distribution is the posterior given the overlapping MLEs  $\tilde{\theta}_{\text{overlap}} \subset \tilde{\theta}_{\text{global}}$ :

$$p(\theta_{\text{MLE}}|d_{\text{posterior}}) = p(\theta|d, \tilde{\theta}_{\text{overlap}}), \quad (17)$$

which has a narrower posterior distribution than the marginalized posterior

$$p(\theta_{\text{MLE}}|d) = \int p(\theta, \tilde{\theta}_{\text{overlap}}|d) p(\tilde{\theta}_{\text{overlap}}) d\tilde{\theta}_{\text{overlap}}. \quad (18)$$

Overlapping signals lead to a joint posterior distribution. To approximate the marginalized posterior for such cases, one approach is to increase the estimated noise by computing the noise of the partial residual  $S_{A,\text{partial}}(f)$ . This is achieved by subtracting the found signals only partially from the original data, leaving some residual signal components in the data:

$$d_{\text{partial}} = d - s_{\text{partial}} \sum_{\theta \in \tilde{\theta}_{\text{recovered}}} s(\theta). \quad (19)$$

where  $s_{\text{partial}} \in [0, 1]$  is a scaling factor which we set to  $s_{\text{partial}} = 0.7$ . By analyzing this partial residual, one can obtain an approximation of the marginalized posterior distribution that takes into account the presence of overlapping signals. In Fig. 3 the difference between  $S_{A,\text{residual}}(f)$

and  $S_{A,\text{partial}}(f)$  is clearly visible for  $f \in [2 \text{ mHz}, 10 \text{ mHz}]$  where most signals are found.

## H. Pipeline

To conclude we present in Algorithm 5 the full pipeline to extract GBs within a given frequency range of  $f_{\min}$  and  $f_{\max}$ . The output is the list of MLEs  $\tilde{\theta}_{\text{recovered}}$  and the list of MCMC chains  $\tilde{\theta}_{\text{posteriors}}$  which provide the posterior distribution.

Algorithm 5. The pipeline to obtain the MLE and posterior distribution of GBs within a large frequency range.

---

```

Function extracting_GBs( $f_{\min}, f_{\max}$ )
 $B_{\text{search}} \leftarrow \text{segmenting}(f_{\min}, f_{\max})$ 
split  $B_{\text{search}}$  into  $B_{\text{even}}, B_{\text{odd}}$ 
 $\tilde{\theta}_{\text{even}} \leftarrow \text{global\_GB\_search}(B_{\text{even}}, 3, d)$ 
 $\tilde{\theta}_{\text{odd}} \leftarrow \text{global\_GB\_search}(B_{\text{odd}}, 10, d - \sum_{\theta \in \tilde{\theta}_{\text{even}}} s(\theta))$ 
 $\tilde{\theta}_{\text{even}} \leftarrow \text{global\_GB\_search}(B_{\text{even}}, 10, d - \sum_{\theta \in \tilde{\theta}_{\text{odd}}} s(\theta))$ 
 $\tilde{\theta}_{\text{recovered}} \leftarrow \tilde{\theta}_{\text{even}} \cup \tilde{\theta}_{\text{odd}}$ 
 $\tilde{\theta}_{\text{posteriors}} \leftarrow \{\}$ 
 $d_{\text{residual}} = d - \sum_{\theta \in \tilde{\theta}_{\text{recovered}}} s(\theta)$ 
for all  $\theta_{\text{MLE}}$  in  $\tilde{\theta}_{\text{recovered}}$  do in parallel
     $d_{\text{posterior}} = d_{\text{residual}} + s(\theta_{\text{MLE}})$ 
    get  $\Theta_{\text{reduced}}$  as described in Sec. III G
     $\tilde{\theta}_{\text{posteriors}}$  append posterior( $\theta_{\text{MLE}}, d_{\text{posterior}}, \Theta_{\text{reduced}}$ )
end for
return  $\tilde{\theta}_{\text{recovered}}, \tilde{\theta}_{\text{posteriors}}$ 
    
```

---

## IV. RESULTS

The analysis of Radler’s, LDC1–4, started with the first 0.5 yr and continued with 1 and 2 yr, where the found signals from the previous analysis are used as initial guesses for the DE algorithm. The Sangria dataset, LDC2a, with the massive black hole binaries subtracted, is analyzed once for the full 1 yr of data. To prioritize high-quality extractions we set  $\rho_{\text{threshold}} = 9$ , which sets the acceptance threshold for found signals in Algorithm 2. The global frequency band is set to  $f_{\min} = 0.3 \text{ mHz}$  and  $f_{\max} = f_{\text{Nyquist}}$ , where  $f_{\text{Nyquist}} = 33.3 \text{ mHz}$  for the Radler challenge and  $f_{\text{Nyquist}} = 100 \text{ mHz}$  for the Sangria challenge.

### A. Computation times

Each segment of  $f_{\text{search}}$  in *global\_GB\_search* can be analyzed in parallel as noted with “do in parallel” in Algorithm 3. Therefore the shortest time to analyze the dataset  $T_{\text{parallel}}$  is determined by the sum of the three frequency segments which take the longest for each sequential analysis of  $B_{\text{even}}, B_{\text{odd}}$  and  $B_{\text{even}}$  segments as listed in Table II. The duration to analyze a segment varies a lot, where segments with no detectable signal are analyzed within 2 min and the longest computation time of a segment containing multiple detectable signals took 126 min.

TABLE III. Computational times of the Radler LDC1–4 data with different  $T_{\text{obs}}$ . The CPU time is the sum of the computational time of all analyzed frequency segments.  $T_{\text{parallel}}$  is the shortest computation time if the segments are analyzed in parallel on multiple CPU threads.

Challenge	$T_{\text{obs}}$ (yr)	CPU core time (h)	$T_{\text{parallel}}$ (h)
Radler	0.5	1607	3.2
Radler	1	2106	4.3
Radler	2	3269	5.5

The data analysis to obtain the MLEs was run on a high-performance computer. In Table III we present the search times for finding the MLE solutions of the Radler dataset. The pipeline demonstrates its efficiency by analyzing the longest observation time of  $T_{\text{obs}} = 2 \text{ yr}$  in only 6 h. In terms of computational cost, the analysis necessitates approximately 3300 h of CPU core hours. If commercial high-performance computing services such as those provided by Google are utilized, the estimated cost would amount to approximately 100 USD [45].

Furthermore, the computation of posterior distributions according to Sec. III G takes 1.8 s per signal on a Quadro RTX 4000 Mobile GPU. Therefore, for example for the 8385 recovered signals of the Sangria challenge it took 4.2 h on a single laptop to compute all posterior distributions.

### B. Matching recovered signals with injected signals

To evaluate the accuracy of the recovered signals  $\theta_{\text{rec}} \in \tilde{\theta}_{\text{recovered}}$ , we are matching them with the injected signals  $\theta_{\text{inj}} \in \tilde{\theta}_{\text{injected}}$  with similar frequencies. In order to determine matches quantitatively, we use the scaled error

$$\delta(s(\theta_{\text{rec}}), s(\theta_{\text{inj}})) = \frac{\langle s(\theta_{\text{rec}}) - s(\theta_{\text{inj}}), s(\theta_{\text{rec}}) - s(\theta_{\text{inj}}) \rangle}{\langle s(\theta_{\text{inj}}), s(\theta_{\text{inj}}) \rangle}, \quad (20)$$

which is dependent on the amplitude of the signals.

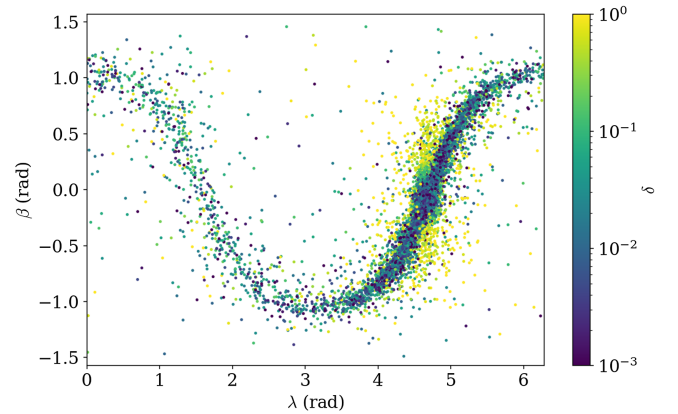


FIG. 4. Scatter plot of the scaled error  $\delta$  across the ecliptic longitude and ecliptic latitude of the 2 yr Radler dataset. The range of the error bar is clipped at  $10^{-3}$  to  $10^0$ .



TABLE IV. The variables of interest include the count of detectable injected GB sources, the count of recovered sources, the count of matches with injected sources, and the match rate. The match rate is determined by dividing the number of matched signals by the total number of recovered signals. The overlap is included to get an evaluation comparable to other analyses [9,10,15].

Challenge	$T_{\text{obs}}$ [yr]	Injected ( $\rho > 10$ )	Recovered	$\delta < 0.3$	Match rate $\delta < 0.3$	$\mathcal{O} > 0.9$	Match rate $\mathcal{O} > 0.9$
Radler	0.5	6813	3937	3418	87%	3407	87%
Radler	1	11 814	7112	6270	88%	6251	88%
Sangria	1	11 814	8375	7176	86%	7184	86%
Radler	2	18 332	11 953	10 372	87%	10 365	87%

In other works the scaled correlation, also called overlap,

$$\mathcal{O}(s(\theta_{\text{rec}}), s(\theta_{\text{inj}})) = \frac{\langle s(\theta_{\text{rec}}), s(\theta_{\text{inj}}) \rangle}{\sqrt{\langle s(\theta_{\text{rec}}), s(\theta_{\text{rec}}) \rangle \langle s(\theta_{\text{inj}}), s(\theta_{\text{inj}}) \rangle}} \quad (21)$$

of two signals  $s(\theta_{\text{rec}})$  and  $s(\theta_{\text{inj}})$  is used [46].

Figure 4 shows the sky locations of all recovered signals in ecliptic coordinates. The recovered signals follow the geometry of the Galaxy, with high  $\delta$  (yellow dots), slightly off the center of the Galaxy. In Table IV we present the number of recovered and matched signals for each analysis where we also include the overlap  $\mathcal{O}$  as a match metric for comparison with other evaluations which used the overlap [9,10,14,15]. The consistently high match rate of all analyses speaks of good quality recoveries. In Fig. 5 we see only small changes in the cumulative distribution function across the analyses. Only the cumulative distribution function of  $\mathcal{O}$  for the Sangria dataset has a higher count for smaller  $\mathcal{O}$ .

Given the potentially high correlation between a low-amplitude signal and a loud signal, even when the scaled error suggests a poor match, we classify the recovered signals with  $\delta < 0.3$  as “matched” signals. For each matched signal, we calculate the error using  $\Delta\beta = |\beta_{\text{rec}} - \beta_{\text{inj}}|$ . In

Fig. 6, we present the error histograms for all parameters. Notably, there is a clear trend of decreasing errors with longer observation times, as expected. The error histograms for the 1 yr analyses of the Radler and Sangria experiments exhibit similar patterns, consistent with our expectations. For the frequency and amplitude parameters, we display the relative errors. The relatively higher errors observed for  $\phi_0$  and  $\psi$  can be attributed to the inherent degeneracy between these two parameters. However, it is evident that the degeneracy diminishes with increasing  $T_{\text{obs}}$ .

There are several instances of recovered signals not aligning with the injected signal. One scenario arises when a recovered signal exhibits a poor match with an injected signal due to excessive noise in the data, which is similar to the second situation where a recovered signal adapts to the combined contributions of multiple injections. A third scenario occurs when multiple recovered signals closely resemble a single injected signal. This third scenario does not yield a robust data analysis and indicates an issue within the pipeline.

In Table V, we provide a quantitative assessment of the third scenario by enumerating the occurrences of multiple recovered signals displaying a significant correlation  $\mathcal{O}$  with the same injected signal. In all observed occurrences, two recovered signals demonstrate a high correlation with

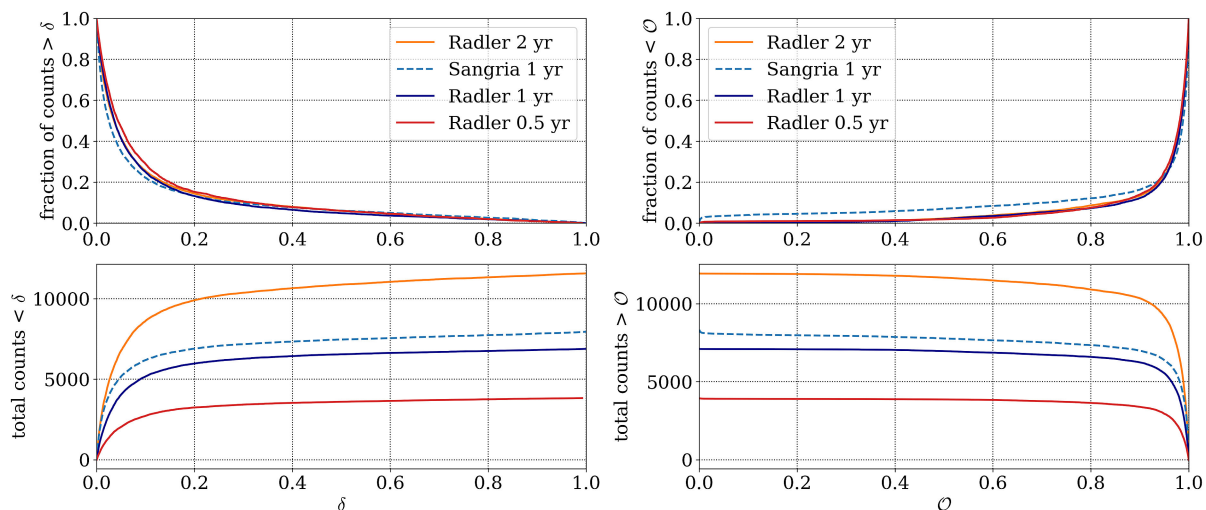
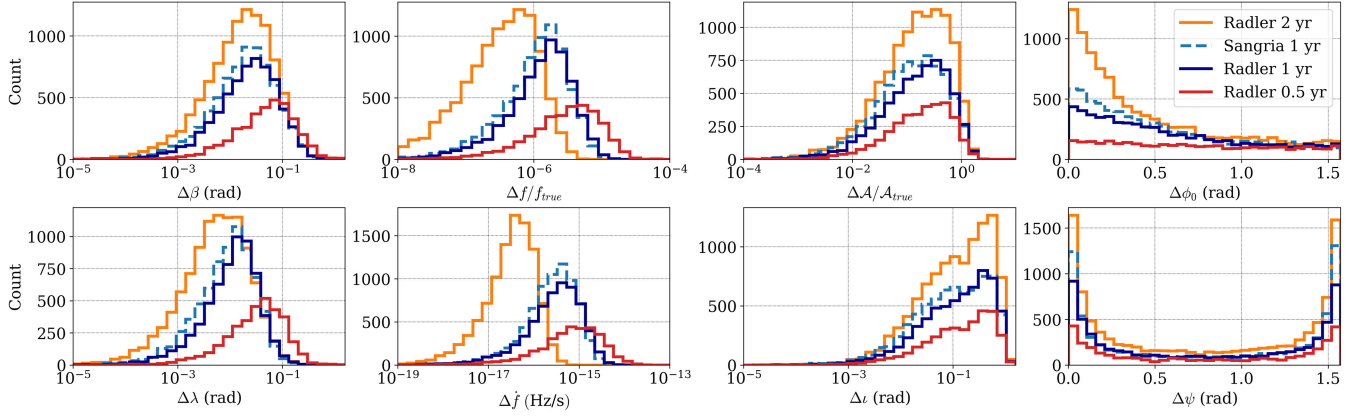


FIG. 5. Cumulative distribution function of  $\delta$  and  $\mathcal{O}$  in the top plots and the survival function in the bottom plots. The plots of the overlap  $\mathcal{O}$ , on the right, are comparable with other analyses such as [14,15].


 FIG. 6. Error histogram of all matched signals with  $\delta < 0.3$ .

one injected signal. The table illustrates various thresholds for defining a high correlation, and it is noteworthy that the occurrence of this problematic scenario is exceedingly rare, even when employing a correlation threshold of  $\mathcal{O} > 0.5$  only around 0.2% of the recovered signals are problematic.

### C. Galaxy

The recovered signals that meet the matching criteria are visualized as green dots in Fig. 7. Additionally, it is evident from the plot that the recovered signals without a satisfactory match predominantly have lower amplitudes  $\mathcal{A}$ . The ability of LISA to recover signals is contingent upon the sensitivity curve, which exhibits lower sensitivity at lower frequencies. Consequently, only signals with higher amplitudes are recoverable at low frequencies. The majority of the recovered signals are concentrated in the central region of the Milky Way, which is also the location of a significant portion of the sources.

For each matched signal with  $\dot{f} > 0$ , where we assume that the evolution of the GB is purely driven by the emission of GWs, we can estimate the luminosity distance [14]:

$$D_L = \frac{5\dot{f}}{48\mathcal{A}\pi^{6/3}f_0^{5/3}}, \quad (22)$$

which is a good estimate of the distance in Euclidean space for objects in the Milky Way. Therefore we are able to convert their GBs to the galactocentric coordinate system and present them in Fig. 8. The upper plot illustrates the

distribution of all injected signals with  $f > 0.3$  mHz and  $\dot{f} > 0$ , while the lower plot depicts the recovered GBs. It should be noted that the number of recovered GBs is lower than the injected ones due to the majority of injected

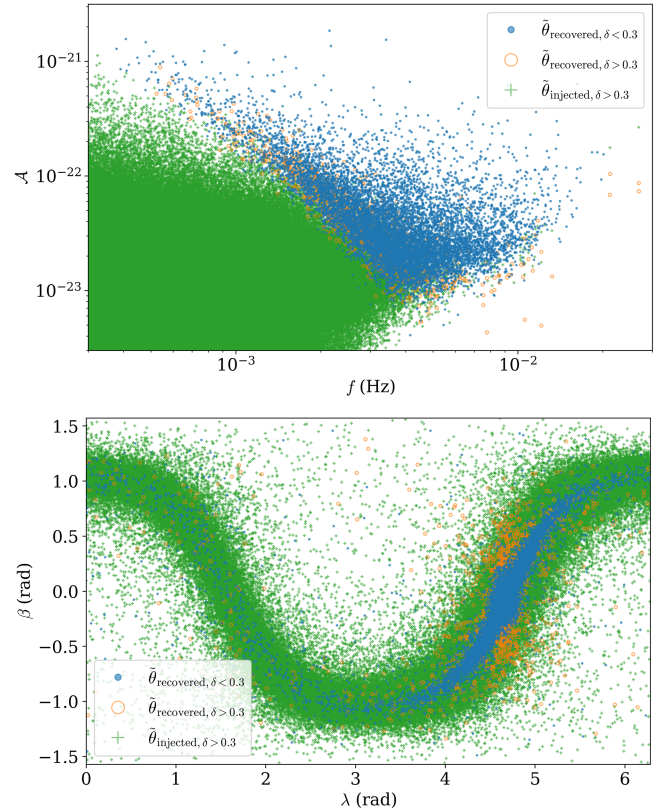


FIG. 7. Scatter plot of recovered GBs and injected GBs of the 2 yr Radler dataset. The upper plot is across the GB amplitude  $\mathcal{A}$  and frequency  $f$  and the lower plot is across the ecliptic sky locations. The green dots are the recovered signals with  $\delta < 0.3$  which are categorized as matched signals. The blue circles on the plot represented the recovered signals that did not have a close match with any of the injected signals. The red crosses represent the injected signals that did not have a good match with any of the recovered signals. These signals were not effectively captured or identified during the recovery process.

TABLE V. The number of occurrences where two recovered signals exhibit a strong correlation with the same injected signal.

Challenge	$T_{\text{obs}}$ [yr]	Recovered	$\mathcal{O} > 0.5$	$\mathcal{O} > 0.7$	$\mathcal{O} > 0.85$
Radler	0.5	3937	3	0	0
Radler	1	7112	2	0	0
Sangria	1	8375	11	3	0
Radler	2	11 953	10	2	0

GBs having a low S/N, rendering them unrecoverable. Notably, a significant number of recovered GBs are located in close proximity to the Sun, which aligns with expectations as closer sources exhibit higher S/N. This trend is also evident in the galactocentric 3D plot shown in Fig. 9.

#### D. Posterior

Assessing the posterior distribution of 10 000 signals presents challenges, particularly in the absence of ground truth for comparison. However, leveraging statistical techniques allows us to evaluate the quality of the uncertainty estimates. Additionally, we can quantify the enhanced precision of the posterior distribution as  $T_{\text{obs}}$  increases. In Fig. 10, we observe the evolution of accuracy and precision for one signal's sky location as a function of  $T_{\text{obs}}$ . Notably, for  $T_{\text{obs}} = 0.5$  yr, the accuracy and

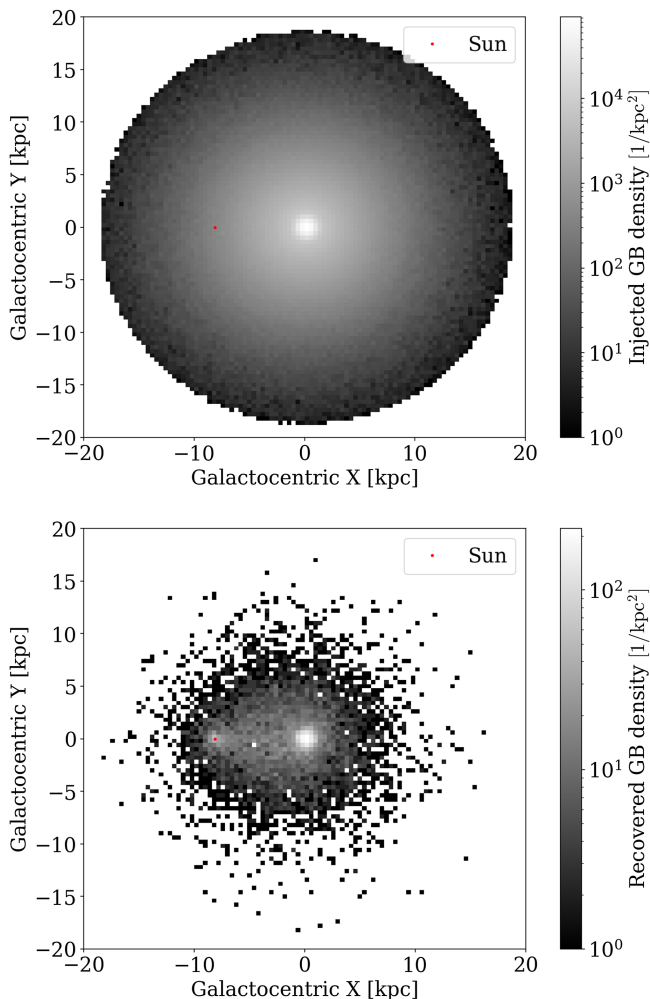


FIG. 8. The GB distribution of the Milky Way galaxy seen perpendicular to the Galactic plane according to the simulated Radler dataset. The red dot marks the Sun. The top plot shows the distribution of the injected GBs and the bottom plot the distribution of the recovered GBs.

precision are comparatively lower than those achieved with longer  $T_{\text{obs}}$ .

Since the posterior distribution for the sky locations is approximate of Gaussian shape, we can estimate the uncertainty by computing the standard deviation  $\sigma_\beta = \sqrt{\frac{1}{N} \sum_{i=1}^N (\beta_i - \mu_\beta)^2}$  with mean  $\mu_\beta = \frac{1}{N} \sum_{i=1}^N \beta_i$ , where  $N = n_{\text{samples}}$  is the length of the MCMC chain and  $\beta_i$  is the  $i$ th sample of the chain. The uncertainty for the other parameters is computed analogously. In the next step, we can estimate the angular confidence area

$$\sigma_{\text{area}} = \int_{\mu_\beta - \sigma_\beta}^{\mu_\beta + \sigma_\beta} \int_{\mu_\lambda - \sigma_\lambda}^{\mu_\lambda + \sigma_\lambda} \sin \beta d\lambda d\beta \quad (23)$$

of the sky location for each signal. Figure 11 displays the histogram of all analyses, revealing a notable trend. As  $T_{\text{obs}}$  increases, the number of posteriors with small confidence areas also increases. This observation aligns with the findings depicted in Fig. 10, showing how the posterior of a signal becomes narrower with longer  $T_{\text{obs}}$ . However, it is important to note that the total number of extracted signals from the data also rises as the S/N of signals improves with longer observation times. Consequently, the number of signals with wider confidence areas also increases.

To assess the quality of the posterior estimate, we can examine whether the true parameters lie within the confidence interval as expected. If the accuracy and precision of the posterior are correct, we would expect to find the true parameters approximately 68% of the time within the interval of  $1\sigma$  standard deviation.

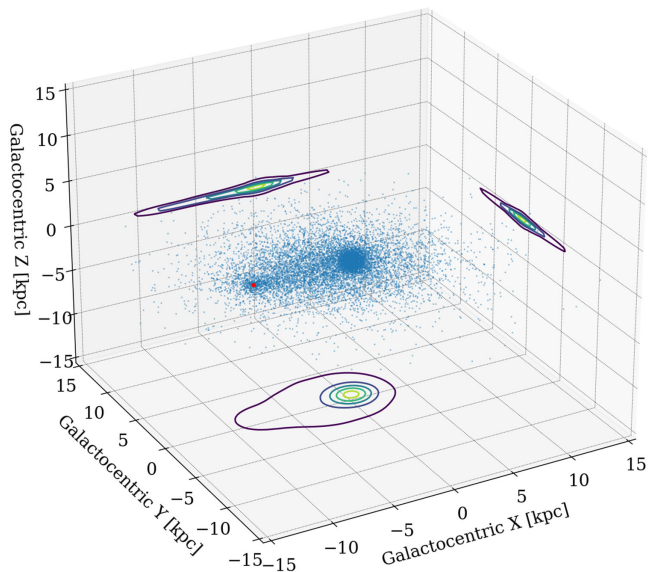


FIG. 9. Recovered GBs plotted as blue dots in the galactocentric coordinate system. The red dot marks the Sun. The density of GBs is represented by 2D contour lines on the planes.

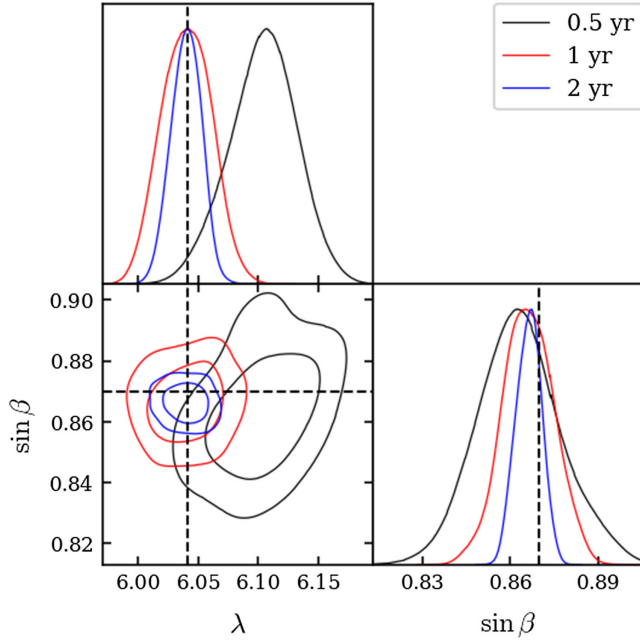


FIG. 10. Posterior distribution of the sky location for the three analyses of the Radler dataset of the signal with  $\theta_{\text{inj}}^f = 4.169906$  mHz. The dashed black lines mark the true values of the matched injected signal  $\theta_{\text{inj}}$ .

If the number of parameters within the confidence interval is higher than expected, it suggests that the precision is worse, meaning the posterior distribution is too wide. On the other hand, if the number of parameters within the confidence interval is lower than expected, it indicates potential inaccuracies in the posterior estimate. This could mean that the posterior distribution is not located at the true parameters, and/or it is excessively precise, where the posterior distribution is too narrow.

The results for individual parameters are presented in Table VI, where we compute the standard deviation and check if the true parameter falls within the 68% confidence interval. Due to degeneracy, the evaluation of  $\phi_0$  and  $\psi$  is

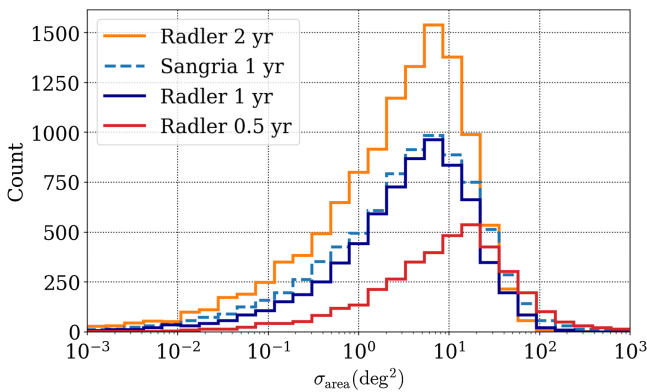


FIG. 11. Angular confidence area histogram of all analyses.

TABLE VI. Ratio of true values within  $1\sigma$  standard deviation.

Parameter	Radler 0.5 yr	Radler 1 yr	Sangria 1 yr	Radler 2 yr
$\mathcal{A}$	23%	29%	34%	33%
$\sin\beta$	52%	61%	67%	69%
$\lambda$	24%	54%	60%	72%
$f$	20%	28%	32%	39%
$\dot{f}$	53%	45%	45%	57%
$\iota$	30%	36%	42%	40%

omitted as it would not yield proper assessment. We observe that the uncertainty estimate for the sky locations, with observation times of 1 yr or more, is close to the expected rate. However, the other parameters exhibit a lower rate than expected. This can be attributed to multiple reasons. Firstly, it could be due to inaccurate estimation, where the true value does not align with the posterior distribution. Secondly, the posterior distribution might be too narrow. Lastly, the assumption of a Gaussian distribution for the other parameters, as used in computing the standard deviation, may not hold true. For multimessenger astronomy, the good agreement between estimated and true uncertainty in sky location is of highest relevance.

## V. CONCLUSION

The extension of the previous pipeline, outlined by [16], allows now to obtain the MLE of GBs in the full frequency range where most of the GBs are overlapping with each other. As detailed in Sec. IV A, the extraction of 18 000 signals from a dataset with  $T_{\text{obs}} = 2$  yr can be accomplished in a mere 6 h. This acceleration also reduces the computational costs significantly to only 100 USD with today's hardware [45], which brings extracting GBs from the full frequency band toward diminishing costs.

Additionally, we have leveraged the power of parallel computation, utilizing GPUs, to compute the posterior distribution for identified MLEs within a remarkable time frame of 2 s per signal. The computation of all posterior distributions can be completed in approximately 9 h on a single laptop-grade GPU of the year 2018. These advancements not only enable efficient analysis of a large number of signals, but also allow for rapid estimation of the posterior distributions.

The next crucial step involves integrating the presented pipeline into a comprehensive global analysis of data encompassing various astrophysical sources and phenomena, such as GBs, MBHBs, extreme mass ratio inspirals, glitches, and data gaps. The incorporation of this pipeline into the development of a global analysis framework offers substantial acceleration in the GB analysis process. This acceleration leads to a notable reduction in the associated costs for future pipeline developments.

The full pipeline and evaluation tools are available at [47].

## ACKNOWLEDGMENTS

We thank the LDC working group [19] for the creation and support of the LDC1–4 and LDC2a. Furthermore, we acknowledge the GPU implementation of FASTLISARESPONSE [48]. The calculations of Algorithm 3 were run on the CPUs of the Euler cluster of ETH Zürich and are gratefully acknowledged. This project is supported by the Swiss National Science Foundation (SNF 200021\_185051).

- 
- [1] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, *Phys. Rev. X* **9**, 031040 (2019).
- [2] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi *et al.*, Laser interferometer space antenna, [arXiv:1702.00786](https://arxiv.org/abs/1702.00786).
- [3] R. Taam, B. Flannery, and J. Faulkner, Gravitational radiation and the evolution of cataclysmic binaries, *Astrophys. J.* **239**, 1017 (1980).
- [4] B. Willems, A. Vecchio, and V. Kalogera, Probing white dwarf interiors with LISA: Periastron precession in eccentric double white dwarfs, *Phys. Rev. Lett.* **100**, 041102 (2008).
- [5] G. Nelemans, L. Yungelson, M. v. d. Sluys, and C. A. Tout, The chemical composition of donors in AM CVn stars and ultracompact x-ray binaries: Observational tests of their formation, *Mon. Not. R. Astron. Soc.* **401**, 1347 (2010).
- [6] T. B. Littenberg and N. Yunes, Binary white dwarfs as laboratories for extreme gravity with LISA, *Classical Quantum Gravity* **36**, 095017 (2019).
- [7] A. L. Piro, Inferring the presence of tides in detached white dwarf binaries, *Astrophys. J. Lett.* **885**, L2 (2019).
- [8] Y. Bouffanais and E. K. Porter, Detecting compact galactic binaries using a hybrid swarm-based algorithm, *Phys. Rev. D* **93**, 064020 (2016).
- [9] X. Zhang, S. D. Mohanty, X. Zou, and Y. Liu, Resolving galactic binaries in LISA data using particle swarm optimization and cross-validation, *Phys. Rev. D* **104**, 024023 (2021).
- [10] P. Gao, X.-L. Fan, Z.-J. Cao, and X.-H. Zhang, Fast resolution of galactic binaries in LISA data, *Phys. Rev. D* **107**, 123029 (2023).
- [11] J. Crowder and N. J. Cornish, Solution to the galactic foreground problem for LISA, *Phys. Rev. D* **75**, 043008 (2007).
- [12] J. Crowder and N. J. Cornish, Extracting galactic binary signals from the first round of Mock LISA data challenges, *Classical Quantum Gravity* **24**, S575 (2007).
- [13] T. B. Littenberg, Detection pipeline for galactic binaries in LISA data, *Phys. Rev. D* **84**, 063009 (2011).
- [14] T. B. Littenberg, N. J. Cornish, K. Lackeos, and T. Robson, Global analysis of the gravitational wave signal from galactic binaries, *Phys. Rev. D* **101**, 123021 (2020).
- [15] T. B. Littenberg and N. J. Cornish, Prototype global analysis of LISA data with multiple source types, *Phys. Rev. D* **107**, 063004 (2023).
- [16] S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler, and D. Giardini, Bayesian parameter estimation of galactic binaries in LISA data with Gaussian process regression, *Phys. Rev. D* **106**, 062003 (2022).
- [17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (MIT Press, Cambridge, MA, 2005), pp. 7–31.
- [18] M. L. Katz, J.-B. Bayle, A. J. K. Chua, and M. Vallisneri, Assessing the data-analysis impact of LISA orbit approximations using a GPU-accelerated response model, *Phys. Rev. D* **106**, 103001 (2022).
- [19] S. Babak and A. Petiteau, <https://lisa-ldc.lal.in2p3.fr/>.
- [20] M. Tinto and J. W. Armstrong, Cancellation of laser noise in an unequal-arm interferometer detector of gravitational radiation, *Phys. Rev. D* **59**, 102003 (1999).
- [21] J. W. Armstrong, F. B. Estabrook, and M. Tinto, Time-delay interferometry for space-based gravitational wave searches, *Astrophys. J.* **527**, 814 (1999).
- [22] F. B. Estabrook, M. Tinto, and J. W. Armstrong, Time-delay analysis of LISA gravitational wave data: Elimination of spacecraft motion effects, *Phys. Rev. D* **62**, 042002 (2000).
- [23] S. V. Dhurandhar, K. R. Nayak, and J.-Y. Vinet, Algebraic approach to time-delay data analysis for LISA, *Phys. Rev. D* **65**, 102002 (2002).
- [24] M. Tinto and S. V. Dhurandhar, Time-delay interferometry, *Living Rev. Relativity* **17**, 1 (2014).
- [25] M. Vallisneri, Synthetic LISA: Simulating time delay interferometry in a model LISA, *Phys. Rev. D* **71**, 022001 (2005).
- [26] N. J. Cornish and T. B. Littenberg, Tests of bayesian model selection techniques for gravitational wave astronomy, *Phys. Rev. D* **76**, 083006 (2007).
- [27] P. A. Mazzali, F. K. Ropke, S. Benetti, and W. Hillebrandt, A common explosion mechanism for type ia supernovae, *Science* **315**, 825 (2007).
- [28] M. Vallisneri, A LISA data-analysis primer, *Classical Quantum Gravity* **26**, 094024 (2009).
- [29] C. E. Shannon, Communication in the presence of noise, *Proc. IRE* **37**, 10 (1949).
- [30] F. Auger and P. Flandrin, Improving the readability of time-frequency and time-scale representations by the re-assignment method, *IEEE Trans. Signal Process.* **43**, 1068 (1995).

- [31] S. A. Fulop and K. Fitz, Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications, *J. Acoust. Soc. Am.* **119**, 360 (2006).
- [32] R. Storn and K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.* **11**, 341 (1997).
- [33] J.-F. Bonnans, J.C. Gilbert, C. Lemaréchal, and C.A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects* (Springer Science & Business Media, New York, 2006), pp. 256–258.
- [34] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [35] R.B. Blackman and J.W. Tukey, The measurement of power spectra from the point of view of communications engineering—part I, *Bell Syst. Tech. J.* **37**, 185 (1958).
- [36] A. Savitzky and M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* **36**, 1627 (1964).
- [37] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [38] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97 (1970).
- [39] T. A. O’Brien, K. Kashinath, N. R. Cavanaugh, W. D. Collins, and J. P. O’Brien, A fast and objective multidimensional kernel density estimation method: fastKDE, *Computational Statistics and Data Analysis* **101**, 148 (2016).
- [40] T. A. O’Brien, W. D. Collins, S. A. Rauscher, and T. D. Ringler, Reducing the computational cost of the ECF using a nufft: A fast and objective probability density estimation method, *Computational Statistics and Data Analysis* **79**, 222 (2014).
- [41] M. L. Katz, C. Danielski, N. Karnesis, V. Korol, N. Tamanini, N. J. Cornish, and T. B. Littenberg, Bayesian characterization of circumbinary sub-stellar objects with LISA, *Mon. Not. R. Astron. Soc.* **517**, 697 (2022).
- [42] N. J. Cornish and T. B. Littenberg, Tests of Bayesian model selection techniques for gravitational wave astronomy, *Phys. Rev. D* **76**, 083006 (2007).
- [43] T. Robson, N. J. Cornish, N. Tamanini, and S. Toonen, Detecting hierarchical stellar systems with LISA, *Phys. Rev. D* **98**, 064012 (2018).
- [44] M. L. Katz, mikekatz04/gbgpu: First official public release (2022).
- [45] Google-Cloud prices, <https://cloud.google.com/pricing/> (accessed on 2023-06-12).
- [46] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Comparison of search templates for gravitational waves from binary inspiral, *Phys. Rev. D* **63**, 044023 (2001).
- [47] S. Strub, LDC-GB Accelerating global parameter estimation of Galactic binaries using a genetic algorithm and GPUs, [10.5281/zenodo.8123815](https://doi.org/10.5281/zenodo.8123815) (2023).
- [48] M. Katz and J.-B. Bayle, mikekatz04/lisa-on-gpu: First official public release, [10.5281/zenodo.5867731](https://doi.org/10.5281/zenodo.5867731) (2022).