

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Fundamental Research

journal homepage: <http://www.keaipublishing.com/en/journals/fundamental-research/>

Review

Extracting overlapping gravitational-wave signals of Galactic compact binaries: A mini review [☆]Rui Niu ^{a,b}, Wen Zhao ^{a,b,*}^a Department of Astronomy, University of Science and Technology of China, Chinese Academy of Sciences, Hefei 230026, China^b School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026, China

ARTICLE INFO

Article history:

Received 16 June 2024

Received in revised form 7 November 2024

Accepted 13 November 2024

Available online 19 November 2024

Keywords:

Gravitational wave

Galactic compact binaries

Global fitting method

Iterative subtraction method

Machine learning technique

ABSTRACT

Gravitational wave (GW) observations have provided a novel tool to explore the universe. In the near future, space-borne detectors will further open the window of low-frequency GW band where abundant sources exist and invaluable information for astrophysics, cosmology, and fundamental physics can be revealed. However, there are various new challenges in data analyses for space-borne detectors coming with the abundance of GW signals. For example, there are Galactic compact binaries (GCBs) with an overwhelming number that can produce continuous GW signals existing the entire mission time of detectors. The enormous overlapping GCB signals tangle and correlate with each other, and blend with other types of sources together in the observed data. Extracting source information from overlapping signals is one of the key problems for data analyses of space-borne detectors. In the paper, we present a review of currently available solutions for extracting overlapping GCB signals as thoroughly as possible aiming at promoting more interest in this question and inspiring further improvements. Current solutions can be roughly categorized by two classes, iterative subtraction and global fitting. There are diverse implementations of both strategies with enhancements focusing on different aspects. Meanwhile, the hybrid approach and the machine learning technique are also used in recent years. In the last, we also present an introduction of the stochastic foreground formed by unresolvable faint GCBs about its separation from extragalactic backgrounds and its utility in exploring the properties of the Galaxy.

1. Introduction

Direct detections of gravitational waves (GWs) by ground-based detectors have ushered in a new era of GW astronomy [1–4]. Fruitful results have been accomplished based on detected events and researches with GWs are explosively growing in recent years. These detections have initiated a paradigm shift in studies of gravity, astrophysics, and cosmology. In the near future, space-borne detectors including LISA [5], Taiji [6], and TianQin [7] will open the new window of low-frequency GW band.

One significant difference between current ground-based detectors and future space-borne detectors is that data given by current ground-based detectors are noise-dominant whereas signals are dominant for space-borne detectors. For current ground-based detectors, the compact binaries will quickly merge after entering the sensitive band, and the transient signals are sparsely distributed in time [1–4]. The situations of signal overlapping are expected to be rare. As reported in the work [8] where the probability of signal overlapping and how severe overlapping can induce significant bias in parameter estimation for second-generation ground-based detectors are thoroughly investi-

gated, it is unlikely to observe overlapping signals by current existing detectors. Whereas, for future space-borne detectors, the enormous sources can persist in the sensitive band during the whole mission period [5–7]. Their GW signals can be heavily overlapping both in the time and frequency domain, which brings new challenges in extracting physical information from the data. Similar issues of signal overlapping also arise in third-generation ground-based detectors. Due to the exquisite sensitivities especially the improvement in low frequency, the visible duration and signal number are both significantly increased. It is unlikely to observe signals without overlapping for third-generation detectors, and the events with very close merger time which may suffer from significant bias in parameter estimation can be common [9,10]. There are extensive works about parameter inference techniques and impacts on researches of scientific problems for overlapping signals in third-generation detectors. For examples, the traditional Bayesian inference framework with strategies of hierarchical subtraction and joint estimation for overlapping signals is investigated in the work [11], and a joint parameter estimation analyzing two overlapping signals simultaneously using normalizing flows is demonstrated in [12]. In the work [13], impacts of signal overlapping on testing General Relativity are

[☆] Peer review under the responsibility of Editorial Board of Fundamental Research.

* Corresponding author.

E-mail address: wzhao7@ustc.edu.cn (W. Zhao).

elaborated. In works [14,15], the impacts of unresolvable overlapping foreground and subtraction residuals induced by parameter estimation bias due to signal overlapping on detecting cosmological stochastic gravitational wave background with third-generation detectors are investigated. A review for signal overlapping problems focusing on third-generation ground-based detectors can be found in [16]. In the rest of this paper, we mainly focus on researches targeting space-borne detectors.

There are diverse types of sources presented in the space-borne detector sensitive band [17], such as massive black hole binaries (MBHBs), extreme mass ratio inspirals (EMRIs), and Galactic compact binaries (GCBs), etc. GCBs are likely to be the most numerous type of GW sources detected in future space GW observation. Tens of millions of such sources may exist in the space-borne detector band. Among them, ten of thousands of bright ones are expected to be independently resolvable, while others will constitute a stochastic foreground as confusion noise [18–21]. Identification and subtraction of the resolvable GCBs are not only important for extracting other kinds of sources but also can be assistance in researches about stellar and Galactic astrophysics with the advantage that the information carried by GWs will not be affected by crowded matters in the Galaxy [22–27].

Although all sources including MBHBs and EMRIs will be blended together in data-stream from space-borne detectors and people may ultimately need algorithms that can separate or simultaneously fit all different sources, GCBs may be the kind that is most heavily overlapping due to their overwhelming amount and the feature of long-living. Therefore, the discussions dedicated to GCBs may be the foundation and starting point of the ultimate full algorithms for separating and fitting all overlapping sources. In this paper, we will focus on GCBs and present a review of current solutions to tackle enormous overlapping GCB sources as comprehensively as possible, aiming at paving for further improvements in the problems of signal overlapping for data analyses of space-borne detectors.

Current solutions of extracting GCB signals mainly work around the simulated data sets including the earlier Mock LISA Data Challenges (MLDCs) [28–34], the recent resurrected LISA Data Challenges (LDCs) [35], and the Taiji Data Challenges (TDCs) [36]. These data sets are released to encourage various efforts for tackling unsolved problems in data analyses of space-borne detectors. These challenges involve data sets not only dedicated to GCBs, but also blended with various other possible sources, and the complexity has been increasing progressively. In the MLDC1, isolated GCBs and moderately overlapping signals with dozens of GCBs are concerned [28–30]. The MLDC2 considers the full population with 26 million GCBs, and includes two data sets where the MLDC2.1 only contains signals of GCBs while MLDC2.2 blends signals of GCBs with MBHBs and EMRIs [31,32]. The data set of GCBs in MLDC3 contains ~ 60 million binaries, which is descended from MLDC2.1 with the improvement of realism by considering two different kinds of GCBs, binaries with two detached components and binaries with interacted components [33,34]. The MLDC4 is the descendant of MLDC2.2, which blends different types of sources concerned separately in sub-challenges of MLDC3 into a single data stream [33].

The new LDCs¹ [35] are resumed in recent years with the new design [5] of LISA. Currently, three challenges have been released, LDC1 Radler, LDC2a Sangria, and LDC2b Spritz. The LDC1 considers various types of possible LISA sources separately like MLDC3, and includes six sub-challenges where LDC1-4 is dedicated to GCBs and contains 26 million signals. The LDC2a is the updated challenge similar to MLDC4, which mixes different types of sources. The LDC2b has the improvement of considering the realistic instrumental and environmental noise, including gaps, glitches, and non-stationary noise. While the GW sources considered in LDC2b are relatively simple, where only MBHBs and verification GCBs are contained. The Taiji project also releases the

data challenge [36] for the configuration of the Taiji detector, which includes sub-challenges concerning various types of GW sources separately and the mixture of these sources.

Numerous endeavors have been made in previous works to address the problem of signal overlapping. The ideas for this problem can be roughly categorized into two groups, the iterative subtraction strategy and the global fitting strategy. The iterative subtraction strategy searches the maximum likelihood estimation for a single source in each step, and this procedure will be performed iteratively with the remaining data after subtracting the identified signals [37–41]. In contrast, the global fitting strategy fits all signals simultaneously in the full Bayesian approach with sophisticated Markov chain Monte Carlo (MCMC) sampling algorithms to obtain the joint posterior distributions [42–49]. Both two strategies are adopted and implemented in diverse full-scale and end-to-end pipelines for extracting overlapping signals of GCBs. The iterative subtraction solutions can extract source candidates quickly but suffer from the correlations among overlapping signals and the contamination of accumulating signal residuals left by each subtraction iteration. The global fitting solutions employ the full Bayesian approach to analyze all sources in a band simultaneously, which can better deal with the source correlations and residual contamination, but with the price of extremely massive demand for computational resources. The hybrid approach combining the maximum likelihood estimation and the Bayesian parameter estimation with MCMC is also proposed for combining the strengths of the two strategies while evading their drawbacks [50–52]. The flourishing machine learning algorithms provide new avenues to solve parameter estimation problems. After training with simulated data, the algorithms can generate posteriors directly without enormously evaluating the computationally expensive likelihood, which can complete the parameter estimation nearly in real time. The techniques have been successfully used in the parameter estimation for current ground-based detector data [53–58]. Machine learning techniques are also considered in the problem of overlapping GCBs, which are expected to be prospective powerful tools in future data analyses of space-borne detectors [59]. The resolvable sources are only a small fraction of the whole GCB population, the remaining faint GCBs can form a stochastic foreground. On the one hand, this Galactic foreground plays the role of noise deteriorating the signal-to-noise ratio (SNR) of individual sources and blending with stochastic signals from extra-galactic sources [60–69]. On the other hand, this foreground also contains information on the GCB population which is correlated to properties of the Galaxy, and offers a unique tool to study Galactic astrophysics [70–72].

The rest of this paper is organized as follows. In the next section, we briefly introduce the characterizations of GW signals from GCBs and how detectors respond to them. Section 3 is the main part of this paper where we present a detailed review of efforts on extracting overlapping GCB signals reported in recent years. After briefly summarizing earlier works around MLDC in Section 3.1, we first introduce a typical implementation of the iterative subtraction scheme and two variants focusing respectively on reducing inaccurate subtraction contamination and improving search efficiency in Section 3.2. Then, we elucidate basic conceptions of the global fitting scheme and introduce two independent implementations of this strategy in Section 3.3. The hybrid Bayesian approach combining the maximum likelihood estimation and the MCMC sampling, and a preliminary attempt of utilizing machine learning techniques in solving the problem of overlapping GCB signals are introduced next in Section 3.4 and 3.5. The discussions about unresolvable GCBs are given in Section 4. The final summary is presented in Section 5.

2. Galactic compact binaries

According to population models [18–21], there are tens of millions of compact binaries in the Galaxy that are slowly inspiraling towards each other with emissions of GWs in the mHz band and might be the type of most numerous sources observed by the space-borne detectors. An illustration of the sky distribution of a simulated GCB population taken

¹ <https://lisa-ldc.lal.in2p3.fr/>.

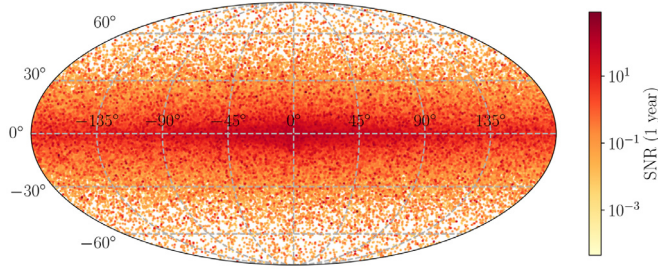


Fig. 1. Illustration for sky distribution of GCBs. The simulated GCB catalog is from LDC2a [35]. The SNRs are calculated with 1-year observation and the sources are displayed in the Galactic coordinate.

Table 1
Expected numbers for the total and detectable sources of various GCB types in the Galaxy. This table is cited from Amaro-Seoane et al. [17].

Types	N_{total}	$N_{\text{detectable}}$
WD+WD	$\sim 10^8$	6000 – 10000
NS+WD	$\sim 10^7$	100 – 300
BH+WD	$\sim 10^6$	0 – 3
NS+NS	$\sim 10^5$	2 – 100
BH+NS	$\sim 10^4 - 10^5$	0 – 20
BH+BH	$\sim 10^6$	0 – 70

from the training dataset of LDC2a [35] is presented in Fig. 1. Among the population, tens of thousands of GCBs are expected to be individually resolvable through the four-year observation time of space-borne detectors. Current electromagnetic observations have identified about dozens of GCBs² [73], while hundreds are predicted to be detected by future observations [20]. These known sources are referred to as verification binaries and the loud ones are guaranteed to be detectable by space-borne GW detectors. The loud verification binaries are expected to be quickly identified by just weeks integration time of observation, thus offering an important tool for functional tests and performance monitoring of the instruments [74]. Most GCBs are binary white dwarfs including detached binaries, as well as interacted binaries that have reached the Roche lobe overflow and started mass transfer [75]. While a small fraction of GCBs may involve with neutron stars or black holes. A summary of expectation of GCBs in the space-borne detector band is present in Table 1 which is cited from Amaro-Seoane et al. [17]. More details about population synthesis simulations or formation scenarios can be found in reviews [17,76].

GCBs in the band of space-borne detectors are far from merger, and will stay in the inspiral phase of slow chirping during the entire observation period. GWs radiated from GCBs can be well described by the quasi-monochromatic waveform which takes the form of [77,78]

$$\begin{aligned} h_+(t) &= \mathcal{A}(1 + \cos^2 \iota) \cos \Phi(t), \\ h_\times(t) &= -2\mathcal{A} \cos \iota \sin \Phi(t). \end{aligned} \quad (1)$$

Here \mathcal{A} is the amplitude, ι is the inclination of binary orbit, and Φ is the GW phase which can be expressed as $\Phi(t) = \phi_0 + 2\pi \int f(t') dt'$ with the arbitrary initial phase ϕ_0 and the frequency evolution $f(t)$. Since the frequency evolution is extremely slow for GCBs in the early inspiral stage, $f(t)$ can be characterized by the central frequency f_0 and the first derivative \dot{f} . The evolution of amplitude is usually neglected and \mathcal{A} is considered as a constant. The effects of mass transfer between binaries can be encoded into the parameter \dot{f} [79]. The phase evolution can be written as

$$\Phi(t) = 2\pi f_0 t + \pi \dot{f} t^2 + \phi_0. \quad (2)$$

² <https://gitlab.in2p3.fr/LISA/lisa-verification-binaries>.

The total matrix perturbation can be assembled by the sum of two polarizations as

$$\mathbf{h}^{\text{TT}} = \epsilon_+ h_+ + \epsilon_\times h_\times, \quad (3)$$

where ϵ_+ and ϵ_\times denote the polarization tensors which depend on the source location (β, λ) and the polarization angle ψ . The response to GWs of a single laser link of space-borne detectors is given by Cornish and Rubbo [80], Rubbo et al. [81], Królak et al. [82]

$$y_{sr} = \frac{1}{2(1 - \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}_{sr})} \hat{\mathbf{n}}_{sr} \cdot [\mathbf{h}^{\text{TT}}(t - L - \hat{\mathbf{k}} \cdot \mathbf{p}_s) - \mathbf{h}^{\text{TT}}(t - \hat{\mathbf{k}} \cdot \mathbf{p}_r)] \cdot \hat{\mathbf{n}}_{sr}, \quad (4)$$

where the subscripts s, r denote the laser sender node and the receiver node respectively, $\hat{\mathbf{k}}$ denotes the unit vector of the GW propagation direction, $\hat{\mathbf{n}}_{sr}$ denotes the unit vector along the direction of the link, \mathbf{p}_s and \mathbf{p}_r are the vectors of positions of the sender node and receiver node in the heliocentric coordinate. Together with the parameters describing GW strains of GCBs as shown in Eqs. 1 and 2, there are 8 parameters $(\mathcal{A}, f_0, \dot{f}, \iota, \lambda, \beta, \psi, \phi_0)$ to fully characterize a signal from GCB.

Since the noise behavior can be conveniently characterized by the power spectral density (PSD) in the frequency domain if the noise is stationary and Gaussian, data analyses of GWs are often performed in the frequency domain. The method proposed in [78] can compute the Fourier transformation of the response quickly and accurately by heterodyning the response signal with a carrier wave of the frequency f_0 . By multiplying with the carrier wave, the response signal can be decomposed into the slow part and the fast part. The Fourier transformation of the fast part can be obtained analytically. The slow part is transformed through fast Fourier transformation numerically, whereas the number of time samples is significantly reduced. For a signal extending within the band of $[f_0, (1 + \eta)f_0]$, the heterodyning operation can shift the required Nyquist frequency from $2(1 + \eta)f_0$ to $2\eta f_0$. Since the GCBs signals are quasi-monochromatic with $\eta \sim 10^{-6}$ [21,83], the number of samples can be much less than the original time samples when numerically computing the fast Fourier transformation for the slow part. There is open source code GBGPU [77,84] that can be used to obtain the response signals of GCBs in practice.

The space-borne detectors are unequal arm interferometers where the laser frequency noise will experience different time delays when traveling along different arms and cannot be canceled out by itself at the photodetector like ground-based detectors. In order to suppress the laser frequency noise which can be stronger than GW signals a few orders, the technique called time delay interferometer (TDI) where the observables are created by time-shifting and combining single link responses has to be used to construct artificial equal arm interferometers [82,85,86]. The 1.5th generation (or 1st generation in some literatures) TDI observable X is given by Babak et al. [87]

$$\begin{aligned} X_{1.5} &= y_{13} + D_{13} y_{31} + D_{13} D_{31} y_{12} + D_{13} D_{31} D_{12} y_{21} \\ &\quad - y_{12} - D_{12} y_{21} - D_{12} D_{21} y_{13} - D_{12} D_{21} D_{13} y_{31}, \end{aligned} \quad (5)$$

where D_{ij} denotes the delay operator defined as $D_{ij} y_{sr} = y_{sr}(t - L_{ij})$ and L_{ij} is the arm-length between the node i and node j of the constellation. (In the construction of TDI combinations, the non-commutativity and variations of arm-length may be taken into consideration, whereas when actually computing the detector responses, the approximation of rigid constellation where all arm-length is equal and constant is usually adopted [83,87,88]. Therefore, we follow this convention and use L_{ij} to denote the arm-length when writing TDI combinations, while using L in other places.) The other two observables Y, Z can be obtained similarly by cyclic permutation of indices. The 2nd generation TDI incorporates that the delay operators are non-commutative for forward and inverse delay of a link due to the rotation of the constellation by compensating more virtual loops in two arms as [87]

$$\begin{aligned} X_{2.0} &= y_{13} + D_{13} y_{31} + D_{13} D_{31} y_{12} + D_{13} D_{31} D_{12} y_{21} \\ &\quad + D_{13} D_{31} D_{12} D_{21} y_{12} + D_{13} D_{31} D_{12} D_{21} D_{12} y_{21} \end{aligned}$$

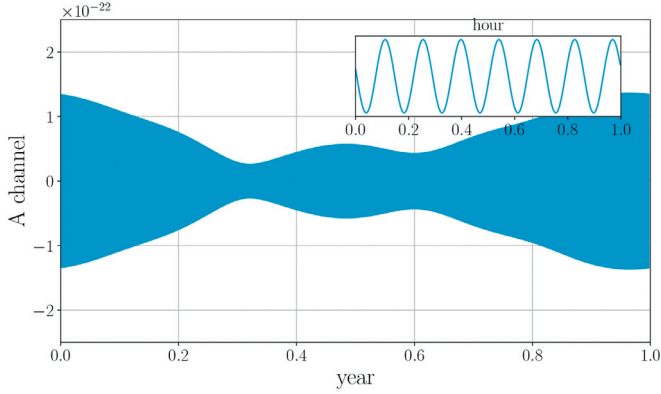


Fig. 2. Time domain signal of a typical GCB source. In the mHz band, the GCBs are at the early inspiral stage of their orbital evolution where the radiated GWs are quasi-monochromatic. However, the motion of detectors can endow signals with annual modulation which depends on the sky locations of the sources. The signal shown here is the A channel of 1.5th generation TDI combination for the responses of LISA to a verification source AM CVn [73]. The signal is generated by *lisaanalysis* tools [93] and *fastlisaresponse* [94].

$$\begin{aligned}
 &+ D_{13}D_{31}D_{12}D_{21}D_{12}D_{21}y_{13} + D_{13}D_{31}D_{12}D_{21}D_{12}D_{21}D_{13}y_{31} \\
 &- y_{12} - D_{12}y_{21} - D_{12}D_{21}y_{13} - D_{12}D_{21}D_{13}y_{31} \\
 &- D_{12}D_{21}D_{13}D_{31}y_{13} - D_{12}D_{21}D_{13}D_{31}D_{13}y_{31} \\
 &- D_{12}D_{21}D_{13}D_{31}D_{13}D_{31}y_{12} - D_{12}D_{21}D_{13}D_{31}D_{13}D_{31}D_{12}y_{21}. \quad (6)
 \end{aligned}$$

The constructions of TDI observable are not unique and have abundant forms. Different constructions can have different sensitivities to GWs of different polarizations and propagation directions [89,90]. The X, Y, Z channels are correlated, and the independent TDI channels can be obtained through the combination of

$$\begin{aligned}
 A &= \frac{1}{\sqrt{2}}(Z - X), \\
 E &= \frac{1}{\sqrt{6}}(X - 2Y + Z), \\
 T &= \frac{1}{\sqrt{3}}(X + Y + Z). \quad (7)
 \end{aligned}$$

In the low-frequency limit where $f < (1/2\pi L)$ with L denoting the arm-length of the detector, signals in the A channel are mainly contributed by the plus polarization, the E channel can be approximated to the cross polarization, and the T channel is approximated to the breath polarization but which is absent in General Relativity. Thus, in data analyses with low-frequency approximation, only A and E channels are considered usually [51,91,92]. We illustrate the time domain signal in the A channel of a typical GCB in Fig. 2 and the frequency domain signals of the GCB population in Fig. 3.

3. Solutions for extracting overlapping GCB signals

This section provides a detailed review of currently available solutions for extracting overlapping GCB signals. We mainly focus on recent works, early efforts with MLDCs are briefly summarized in Section 3.1, and diverse innovations and new implementations in recent years are introduced in the subsequent sections. A summary of solutions mentioned here and corresponding references are presented in Table 2.

3.1. Early researches

Problems of extracting overlapping GCB signals have been discussed for more than two decades. The iterative subtraction strategy was proposed at the beginning of this century [97], where the brightest sources

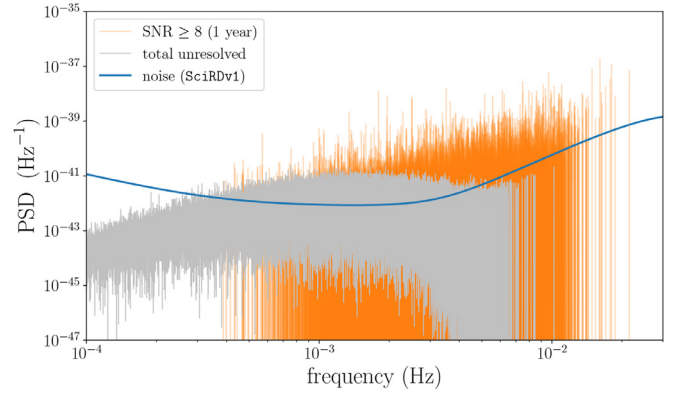


Fig. 3. Signals of the simulated GCB population in frequency domain. The GCB catalog is from the training dataset of LDC2a [35]. The displayed signals are responses in the A channel of 1.5th generation TDI combination of LISA generated by *gbgpu* [77,95]. The orange lines denote the sources that have $\text{SNR} \geq 8$ for 1-year observation and are expected to be resolvable individually. The total contribution of remaining faint sources is denoted by gray lines which can form a stochastic foreground as confusion noise. The noise of the instrument is indicated by the blue line, which is given by the noise model *SciRDv1* [87,96].

are iteratively identified and subtracted. The global fitting strategy can be traced back to the early researches [42,43] which employs the trans-dimensional MCMC sampling algorithm to simultaneously infer the source number, joint posteriors of source parameters, and noise levels.

Diverse solutions for overlapping GCB signals are explosively presented working around the MLDCs. For example, the blocked annealed metropolis (BAM) algorithm developed in [98–100] is a quasi-Bayesian approach with a \mathcal{F} -statistic likelihood and a customized MCMC sampling strategy. The subsequent work [101] extends the BAM algorithm through introducing parallel tempering, the reversible jump MCMC (RJMCMC), and the fast-slow decomposition waveform model, which is the basis of the recent full-scale global fitting pipeline GBMCMC introduced in Section 3.3. The search methods using \mathcal{F} -statistic and various optimal algorithms [102–107] are also the foundation of diverse iterative subtraction solutions in recent years. Additionally, various intelligent ideas besides the above two schemes are also widely discussed, including a tomographic approach [108,109], genetic searches [110], the two-stage strategy [111], and a sophisticated MCMC walking strategy [112,113], etc. More details about early efforts can be found in reports of each MLDC [30,32–34], and a comprehensive review [114].

Based on early efforts, various enhancements, new methods and implementations have emerged with the resurrected LDCs in recent years, which will be detailed in the following.

3.2. Iterative subtraction using \mathcal{F} -statistic and particle swarm optimization

One of widely considered strategies for extracting overlapping signals is iterative subtraction where the search algorithm for one individual source will be run to identify the brightest source, then the identified signal will be subtracted from the data, and this procedure will be iteratively executed with the residual data until some stopping criteria are satisfied. The idea of iterative identification and subtraction for overlapping signals in mHz GW band is proposed early in [97], and is implemented extensively in previous works. In this review, we present an introduction of a typical iterative subtraction scheme developed in recent years [37–39], which employs \mathcal{F} -statistic to construct likelihood and uses particle swarm optimization (PSO) to search the optimal estimations.

The framework developed in [37] is referred to as Galactic Binary Separation by Iterative Extraction and Validation using Extended Range

Table 2

A collection of currently available solutions for extracting GCB signals. Relevant codes if open source are listed in the tablenote for convenience of reference. Most equations appearing in the corresponding chapters can be found in the references listed above.

	Solution	Key techniques	References
Iterative subtraction	GBSIEVER	\mathcal{F} -statistic; PSO.	[37–39]
	LMPSO-CV	\mathcal{F} -statistic; PSO with special particle movement rules; Tuned rules for removing false candidates.	[40]
	Combination of coarse template search and fine PSO search GBMCMC ¹	\mathcal{F} -statistic; PSO with constrained search ranges; Template search.	[41]
Global fitting	Eryn ²	RJMCMC; Parallel tempering; Blocked Gibbs update scheme; \mathcal{F} -statistic proposal; Multimodal proposal; Posterior-based proposal.	[44–47]
	Hybrid Bayesian approach ³	RJMCMC; Parallel tempering; Blocked Gibbs update scheme; Ensemble sampling; Delay rejection; Multiple try metropolis Stretch-move proposal Proposal built from available samples; GPU acceleration.	[48,49]
	Machine learning techniques Early efforts with MLDCs	DE; SLSP; MCMC sampling within promising regions; Gaussian process regression. Normalizing flows	[50–52] [59] [30,32–34,114]

¹ <https://github.com/tlittenberg/ldasoft>.
² <https://github.com/mikekatz04/Eryn>.
³ <https://github.com/stefanstrub/LDC-GB>.

(GBSIEVER). In the iterative subtraction scheme, the single brightest source needs to be identified in each iteration. GBSIEVER implements this through maximum likelihood estimation where the likelihood is constructed by \mathcal{F} -statistic. The data from detectors are composed of noise $n(t)$ and GW signals $h(t)$, which can be expressed as $d(t) = n(t) + h(t)$. Assuming the noise $n(t)$ is Gaussian and stationary, the probability of a realization of $d(t)$ given a specific GW signal $h(t)$ described by a set of parameters θ can be written by

$$\ln p(d(t)|h(t, \theta)) = -\frac{1}{2} \langle d(t) - h(t, \theta), d(t) - h(t, \theta) \rangle, \quad (8)$$

where the angle brackets denote the noise weighted inner product defined as

$$\langle a(t), b(t) \rangle = 4\text{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \quad (9)$$

Here, $\tilde{a}(f)$ and $\tilde{b}(f)$ is the Fourier transformation of time series $a(t)$ and $b(t)$, $S_n(f)$ denotes the PSD of the noise. To identify the signal in data through maximum likelihood estimation, one needs to explore the parameter space of θ to find the parameters $\hat{\theta}_{\text{MLE}}$ where the probability (Eq. 8) has the maximum value. Drop the term independent with the source parameters, the likelihood can be written by

$$\ln \Lambda = \langle d, h \rangle - \frac{1}{2} \langle h, h \rangle. \quad (10)$$

For multiple independent measurements of the same signal, such as the independent TDI channels or different detectors, the final log-likelihood is the sum of log-likelihoods for each independent measurement. The estimation of source parameters $\hat{\theta}_{\text{MLE}}$ is given by

$$\left. \frac{\partial \ln \Lambda}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{MLE}}} = 0. \quad (11)$$

Here, we use the hat to denote parameters with the maximum likelihood value.

Exploring the high-dimension parameter space to solve Eq. 11 is computationally expensive. In the framework GBSIEVER, the reduced likelihood is constructed through \mathcal{F} -statistic, where only four intrinsic parameters are searched and the extrinsic parameters can be obtained analytically. The full set of parameters to describe GCB signals is separated into intrinsic parameters $\{f_0, \dot{f}, \lambda, \beta\}$ and extrinsic parameters $\{\mathcal{A}, \iota, \psi, \phi_0\}$. The GW signal can be decomposed by a linear combination of templates only depending on intrinsic parameters as

$$h(t, \theta) = \sum_i a_i X_i(t, \kappa), \quad (12)$$

where κ denotes a set of intrinsic parameters, a_i is a reparametrization of extrinsic parameters, and basis templates $X_i(t, \kappa)$ can be defined by

the GW signals at 4 sets of specific extrinsic parameters. Using this decomposition, we can construct the vector U which has the element of

$$U_i = \langle d(t), X_i(t, \kappa) \rangle, \quad (13)$$

and the matrix W which has the element of

$$W_{ij} = \langle X_i(t, \kappa), X_j(t, \kappa) \rangle. \quad (14)$$

Using vector U and matrix W together with the vector A consisting of a_i , the likelihood Eq. 10 can be rewritten as

$$\ln \Lambda = AU - \frac{1}{2}AWA^T. \quad (15)$$

Maximizing the original likelihood $\ln \Lambda$ can be substituted by maximizing the \mathcal{F} -statistic defined as

$$\mathcal{F}(\kappa) = U^T W^{-1}U, \quad (16)$$

from which one can obtain the maximum likelihood estimation for intrinsic parameters $\hat{\kappa}$. The estimation of \hat{a} can be obtained analytically through

$$A^T = W^{-1}U. \quad (17)$$

The full set of parameters with the maximum likelihood can be recovered through $\hat{\kappa}$ and \hat{a} .

The remaining question is how to efficiently explore the intrinsic parameter space to maximize $\mathcal{F}(\kappa)$. In the framework of GBSIEVER, PSO is performed to accomplish this task. PSO is proposed by Kennedy and Eberhart [115] and widely used in various optimal problems. PSO is inspired by the social behavior of organisms like bird flocks or fish schools. It utilizes a population (referred to as the swarm) of candidate solutions (referred to as particles) which are moved by the guidance of own experience of each particle and the collective knowledge of the entire swarm to find the optimal solution in the parameter space. The interaction among particles endows the PSO method with more capability for global searching and against trapping in local optima.

Initially, the particles are randomly drawn in the parameter space and assigned with random velocities. The subsequent movement of particles is guided by Gao et al. [40]

$$\begin{aligned} v_i^{t+1} &= \omega v_i^t + c_1 r_1 (P_i^t - x_i^t) + c_2 r_2 (G^t - x_i^t), \\ x_i^{t+1} &= x_i^t + v_i^{t+1}. \end{aligned} \quad (18)$$

In the above equations, x_i^t and v_i^t denote the position and velocity of i th particle at the iteration step of t . P_i^t is the best position of the i th particle found in previous iterations, which represents the experience of individual particles, and G^t is the best position found by the entire swarm in previous iterations, which represents the collective knowledge

of the swarm. The terms corresponding to P^i and G^i are usually called the personal term and the social term respectively in literature. ω , c_1 , and c_2 are constant coefficients that need to be tuned according to the specific problem through experimental runs or empirical knowledge.

The process of optimization can be viewed as two phases, the exploration phase where the particles explore the parameter space expansively and quickly to find better locations, and the exploitation phase where the particles have converged within a promising region and updates of better position will be relatively slow. ω is called the inertia weight controlling the balance of exploration and exploitation in the optimization process. A lower inertia weight favors exploitation and allows particles to quickly converge toward promising regions. Conversely, a higher inertia weight can help particles resist attractions of previous best positions of personal and social terms, explore the parameter space more expansively, and potentially escape from local optima. The coefficients c_1 and c_2 are called acceleration coefficients controlling the balance between the personal term and social term of the particles in their movement. The setting of these coefficients has significant impacts on the performance and efficiency of PSO. A typical configuration in the context of GW data analyses can be found in [116], which is also adopted by GBSIEVER. After sufficient iterations, the particles are expected to converge to the optimal solution.

Some practical techniques are used in GBSIEVER to improve search efficiency and eliminate false candidates. As introduced in Section 2, GCB signals have the characterization of narrowband. Therefore, in practice, it is usually to divide the whole frequency band into small bins. Independent analyses are performed in different frequency bins, and the edge effects are carefully addressed in the meanwhile. Furthermore, another practical technique in GBSIEVER is a special downsampling operation which can reduce the number of samples thus relieve the computational cost in likelihood evaluation. To eliminate spurious sources, GBSIEVER employs a cross-validation procedure to compare the extracted candidates in two independent runs. Only the candidates with similar estimation results in two independent runs are considered to be genuine.

The performance of GBSIEVER is verified with the dataset of LDC1-4 and modified MLDC3.1, where it is demonstrated that $\mathcal{O}(10^4)$ GCBs can be successfully identified [37]. The residual after subtracting identified signals is shown in Fig. 4. In the subsequent work [38], the framework GBSIEVER is extended by incorporating the network of detectors. Furthermore, in the work [39], the fact that the data collection is gradually incremental in real observations is considered. The search results obtained in ahead short period observations can be used to reduce the parameter space in following searches, which can help enhance the efficiency of the algorithm.

Although the framework GBSIEVER has been demonstrated to be capable of extracting a large population of overlapping GCB signals, there are various aspects which can be further improved. In the following, other two implementations of iterative subtraction focusing on reliving inaccurate contamination and enhancing search efficiency will be introduced.

3.2.1. Local maxima particle swarm optimization algorithm

In the iterative subtraction scheme, the inevitable errors of parameter estimation in each iteration will leave residual signals that can continuously accumulate and contaminate remaining data as noise. The work [40] improves the previous GBSIEVER framework and develops the new approach of local maxima particle swarm optimization algorithm with a special strategy of creating voids referred to as LMP-PSO-CV for dealing with inaccurate subtraction contamination, especially for the low SNR sources. LMP-PSO-CV starts with the remaining data assuming all sources with SNR > 15 have been identified and subtracted. The algorithm aims at identifying all local maxima of \mathcal{F} -statistic in the parameter space, and the source parameters will be extracted from these local maxima.

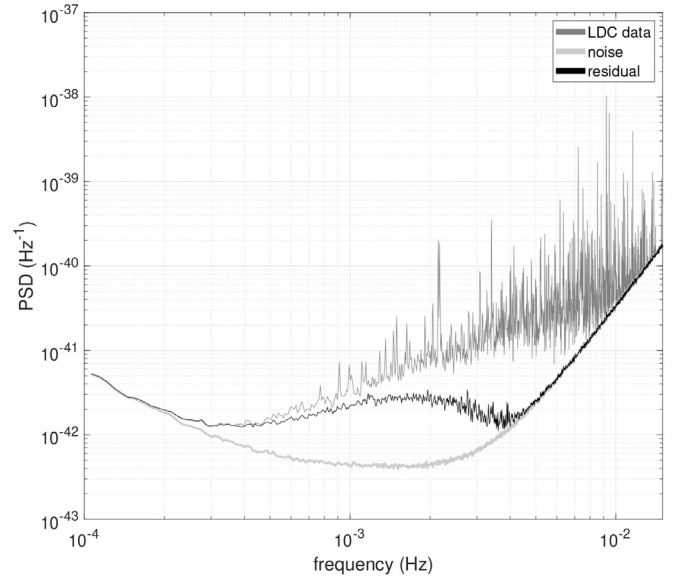


Fig. 4. The residual after subtracting identified signals given by GBSIEVER. Utilizing the iterative subtraction method with the \mathcal{F} -statistic likelihood and the PSO algorithm, GBSIEVER can successfully identify $\mathcal{O}(10^4)$ GCB signals. This figure shows the residual of the A channel of the TDI combination of simulated LISA dataset LDC1-4 which is used to verify the performance of GBSIEVER. This figure is cited from Zhang et al. [37].

In LMP-PSO-CV, to identify the local maxima, the configuration of PSO is adjusted by setting $c_1 = \omega = 0$. The equations guiding the movement of particles turn into the form of [40]

$$\begin{aligned} v_i^{t+1} &= c_2 r_2 (G^i - x_i^t), \\ x_i^{t+1} &= x_i^t + v_i^{t+1}. \end{aligned} \quad (19)$$

As mentioned in Section 3.2, a lower inertia weight favors the exploitation where particles tend to move within a local promising region, and emphasizing the global term will guide particles to move toward the same position. It can be expected that the setting of entirely dropping the inertia weight and the individual term will extremely improve the convergence towards local maxima.

Once a local maximum is identified, to avoid the search algorithm picking the same or too close local maxima multiple times, a void in the parameter space will be created and excluded in subsequent searches. The void is modeled by a spheroid as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad (20)$$

where x, y, z correspond to the source parameter f_0, β, λ . The size of the spheroid is determined by the frequency resolution and the degeneracy of β and λ through experimental computations. In the process of searching for local maxima, if a particle moves into voids, it will be placed at a new random position outside voids in the next iteration to avoid redundancy.

The search of local maxima and creation of voids will be performed repeatedly until the termination rules designed for identifying all local maxima beyond the threshold are satisfied. The source parameters are extracted among the identified voids. As can be seen from Fig. 5, there are enormous local maxima in likelihood surfaces corresponding to false signals which need to be eliminated when extracting real candidates. The extraction consists of three parts. The first part aims at removing local maxima induced by degeneracy noise of individual signals. From the observation of experimental evaluation of the likelihood for an individual signal as shown in the top row of Fig. 5, the local maxima corresponding to false signals have smaller likelihood values than the

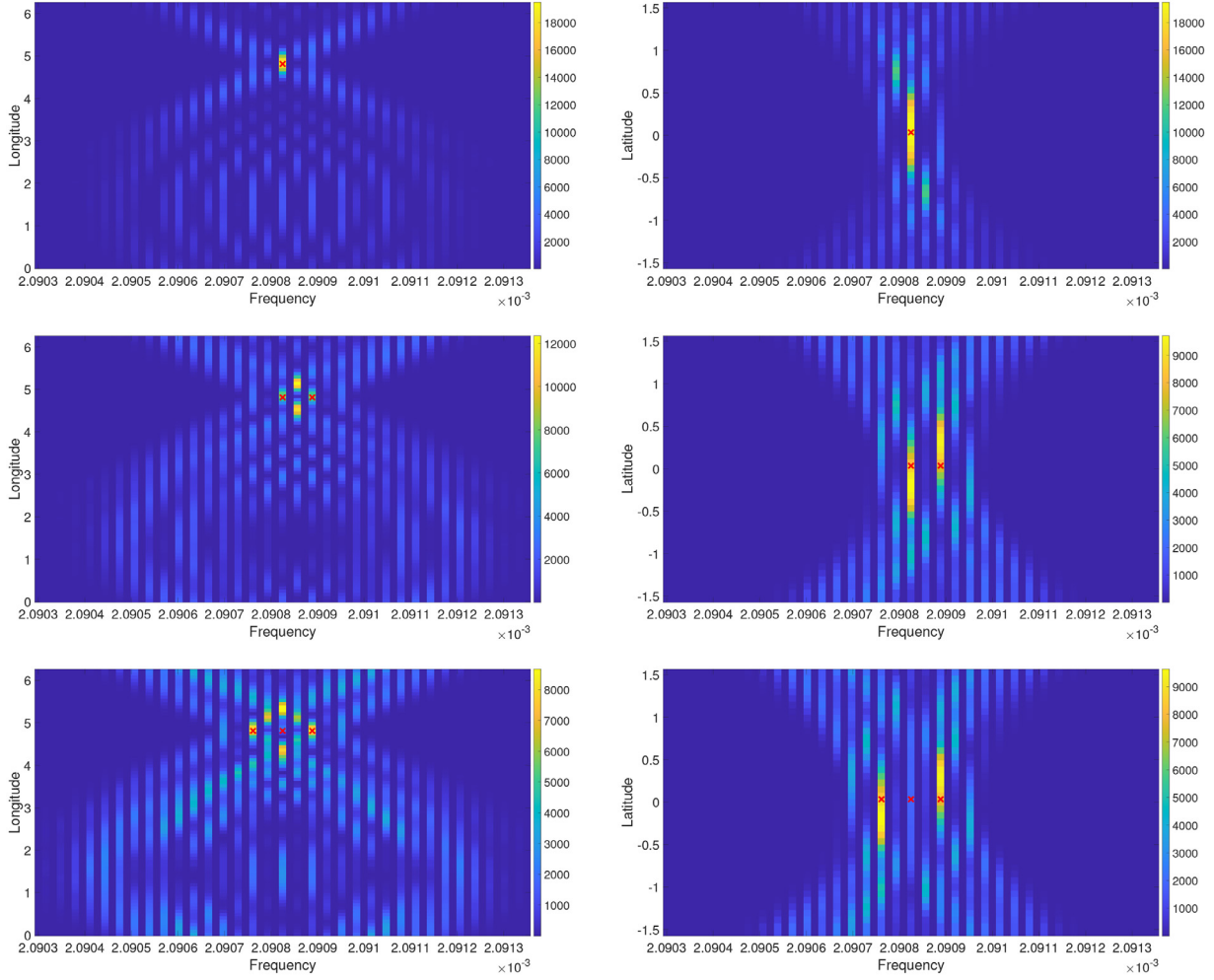


Fig. 5. The likelihood surfaces for cases of an individual signal and overlapping signals. These figures show the value of the \mathcal{F} -statistic likelihood in the parameter space of (f_0, λ, β) . The red crosses denote the injected values, and the colorbars indicate the \mathcal{F} -statistic values. The top row is for the case of an individual signal, the middle and bottom rows are for the cases of two and three overlapping signals. For the overlapping signals with enough small frequency intervals, there will be local maxima in likelihood surfaces corresponding to false candidates induced by correlations among overlapping signals, which may have higher likelihood values than the real candidates. These figures are cited from Gao et al. [40].

real signal. Therefore, the local maxima are sorted in descending order, and the extraction starts with the local maxima of the highest \mathcal{F} -statistic value which is assumed to be a real signal. Then, the quantities of $\Delta\mathcal{F}$ defined as [40]

$$\Delta\mathcal{F}(\theta_i) = \mathcal{F}(\theta_i, d(t)) - \mathcal{F}\left(\theta_i, \sum_{m=1}^{i-1} h(t, \theta_m)\right) \quad (21)$$

are computed for subsequent local maxima. In the above equation, the first term $\mathcal{F}(\theta_i, d(t))$ is the \mathcal{F} -statistic value of the i th local maximum with parameter θ_i with respect to the raw data, which is same to the \mathcal{F} -statistic value used in the search process. The second term $\mathcal{F}\left(\theta_i, \sum_{m=1}^{i-1} h(t, \theta_m)\right)$ is the \mathcal{F} -statistic value of the i th local maximum with respect to the summation of candidate signals previous extracted, which characterizes the degeneracy with previous extracted signals. If the \mathcal{F} -statistic value of the i th local maximum is mainly contributed by the degeneracy with previous extracted signals. The quantity of $\Delta\mathcal{F}$ is expected to be small. Conversely, large $\Delta\mathcal{F}$ indicates the local maximum is likely to be a real signal. The local maxima that satisfy the criterion of $\Delta\mathcal{F} > 53$ will be retained and passed into the second part for further removing false signals.

The second part of signal extraction and false signal elimination is based on the astrophysical properties of GCBs. The second part requires a new independent search of local maxima which is similar to the pro-

cess discussed above except using a different range for f . In the first search, the search range for f is $[-10^{-16}, 10^{-15}]$ when $f \leq 4$ mHz and $[-10^{-14}, 10^{-13}]$ when $f > 4$ mHz. Whereas, in the second search, the range of f is determined by the mass parameters of binaries. Since the majority of GCBs are consisted of binary white dwarfs which have the mass ranging from 0.1 to $1.4M_\odot$. In the second search, the values of f are constrained within the range corresponding to the mass range of white dwarfs. After two independent searches, the candidate signals need to be compared through the correlation coefficient defined as

$$R(\theta, \theta') = \frac{C(\theta, \theta')}{[C(\theta, \theta)C(\theta', \theta')]^{\frac{1}{2}}}, \quad (22)$$

where $C(\theta, \theta') = \langle h(\theta, h(\theta')) \rangle$. This correlation coefficient quantifies how similarity between two signals identified in two independent searches. Only candidates with enough large R are considered as reliable sources and are retained for further processing. Another characterization of GCBs is the concentration in the Galactic disk. Therefore, one can simply remove candidates with Galactic latitude outside of $[-0.5\text{rad}, 0.5\text{rad}]$ to enhance the reliability.

The third part of signal extraction focuses on the false candidates induced by overlapping signal degeneracy noise. In the experimental computation of the likelihood surface for the situation of overlapping signals as shown in the middle and bottom rows of Fig. 5, it is found that the \mathcal{F} -statistic values of local maxima corresponding to false candidates

can be even larger than real signals. To remove false candidates of this kind, a similar process of the first part needs to be performed again but in ascending order of local maxima.

The process of LMPSO-CV can be summarized in two steps. In the first step, the stochastic optimization algorithm is tuned to specializing in finding local maxima. A void in parameter space will be created to avoid redundant search once a local maximum is identified. The second step extracts signals and eliminates false candidates from the identified local maxima, which consists of three parts: removing false candidates induced by degeneracy of individual signals, applying the astrophysical properties of GCBs, and removing false candidates induced by degeneracy of overlapping signals. LMPSO-CV mainly focuses on the sources with SNR between 7 and 15. Since the LMPSO-CV approach searches all local maxima simultaneously and filters out false candidates afterward, rather than iteratively performing search and subtraction for a single signal, it can better deal with inaccurate subtraction contamination in the traditional iteration subtraction scheme.

3.2.2. Combination of coarse template search and fine PSO search

Another implementation of the iterative subtraction scheme focuses on improving search efficiency by the strategy of combination with the coarse search based on templates for quickly identifying rough parameters of candidates and the fine search based on F -statistic and PSO for detailed exploring small promising regions [41].

The strategy is composed of two steps. In the first step, candidates are searched by a template bank. A stochastic template bank [117] is constructed, where the places of template parameters are randomly drawn in the parameter space and additional pruning operation is performed to remove templates with too small separation. The number of templates can be approximated by Lu et al. [41]

$$N(\eta, m_*, S_n) \approx \frac{V_{S_n}}{V_n} \ln \left(\frac{1}{1-\eta} \right) m_*^{-n/2}, \quad (23)$$

where n denotes the dimension of the parameter space S_n , η denotes the desired level of coverage confidence, m_* is the mismatch criterion, V_n is the volume of unit sphere in n -dimension, and V_{S_n} is the volume of the whole parameter space. The construction of the template bank starts with randomly generating N waveforms. Then, the KDtree (k-dimensional tree) algorithm [118,119] is used to find the nearest neighbor for each waveform. If the mismatch between two waveforms is less than m_* , one of them will be removed. The above two procedures are performed repeatedly until the total number of templates gets to stable. The F -statistic likelihood is also used in the coarse search, the threshold for coarse search is set to $F_{\text{th,coarse}} = 15$ which corresponds to SNR ≈ 5.1 . The templates whose F -statistic values exceed the threshold are recorded as candidates. Due to the annual orbital motion of detectors, modulation is imposed on the quasi-monochromatic signals of GCBs. The modulation can induce sidebands around the central frequency f_0 of GCB signals. A clustering operation will be performed to gather candidates with close central frequency. Among all candidates within the extended Doppler window, only the sources with the largest F -statistic values are retained for fine search in the next step.

In the second step, the parameters of candidates identified by the template search are accurately estimated using PSO, which is similar to the procedure discussed before, while the search ranges for the central frequency are tuned according to results given by coarse search.

The performance of this methodology is demonstrated with MLDC3.1. It is shown that $\mathcal{O}(10^4)$ sources can be successfully identified, and nearly 90 percent of them are well aligned with injected signals. This method performs a coarse template search in the first step, which can provide priori information guiding the PSO search thus enhancing the efficiency of the fine parameter estimation.

3.3. Global fitting with trans-dimensional Bayesian inference

The iterative subtraction method has advantages of rapidity and efficiency but also suffers from problems of correlations among overlapping

signals and contamination of inaccurate subtraction. In each step of iterative subtraction, errors in parameter estimation are unavoidable, which can yield signal residuals left behind the subtraction. The residuals will contaminate remaining data and accumulate along with iterations. Furthermore, in each iteration only the parameter values with maximum likelihood are used in subtraction, the uncertainty of parameter fitting will propagate alone iterations. The uncertainty of parameter estimation for identified candidates is difficult to ascertain. Additionally, GCB signals are heavily overlapping in both time and frequency, which can induce high correlations among signals. The correlations may lead to bias in the parameter estimation of each iteration, and this will further intensify the problem of inaccurate subtraction contamination.

To alleviate these problems, a commonly used operation is dividing the whole frequency band into small bins and performing analyses independently in different bins while carefully addressing signals residing at edges, which can reduce the number of iterations and reduce the accumulation of inaccurate subtraction contamination. Besides, as introduced in Section 3.2.1, the method of modifying the particle movement rule in PSO aiming at identifying all local maxima on the likelihood surface simultaneously is also an effective method to address inaccurate subtraction contamination for the low SNR sources.

The global fitting is another route that can effectively deal with the problems of overlapping signal correlations. In the iterative subtraction scheme, only parameters for one source are estimated in each step. In contrast, parameters for all sources together with the source number are estimated simultaneously in the global fitting scheme. The global fitting with full Bayesian parameter estimation can obtain joint posterior distributions rather than just the maximum likelihood estimation, thus the uncertainty of fitting can be read out from posteriors straightforwardly. Since all sources are fitted simultaneously, the correlations of overlapping signals are taken into account and can be reflected in joint posteriors of multiple overlapping signals.

As mentioned in 3.1, the strategy of global fitting and the Bayesian approach have been proposed and implemented early in various works focusing on MLDCs [42,43,98–101]. Further improvements like incorporating the reality that the data collection is time evolving, various sophisticated methods for increasing the efficiency of MCMC sampling, and new implementations of RJMCMC have been continuously presented in recent years [44–46,48,49].

3.3.1. RJMCMC

For the Bayesian inference of global fitting, One of the key differences with the Bayesian parameter estimation commonly used in data analyses of current ground-based detectors is that the number of sources is uncertain and has to be inferred from the given data. A single GW source signal $h(t, \theta)$ in the likelihood shown in Eq. 8 turns to a summation of multiple signals $\sum_k h(t, \theta_k)$ where the number of source is uncertain. The sampling algorithms to solve inference problems with uncertain dimensionality are referred to as trans-dimensional MCMC or RJMCMC.

In the parameter estimation problem of the Bayesian framework, the estimation results are represented by posteriors through the Bayes theorem

$$p(\theta_n | d(t)) = \frac{\pi(\theta_n) p(d(t) | \theta_n)}{\mathcal{Z}}, \quad (24)$$

where $d(t)$ denotes observed data, θ_n represents parameters with the dimensionality of n , $\pi(\theta_n)$ is the prior representing the knowledge before observations, $p(d(t) | \theta_n)$ is the likelihood representing the probability of noise realization that can just present the time series $d(t)$ observed on detectors when adding the GW signals described by parameters θ_n and depending on the models of noise behavior and GW signals, \mathcal{Z} is the normalization factor called as evidence. Since the dimension of parameters θ_n for describing GW signals is usually high, it is impractical to compute posterior on a grid of the parameter space. Stochastic sampling algorithms are usually employed to perform random walking in the pa-

parameter space, and use the density distributions of random samples to approximate the probability distributions of posteriors.

The MetropolisHasting MCMC (MHMCMC) [120,121] is one of the well known and widely used algorithms to perform sampling for fixed dimensional problems. After setting the initial state of a random walker, the subsequent movements are guided by the rules introduced below. If current state is θ_n^i , a new state can be drawn from a proposal distribution $q(\theta_n^{i+1}|\theta_n^i)$. Then the acceptance probability of the new state is evaluated by

$$\alpha = \min \left[1, \frac{p(\theta_n^{i+1}|d) q(\theta_n^i|\theta_n^{i+1})}{p(\theta_n^i|d) q(\theta_n^{i+1}|\theta_n^i)} \right]. \quad (25)$$

The probability of whether actually moving from current state θ_n^i to new proposed state θ_n^{i+1} is determined by α . It can be proven that after sufficient walking, the equilibrium distribution of the walker will converge to the target posterior distribution $p(\theta_n|d)$. Although, in practical problems, the MHMCMC usually needs to be modified in various aspects to increase sampling efficiency or avoid bias, the above description presents the most basic conception of the MCMC sampling algorithm for fixed dimensional problems.

In problems of uncertain dimensionality, to allow the walker to jump between different parameter spaces, the RJMCMC algorithm draws new proposals by a different procedure [122,123]. Firstly, a random vector \mathbf{u} with the dimension of r is generated from a chosen probability distribution $g(\mathbf{u})$. Then, the proposal state is generated through a transformation

$$\theta_{n'}^{i+1} = f(\theta_n^i, \mathbf{u}). \quad (26)$$

The corresponding inverse transformation is given by

$$\theta_n^i = f^{-1}(\theta_{n'}^{i+1}, \mathbf{u}'), \quad (27)$$

where \mathbf{u}' is a vector with size r' and satisfying $n+r = n'+r'$. The only requirement for the transformation is that f and f^{-1} need to be differentiable. The dimensions of parameter space n and n' do not need to be the same, and even the parameterization associated with θ_n^i and $\theta_{n'}^{i+1}$ can be different. The acceptance ratio for the new proposed state is given by

$$\alpha = \min \left[1, \frac{p(\theta_{n'}^{i+1}|d) g'(\mathbf{u}')}{p(\theta_n^i|d) g(\mathbf{u})} |\mathbf{J}| \right]. \quad (28)$$

The term \mathbf{J} is the Jacobian defined by

$$\mathbf{J} = \left| \frac{\partial(\theta_{n'}^{i+1}, \mathbf{u}')}{\partial(\theta_n^i, \mathbf{u})} \right| \quad (29)$$

to account for the change of volume element under the transformation of parameter spaces. In general, the Jacobian can be difficult to compute. However, in the situation of global fitting for GCB signals, where it only involves adding or removing GCB signals with the same set of parameters, the Jacobian can be easily computed by the ratio of prior volumes of parameter spaces before and after the transformation. The walker of RJMCMC can freely explore the entire possible parameter space including not only parameter spaces of individual sources but also parameter spaces for all possible source numbers. The estimation of source number can be naturally obtained through comparing the iteration numbers of the walker lingered in the different parameter spaces, and the Bayesian factor for the assumptions of different source numbers can be given by the ratio of the iteration numbers in different parameter spaces.

Although the movement rule of RJMCMC can be explicitly presented as above, the implementation of the RJMCMC algorithm in practice to solve the problem of overlapping GCB signals is very challenging. On the one hand, since all sources are simultaneously fitted, the parameter spaces are extremely high dimensional. Even the whole frequency band is usually divided into small bins and different bins can be analyzed independently after carefully addressing edge effects. The source number in a single small bin can still be considerable. For example, the default setting of the prior range for the source number in each bin is [0, 30] in

the work [44]. As mentioned in Section 2, a GCB signal is characterized by 8 parameters. The maximal dimension of parameter space can be 240. And the posterior can have the feature of multimodality, which requires the sampler has the ability to deal with complicated likelihood surfaces in high-dimensional parameter spaces. On the other hand, the uncertainty of source number requires the sampler to jump between different parameter spaces, which further extremely expands the space needed to explore. Furthermore, since the likelihood in vast regions may be very small. Only if the proposal is enough close to the true values, it can likely be accepted. Generating proposals entirely uniformly may lead extremely low acceptance rate, especially for between-model proposals. One need to design good strategies for random walking to increase the sampling efficiency.

3.3.2. Global fitting pipelines

Currently, available full-scale and end-to-end global fitting pipelines for GCBs include GBMCMC [44–46] and Eryn [48,49]. Various intelligence methods are utilized to overcome difficulties in the global fitting with RJMCMC. For example, the method of parallel tempering [124–126] is used in both GBMCMC and Eryn. In parallel tempering, multiple walkers are randomly moving in parallel at different temperatures T . The likelihood is modified as

$$p_T(\theta|d) \propto \pi(\theta)p(d|\theta)^{1/T}, \quad (30)$$

where $\pi(\theta)$ is the prior and $p(d|\theta)$ is the original likelihood function. When $T = 1$, $p_T(\theta|d)$ returns to the target posterior distribution. When $T \rightarrow \infty$, $p_T(\theta|d)$ approaches to the prior distribution. Higher temperatures correspond to more exploration where walkers can escape from local optima more easily, while lower temperatures correspond to more exploitation where walkers can sample the small promising regions in more detail. During the random walking, state exchanges are attempted between walkers at different temperatures. The probability of acceptance of exchange proposals is determined by the ratio

$$\alpha = \min \left[1, \frac{p_{T_i}(\theta_i)p_{T_j}(\theta_j)}{p_{T_j}(\theta_j)p_{T_i}(\theta_i)} \right]. \quad (31)$$

The detailed balance is maintained under this state exchange between walkers, which ensures that the equilibrium distribution of walkers can converge to the target posterior distributions. The temperature ladder needs to be chosen for maximizing the information flow among walkers at different temperatures. Ideally, one may expect an equal acceptance ratio between every pair of walkers with neighboring temperatures. The GBMCMC and Eryn adopt the scheme presented in [124] to set the temperature ladder, where the temperature space is dynamically adjusted in the initial burn-in phase to obtain a stable configuration for the rest of random walking. The parallel tempering mechanism can help samplers efficiently explore the complicated likelihood surface with high multimodality.

The strategies for drawing proposals play a crucial role in the sampling efficiency. Customized proposal distributions are developed in GBMCMC to enhance the sampling efficiency. Ideal proposal distributions would be identical to posterior distributions or likelihood functions. The likelihood of multiple overlapping signals consists of three parts, the correlation between each signal and noise, the correlation of noise itself, and the cross-correlations among overlapping signals. As discussed in [44], high correlations among overlapping signals are relatively rare. Therefore, one can enhance proposal distributions focusing on individual signals. Although the cross-correlations of overlapping signals are neglected when designing the proposal distributions, the sampler still explores the joint parameter space of multiple overlapping signals which incorporates cross-correlations of overlapping signals.

The design of proposal distribution in GBMCMC utilizes the F -statistic, the feature of multimodality due to orbital motion of detectors, and posteriors from former epochs of observation. The method of F -statistic is widely used in various implementations of iterative subtraction as presented in previous sections, it can also be used here to

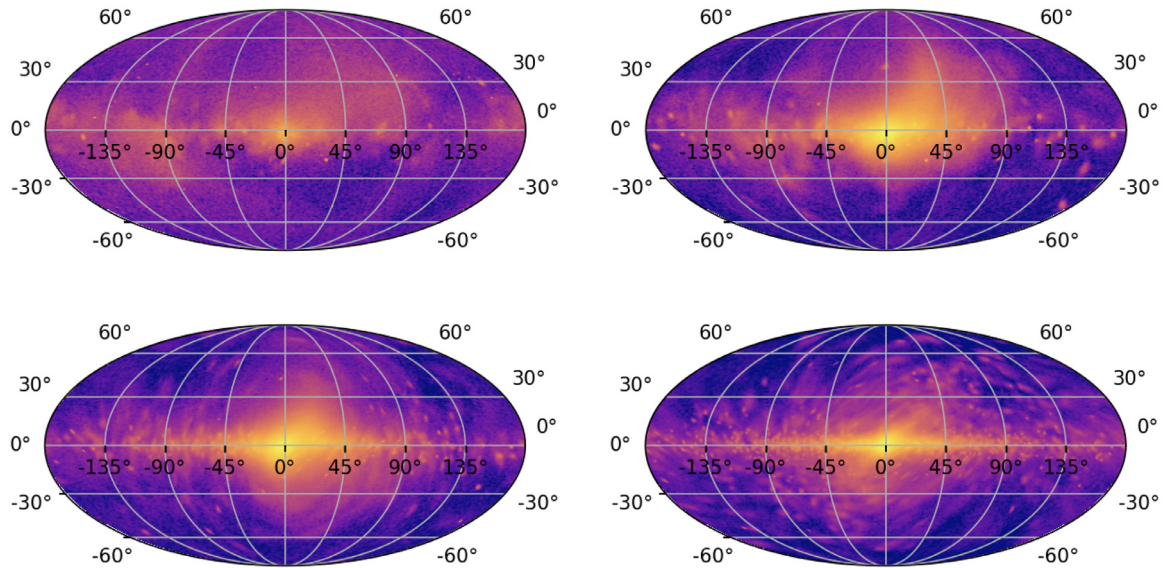


Fig. 6. Posteriors of sky location of identified GCBs in the Galactic coordinate given by GBMCMC. The subplots correspond to the observations of 1.5 (top left), 3 (top right), 6 (bottom left), and 12 (bottom right) months. These figures are cited from Lackeos et al. [46].

construct the proposal distribution for a single source. GBMCMC uses the \mathcal{F} -statistic to build proposals for parameters (f_0, λ, β) which are pre-computed on a grid with spaces determined by estimation of the Fisher matrix. The values of \mathcal{F} -statistic can be approximated to the original likelihood values, thus the \mathcal{F} -statistic proposals are expected to have a high accepted rate. As mentioned in Section 3.2.2, the modulation induced by orbital motion can cause sidebands around the central frequency f_0 of quasi-monochromatic GCB signals. The likelihood surface have the feature of multiple modes around f_0 . GBMCMC designs a dedicated proposal to update the frequency parameter by shifting f_0 with modulation frequency f_m as $f_0 \rightarrow f_0 + n f_m$, where $f_m = 1/\text{year}$. The multimodal proposal can address this known degeneracy and help to improve the convergence of the sampler. In reality, the data collection is gradually incremental. It is impractical to wait until after the finish of entire observation to analyze the obtained data. Data analyses must be performed accompanying with data accumulation. Therefore, one can utilize posteriors obtained from former periods of observation to construct posterior-based proposal distributions in parameter estimation with more observed data, which can significantly improve the sampler convergence. The results of GBMCMC with simulated GCB dataset are demonstrated in Fig. 6 where the posteriors of sky location of identified GCBs are plotted. It can be seen that with the incremental data, the GCBs can be better localized and the structure of the Galaxy is clearer indicated.

Another independent implementation of RJMCMC referred to as Eryn [48,49] which adopts different sophisticated mechanisms in stochastic sampling to overcome poor convergence in RJMCMC. Eryn is based on the ensemble sampler emcee [127] which uses multiple interacting walkers exploring the parameter space simultaneously with the so-called stretch-move proposal to enhance the efficiency and convergence of the sampling. Whereas, in order to address the trans-dimensional movement and the heavily multimodal likelihood surface of overlapping GCBs, Eryn extends the stretch-move proposal in origin emcee sampler to the group proposal which sets a stationary group of walkers and makes the state updates likely to happen within one same mode in the likelihood surface. Furthermore, Eryn build efficient proposals through a data-driven approach, where proposals can be drawn from a fitted distribution constructed by posteriors obtained in burn-in runs with residual data after subtracting bright sources. To overcome the problem of low acceptance rate, Eryn implements two mechanisms called delayed rejection [128,129] and multiple try metropolis

[130–133]. However, due to the high computational requirement of delayed rejection, although this mechanism is implemented in Eryn, it is not currently used in solving the problem of global fitting for GCBs. Another important feature of Eryn is the utilization of GPU which reduces the wall-time to perform the whole analysis by parallel computing on contemporary computational hardware.

The ultimate goal of global fitting is simultaneously analyzing all kinds of sources contained in data including MBHBs, EMRIs, and even unmodeled signals, etc. In the work [45] and [49], GBMCMC and Eryn are incorporated within global fitting pipelines that can handle blended data of MBHBs and GCBs. Global fitting for multiple source types can be realized by the blocked Gibbs scheme considering that correlations among different source types are relatively small. Parameter estimations for different source types can be implemented in separate modules, and these modules are assembled by a wheel update strategy where one can perform updates in one module conditioning on fixed other modules and all modules are updated periodically. Fig. 7 shows the results given by a global fitting pipeline incorporating GCBs and MBHBs where the fitting of GCBs is accomplished by Eryn.

In a short summary, fitting overlapping GCB signals through the full-Bayesian approach can be realized by RJMCMC. Practical implementations of RJMCMC have to overcome various difficulties such as the heavy multimodality of likelihood surface in vast parameter space, low acceptance rate of trans-dimensional proposals, and poor convergence of random walker, etc. which requires elaborate MCMC algorithms. Prototype global fitting pipelines for GCBs have been accomplished by GBMCMC and Eryn which can well address simulated data and are successfully incorporated within the global fitting pipelines for blended data of MBHBs and GCBs. The global fitting with RJMCMC is a full Bayesian approach and can provide joint posteriors of overlapping signals, which can well account for correlations among overlapping GCBs and avoid inaccurate subtraction contamination in the iterative subtraction scheme. Whereas the full Bayesian approach requires massive computational cost and complicated MCMC algorithms which are difficult to implement.

3.4. Hybrid Bayesian approach

The iterative subtraction strategy offers solutions with high efficiency where the computation burden is modest and analysis can be finished in a relatively short time. Whereas, in each iteration, only the maximum likelihood estimation for a single source is extracted, the cor-

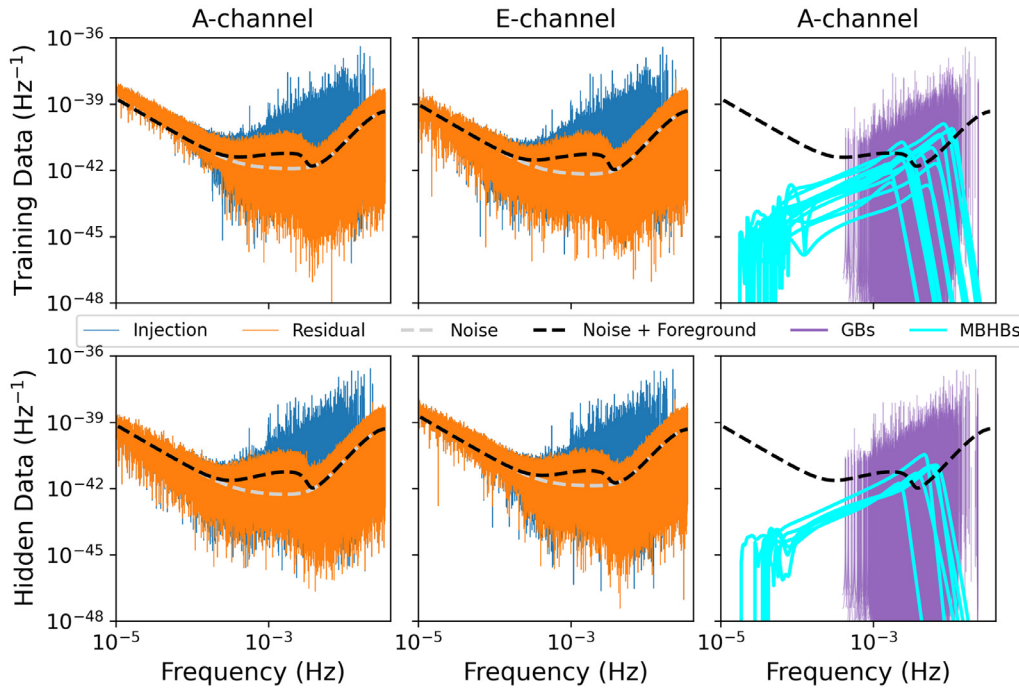


Fig. 7. Results given by a global fitting pipeline incorporating GCBs and MBHBs. The top and bottom rows show the results of training and hidden datasets in the LDC2a respectively. The left and middle columns present the injection data, the residual after subtracting identified signals, and the fitting of noise curves. The identified signals are illustrated in the right column. This figure is cited from Katz et al. [49].

relation of overlapping signals can not be well accounted for. The inaccuracy of estimation in each iteration will contaminate the remaining data. And uncertainty analyses for the identified signals are difficult. The full Bayesian approach with RJMCMC where overlapping signals are fitted simultaneously can effectively overcome these difficulties, but with the cost of extremely massive demand on computational resources. In the works [50–52], a hybrid approach where the maximum likelihood estimation is performed first to find the approximate values of signal parameters, and then the MCMC sampling is used to obtain posteriors of signal parameters is proposed aiming at combining the advantages of the maximum likelihood estimation and the Bayesian parameter estimation while evading their drawbacks.

Similar to the solutions of iterative subtraction introduced previously, the hybrid Bayesian approach also needs to identify the maxima of likelihood in the first step. However, the methods used for searching maxima are different. In this hybrid Bayesian approach, the off-the-shelf algorithms of the differential evolution (DE) [134,135] and the sequential least squared programming (SLSP) [136] implemented in the `scipy` library [137] are adopted for searching the maxima in the likelihood surface. The DE method is used to identify the candidates at first through the iterative subtraction scheme, where only one signal is fitted each time the fitting is repeatedly performed with the remaining data after subtracting the best-fit signal. Then, in order to address the correlation of overlapping signals, using all found candidates as the start, the global optimizations are performed by the SLSP method to search the maximum likelihood in the joint parameter space of all candidates.

The DE algorithm is a simple and efficient method for optimization problems that searches the optima by manipulating a population of potential solutions similar to PSO introduced in Section 3.2. The DE algorithm is initialized by a population of candidate solutions drawn randomly within the search space. Then, in following iterations, the algorithm generates new potential solutions by combining and mutating existing individuals in the population. There are various strategies [135] for generating new solutions and updating populations, which may suit different problems and need to be determined through experimental runs. The populations will be continuously updated by replac-

ing individuals with better fitness and eventually converge to the optima. The SLSQ method is a gradient-based optimization method suits for problems with smooth and continuously differentiable likelihoods. After the iterative subtraction search with the DE algorithm, the SLSQ method starts a global search with the initial values given by the DE algorithm, and iteratively steps towards a better solution which incorporates correlations of overlapping signals by the line search in a direction found through derivatives of the likelihood surface. As examples, the true and recovered signals on a small frequency band are illustrated in Fig. 8.

In the second step, The MCMC sampling algorithm is performed to obtain the posterior distribution for each identified signal. The MCMC algorithm for individual sources is similar to methods widely used for parameter estimation in data analysis of ground-based detectors [138,139], except that the data used in likelihood evaluation are the remains after the subtraction of identified signals in the first step excluding the signals to be analyzed, which can be expressed by

$$d_{\text{posterior}}^{(i)} = d_{\text{original}} - \sum_{\hat{\theta} \in \hat{\theta}_{\text{recovered}}} h(\hat{\theta}) + h(\hat{\theta}_i), \quad (32)$$

where $d_{\text{posterior}}^{(i)}$ is the data used in likelihood evaluation for the source i , and $\hat{\theta}_i$ denotes the parameters of maximum likelihood estimation obtained in the first step for the source i . One drawback is that the above MCMC sampling method for individual signals cannot account for the correlation of overlapping signals, which may lead to overoptimistic posterior distributions. To address this, some level of residual signals are intentionally left in the data used for estimating the noise characterization. The partial residual data used to obtain the noise PSD can be expressed as

$$d_{\text{partial}} = d_{\text{original}} - s_{\text{partial}} \sum_{\hat{\theta} \in \hat{\theta}_{\text{recovered}}} h(\hat{\theta}), \quad (33)$$

where s_{partial} is a factor from the the range of [0, 1] which has to be determined experimentally and is set to $s_{\text{partial}} = 0.7$ in the work [52].

Various methods are used to improve the efficiency of MCMC sampling in the second step, for example, constraining the search space only

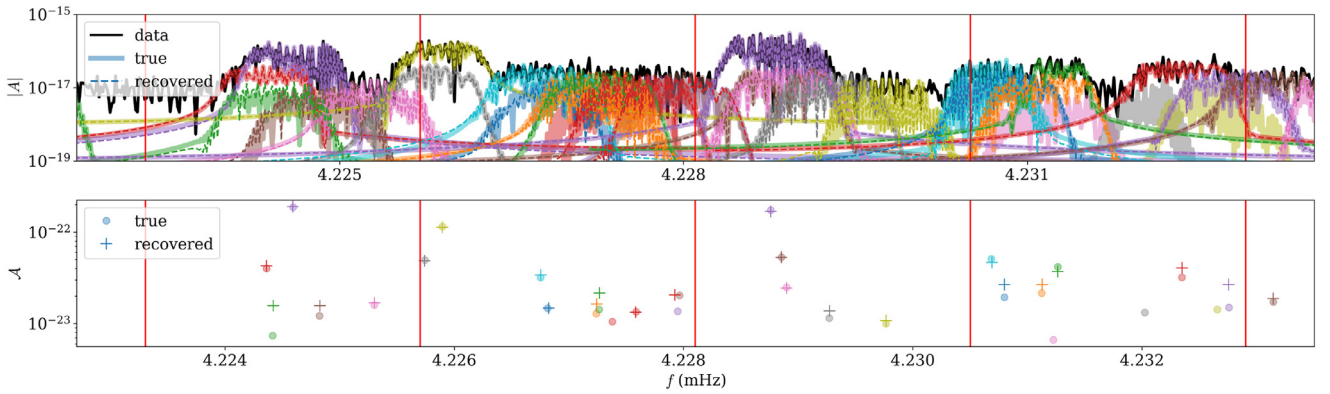


Fig. 8. Recovered GCBs by the hybrid Bayesian approach on four frequency segments. The top panel shows original data, injected signals, and recovered signals in the A TDI channel. The bottom panel illustrates the true and recovered values for the amplitude parameter of each source. The red vertical lines denote the boundaries of frequency segments which are divided for the convenience of extraction of GCBs considering the quasi-monochromatic feature of GCB signals. This figure is cited from Strub et al. [52].

within the promising region [52], or using Gaussian process regression to model the likelihood [51]. The posterior distributions are typically concentrated within small regions of parameter space. To relieve the computational burden, one can only explore the reduced parameter space based on the maximum likelihood estimation obtained in the first step. The Fisher matrix is used to determine the boundaries of the reduced parameter space. The Fisher matrix is widely used in forecast works [140,141], which is an approximation of the Bayesian method under the assumption of high SNR and can provide estimations for parameter measurement uncertainty. The Fisher matrix is given by

$$F_{ij} = \left\langle \frac{\partial h(\hat{\theta})}{\partial \theta_i}, \frac{\partial h(\hat{\theta})}{\partial \theta_j} \right\rangle, \quad (34)$$

where $\hat{\theta}$ is the maximum likelihood estimation obtained in the first step, and the angle brackets are the inner product defined as Eq. 9. The inverse of the Fisher matrix presents the estimation of the covariance matrix of parameter measurement. The boundaries of reduced parameter space for the MCMC sampling are determined by the variance of parameters estimated through the Fisher matrix. Typically, $3\text{-}\sigma$ regions are sufficient, while the practical settings need to be adjusted according to the features of posterior distributions in experimental runs, and the tolerance of computational burden or the desired coverage of the parameter space.

Additionally, in the procedure of MCMC sampling, the likelihood has to be evaluated a huge number of times. The computational cost for likelihood evaluation is one of the main barriers to MCMC sampling. In the work [52], the contemporary computational hardware is used to perform the likelihood evaluation. The likelihood can be computed on GPU in massively parallel, which can significantly reduce the wall-time required for finishing the MCMC sampling [77,94]. On the other hand, as shown in the work [51] the likelihood can be modeled by the Gaussian process regression, where the likelihood is modeled by a joint Gaussian distribution whose mean vector and covariance matrix are determined by training samples [142]. The computation of Gaussian distributions can be much faster than the likelihood defined through the inner product in Eq. 8. In the work [51], 1000 random samples are first drawn for training the Gaussian process regression model, and 500 samples are used for verification. After this, the likelihood is replaced by the Gaussian process regression model in following MCMC sampling, which can reduce the computational cost of the entire sampling process. The training and evaluating of the Gaussian process regression model are implemented through the `scikit-learn` package [143].

The hybrid Bayesian approach combines the maximum likelihood estimation and the MCMC sampling. A search of the iterative subtraction scheme is first performed to identify potential candidates which are used as the start in the subsequent global optimization in the joint

parameter space for all overlapping signals. Then the MCMC sampling algorithm is performed for each identified signal to obtain the posterior distributions. This hybrid Bayesian approach uses the strategy of iterative subtraction to determine the source number and the point estimation for source parameters, which can avoid the computationally expensive trans-dimensional MCMC sampling. Meanwhile, after the iterative subtraction search, the source parameters are again globally optimized around the identified candidates, which can rectify errors induced by inaccurate subtraction contamination and correlations of overlapping signals. However, since the MCMC sampling is performed for individual signals, only the marginalized posteriors can be obtained for each source. Incorporating additional uncertainty induced by signal overlapping requires tuning the algorithm configuration empirically.

3.5. Utilization of machine learning techniques

The development of machine learning techniques has led to revolutions in many fields including the data analysis of GWs. Various machine learning algorithms have been successfully used in tasks of signal identification and classification [144–146], parameter estimation [53–58], noise reduction [147], waveform modeling [148,149], etc. Comprehensive reviews can be found in [150,151]. Utilizing machine learning techniques in data analyses of space-borne detectors is also actively discussed [152–156]. A preliminary attempt of using the machine learning method to extract overlapping GCB signals is presented in the work [59] where normalizing flows are used to build proposals for improving the convergence of MCMC sampling and offer a new method for sharing analysis results.

Normalizing flows are a class of machine learning algorithms that can model complicated distributions and are widely used in density estimation problems [157–160]. Normalizing flows model a probability distribution through an invertible and differentiable transformation $f: \mathcal{X} \rightarrow \mathcal{Y}$ that can map the simple and tractable base probability distribution $P_{\mathcal{X}}(\mathbf{x})$ to the complicated target probability distribution $P_{\mathcal{Y}}(\mathbf{y})$. The target distribution can be given by

$$P_{\mathcal{Y}}(\mathbf{y}) = P_{\mathcal{X}}(f^{-1}(\mathbf{y})) |\det J_{f^{-1}}(\mathbf{y})| \quad (35)$$

where $|\det J_{f^{-1}}(\mathbf{y})|$ is the determinant of the Jacobian of the transformation f^{-1} accounting for the change of volume elements of parameter spaces. The transformation is parameterized by neural networks and can be composed of a sequence of transformations. Various kinds of transformations are developed, a comprehensive review can be found in [158]. In the training procedure, the random samples are drawn from the known target distribution, and the neural network is trained by mapping these samples into the simple base distribution. In the inference

procedure, the above process is performed inversely, the target distribution is obtained by transforming the random samples drawn from the base distribution through the trained neural network. Drawing samples from the base distribution and computing the transformation can usually be much faster than computing the likelihood defined by the inner product. Once the training procedure is finished, normalizing flows can generate posterior distribution more efficiently than the MCMC sampling.

In the work [59], the neural density estimation with normalizing flows is used in three different aspects. In the first, the normalizing flows are used to build physical priors for amplitude and sky location based on population models. The spatial distribution of GCBs mainly concentrates on the Galactic disk, rather than uniformly distributes on the whole sky. The amplitude depends on the binary masses, distance, and orbital period which can also be provided by population synthesis models [21,161–163]. Utilizing available information can help to constrain the parameter space and improve the efficiency of MCMC sampling. Training with the simulated GCB catalog in the LDC2a [35], the physical priors for amplitude and sky location can be constructed through normalizing flows.

In the second aspect, normalizing flows are used to construct proposal distributions from results given by former epoch observations. Constructing proposal distributions from available samples to improve sampling efficiency has been proposed earlier in [164] where the density fit method of kernel density estimation (KDE) is used to build proposal distributions. However, the KDE method has drawbacks for high-dimensional problems where one needs to divide parameters into different groups and build KDE proposal distributions in low-dimensional parameter subspaces. Normalizing flows provide an alternative way to build proposal distributions from available samples with its capability of modeling complicated distributions. Available samples used to build proposal distributions can be obtained in two cases. As mentioned before, the data are incrementally collected, and it is impractical to analyze data waiting until the end of space-borne detector missions. In reality, data analyses need to be repeatedly performed with growing data. It is essential to fully utilize the results obtained previously when analyzing updated data. Normalizing flows can fit the posterior distributions obtained in previous analyses, and be used to draw proposals in subsequent analyses, which is expected to have high acceptance rate and low autocorrelation of chains. Meanwhile, for the trans-dimensional MCMC, by utilizing the capability of normalizing flows to model complicated distribution, the proposal distributions can be constructed through candidates identified by an ahead iterative subtraction procedure.

Thirdly, the result posteriors can be published through normalizing flows. As mentioned in Section 2, the number of resolvable GCBs is expected to be $\sim \mathcal{O}(10^4)$. Moreover, due to the overlap among signals, parameters of different sources may have correlations which can not be represented by marginalized posteriors for individual sources. One may need a joint posterior for multiple sources. Using samples to represent posteriors may have higher demands for data transfer and storage, and will be inconvenient when sharing the analysis results with the community. Normalizing flows offer an effective way to model complicated distributions. Therefore, the results sharing and data product publishing can be in the form of normalizing flow models trained by original posterior samples, from which users can generate samples of the identical distribution for any number needed.

In summary, normalizing flows offer a tool to fit arbitrary complicated distributions, which can be used to construct physical priors from simulated GCBs catalogs, build proposal distributions from available samples, and as a new representation of posteriors to share with the community substituting posterior samples. Although a full-scale and end-to-end search pipeline for GCBs based on machine learning techniques is not available up to now. Related algorithms are demonstrated to be effective and are being incorporated into global fit pipelines as reported in the roadmap mentioned in [59].

4. Unresolvable sources

As mentioned in Section 2, the individually resolvable GCBs are only a small fraction of the total sources. The remains will form a stochastic foreground and contribute to the confusion noise in the data of space-borne detectors. On the one hand, the unresolvable GCBs play the role of noise and affect the observations of other types of sources. On the other hand, the unresolvable sources can still provide invaluable information about the Galaxy.

4.1. Separation of the stochastic Galactic foreground

The Galactic confusion noise will degrade the sensitivity of detectors and reduce the SNR of other sources [60–63]. Additionally, due to the orbital motion of detectors, the pointing of constellations will constantly change relative to the Galactic center where GCBs are concentratedly distributed, which influences the response to the population of unresolvable GCBs and induces the modulation of the Galactic confusion noise. Therefore the Galactic confusion noise is nonstationary, adding more challenges in data analyses of space-borne detectors [165,166]. Furthermore, the foreground of unresolvable GCBs will blend together with the extra-galactic stochastic background of astrophysical or cosmological origin. It is important to separate or subtract the Galactic foreground for studies of other stochastic GW background signals [64–69].

The Galactic stochastic foreground can be separated from other components of stochastic signals or noise thanks to its distinct spectral shape and the modulation induced by the orbital motion of detectors. For the separation, one needs to first properly model the different components. The stochastic signals can be characterized by the cross spectrum $\langle \tilde{s}_i(f) \tilde{s}_j^*(f) \rangle$ where i and j denotes different TDI channels. Here the angle brackets denote the ensemble average. The total stochastic signals may be contributed by three components including the confusion foreground originating from unresolvable GCBs, the stochastic background of extra-galactic origin, and the instrument noise as

$$\langle \tilde{s}_i \tilde{s}_j^* \rangle = \langle \tilde{h}_i \tilde{h}_j^* \rangle_{\text{gal}} + \langle \tilde{h}_i \tilde{h}_j^* \rangle_{\text{extra-gal}} + \langle \tilde{n}_i \tilde{n}_j^* \rangle_{\text{ins}}. \quad (36)$$

There are various models developed to describe these components. For example, a typical model for extra-galactic stochastic background reads

$$\langle \tilde{h}_i \tilde{h}_j^* \rangle_{\text{extra-gal}} = \frac{3H_0^2 A_*}{4\pi^2 f^3} \left(\frac{f}{1\text{mHz}} \right)^m R_{ij}, \quad (37)$$

where R_{ij} is the response functions of detectors for different channels averaged over sky locations and polarizations, H_0 is the Hubble constant, m is the spectral index, and A_* is the amplitude at the reference frequency 1 mHz. This power-law model is widely used in both stochastic signals of cosmological origin [167] and astrophysical origin [168]. More models for extra-galactic stochastic background can be found in [64].

Modeling instrument noise can be extremely complicated involving detailed studies about the electronic systems, optical systems, and space environments of detectors, etc. However, in discussions of data analyses or science problems, a simplified noise model can be used, which groups noise into two components, Interferometry Metrology System (IMS) noise and acceleration noise, characterized by two quantities S_{IMS} and S_{acc} respectively [87,169]. Detailed derivation for noise spectral densities and correlations of different channels for different TDI generations or levels of approximation can be found in [64,87].

For modeling the stochastic Galactic foreground, one simplified method is utilizing the analytic fitting [170–172] which reads

$$\langle \tilde{h}_i \tilde{h}_j^* \rangle_{\text{gal}} = B_* f^{-7/3} \exp \left[-f^\alpha - \beta f \sin(\kappa f) \right] \left\{ 1 + \tanh[\gamma(f_k - f)] \right\} R_{ij}, \quad (38)$$

where fitting parameter B_* controls the overall amplitude, f_k controls the position of the knee-like feature in the spectrum of Galactic foreground, together with α , β , γ , and κ describe the spectral shape of Galac-

tic confusion foreground. In practice, one can only vary the overall amplitude in parameter estimation while taking the values of other parameters fitted through simulations of GCB population [64].

However, in this simplified model, the features of anisotropy and non-stationary are averaged and not incorporated. To fully exploit the distinctive features of the Galactic confusion foreground, a numerical model [66] can be used. In this numerical model, the modulated spectrum of the Galactic foreground is characterized by 17 Fourier coefficients [66,173,174] which are treated as free parameters varied in parameter estimation, and their prior ranges are obtained through multiple runs of GCB population simulation. The simulations of GCB population can be based on the Galaxy model constrained by the bright and individually resolvable GCBs [175]. To properly incorporate the modulation induced by detector orbital motion, the entire data are divided into segments of week-long during which the responses of detectors to Galactic confusion foreground are considered to have no appreciable changes.

After properly modeling all components of stochastic signals, one can use Bayesian inference to obtain estimations of free parameters in models. The likelihood is given by

$$p(s|\theta) = \prod_{n,d} \frac{1}{(2\pi)^{N/2} |C_{d,ij}|} \exp\left(-s_{d,i} C_{d,ij}^{-1} s_{d,j}\right), \quad (39)$$

which is the probability of occurrence of observed data if the stochastic signals and noise are governed by the chosen models with parameters θ . Here, s denotes the entire observed data, θ denotes parameters required to describe models for all components of stochastic signals, n labels the data samples, d labels different segments used for addressing the modulation of Galactic confusion foreground, i and j represent the different TDI channels. $C_{d,ij}$ is the correlation matrix depending on chosen models for each components in stochastic signals. The product needs to run all data samples and all segments. Posterior distributions for model parameters can be obtained through MCMC sampling, from which one can get the separation of different components of stochastic signals and noise.

Except for the stochastic background signals of extra-galactic origin, subtracting Galactic confusion foreground also benefits observations of individual sources as presented in [176] where the technique referred to as dictionary learning is used to reconstruct MBHB signals with low-SNR. The basic idea of dictionary learning [177] is representing a target signal $h(t)$ through a set of basis elements called dictionary \mathbf{D} and a sparse coefficient vector α as $h \sim \mathbf{D}\alpha$. The dictionary can be predefined by training dataset created through signals of individual MBHBs without noise, and the coefficient vector α for a given MBHB signal is searched in the presence of noise. Their combination provides a reconstruction of the signal hidden in the noisy data. As demonstrated in the reference [176], low-SNR massive black binaries can be successfully separated in the presence of Galactic noise through this method.

4.2. Tools for studying the Galaxy

The old stellar population is one of the excellent tools to trace the dynamical evolution of the Galaxy [178]. However, these sources are usually dim and difficult to be observed in the electromagnetic band. GCBs make up the majority of the total old stellar population and their GW signals are not affected by crowded matters in the Galaxy. GW Observations of GCBs offer a unique tool to study the Galaxy. Although, it has been demonstrated that the individual resolvable GCBs which have better measurement and localization can already trace the structure of the Galaxy [26,27,179]. These resolvable sources are usually more massive or nearby, which may require careful notice of potential bias [162]. The stochastic foreground containing the contributions of the full GCB population can also reveal various properties of the Galaxy [70–72].

In one aspect, one can extract information about the Galaxy from the spherical harmonic decomposition of the Galactic confusion foreground [70]. The angular power spectrum of the foreground that encodes information of the GCB spatial distribution can be obtained from the observed

cross spectrum of different TDI channels by the framework presented in [180]. It is pointed out that in [70] the scale height of GCB distribution can be effectively constrained using the hexadecapole moment of the spherical harmonic expansion. Although the constraint provided by this methodology has limited sensitivity comparing to the method with resolvable GCBs, this approach may be less affected by the observation bias and provides a complementary way to measure the structure of the Galaxy. Another approach for obtaining information about the Galaxy is through the spectral shape and amplitude of the Galactic confusion foreground [72]. This methodology is similar to fitting Galactic confusion foreground with the analytic model of Eq. 38 mentioned in the last section. Whereas the fitting parameters can be mapped to the physical parameters describing properties of the Galaxy. Thus, one can obtain information of the Galaxy like the total stellar mass as discussed in [72] through the similar parameter estimation procedure discussed in the last section from the Galactic confusion foreground.

5. Summary

Space-borne detectors will open the windows of the low GW band in the near future, which can provide new tools to explore the Universe. In contrast to ground-based detectors whose data is noise-dominant, data from space-borne detectors will be signal-dominant where signals are more crowded and various sources are tangled together. The abundance of sources can present plentiful invaluable information on the one hand, but also play the role of noise to disturb source detections and measurements on the other hand. The heavily overlapping signals pose new challenges for the data analyses of space-borne detectors.

Among various source types targeted by space-borne detectors, the vast population of GCBs is likely the type having the most number to be detected. It is expected that there are tens of millions GCBs in the mHz band, while tens of thousands massive or nearby sources among them are individually resolvable and the remains will form a stochastic foreground contributing to the confusion noise. The GWs from GCBs are continuous signals and have the feature of quasi-monochrome. In the time domain, GCB signals with overwhelming numbers exist in the data of space-borne detectors simultaneously during the entire mission period. In the frequency domain, although a single GCB signal is only extended within a narrow band, due to their vast number, GCB signals can still heavily overlap in frequency. GCBs are likely the type that has the most heavy overlap and correlation. Separation and extraction of overlapping signals focusing on GCBs may be the foundation of the ultimate data analysis pipelines for globally separating all kinds of sources. Therefore, in the paper, we present a comprehensive review of precious efforts dedicated to the separation and extraction of GCB signals.

Current solutions for separating overlapping GCB signals can be mainly categorized into two classes, iterative subtraction and global fitting. The strategy of iterative subtraction searches for the optimal fitting for just one signal in each iteration and the same procedure will be performed repeatedly with the remaining data after the subtraction of identified signal. A typical iterative subtraction solution referred to as GBSIEVER employs the F -statistic likelihood and PSO to search the optimal fitting in each step. To address the problem of inaccurate subtraction contamination for low SNR sources, a different implementation of iterative subtraction uses a tuned particle movement rule of PSO aiming at identifying all local maxima in the likelihood surface. Then, the local maxima corresponding to false candidates induced by degeneracy of individual signals and overlapping signals are removed. Astrophysical properties of GCBs are incorporated to further filter out the low credible candidates. Another implementation of iterative subtraction focuses on improving the efficiency in the PSO searching. Before the fine PSO search procedure, a template search is performed to obtain the coarse estimation of existing signals.

Another strategy is global fitting which estimates the parameters of all sources together with the source number simultaneously through the full Bayesian approach. The full Bayesian approach can provide the joint

posterior distributions for multiple overlapping signals rather than the only point estimation given in the iterative subtraction scheme, which can well describe the correlation among overlapping signals and provide a straightforward way for uncertain analysis and model selection. Posterior estimation over the parameter space with an uncertain dimension is performed through RJMCMC. To overcome the poor convergence in the trans-dimensional MCMC sampling, various tuned proposal distributions according to the features of GCB signals and sophisticated random sampling mechanisms are adopted in different implementations.

Besides typical solutions of above two schemes, a hybrid approach that combines the point estimation of maximum likelihood and Bayesian posterior estimation is also proposed. A procedure of iterative subtraction is first performed to identify potential candidates, and a global optimization of all identified signals is performed again to relieve errors induced by correlations of overlapping signals. Then, the MCMC sampling is performed for each found signal to obtain the marginalized posterior distributions of individual sources. Machine learning techniques are considered as very promising tools in future GW data analyses. Neural density estimation methods are investigated for assistance in drawing proposals and sharing the posterior results with the advantage of normalizing flows at efficiently modeling complicated distributions.

The resolvable GCBs are only a small fraction of the entire population. The remaining sources will form a stochastic foreground contributing to the confusion noise. The Galactic confusion foreground on the one hand is a type of noise affecting the observation of other sources, on the other hand provides a powerful tool to research the structure of the Galaxy with the advantage that GWs will not be suppressed or obscured by crowded matters in the Galaxy. Thanks to the features of the different spectral shape and the time-evolving modulation, the stochastic Galactic foreground is distinguishable with the instrument noise and the extra-galactic astrophysical or cosmological original background. And the distinct spectral shape of Galactic foreground can also be used in the measurement of properties of the Galaxy. The anisotropy of the Galactic foreground is associated with the stellar population distribution of the Galaxy, thus the Galactic structure can be constrained through the spherical harmonics decomposition of the Galactic foreground.

For future works, although there are already diverse end-to-end prototype pipelines for extracting overlapping GCBs, various problems still need to be addressed for actually handling real data. For example, in current solutions the noises are usually assumed to be stationary and Gaussian, while in reality there is slow drift of instrument noise in the long period and glitches in the short period which are neither stationary nor Gaussian [181,182]. Better noise modeling may required for future works. Besides, gaps will exist unavoidably in the data stream due to various reasons including scheduled maintenance or unplanned random events [183,184]. Future pipelines may need to incorporate processing of various potential defects existing in data. Furthermore, the data from space-borne detectors blend all sources of different types, the ultimate pipelines need to be capable of addressing different kinds of overlapping signals. New independent implementations or enchantments of current pipelines are anticipated for preparation of the launch of future space-borne GW observation missions.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

The authors have no conflict of interest. W. Z. is supported by the National Key R&D Program of China (2022YFC2204602 and 2021YFC2203102), Strategic Priority Research Program of the Chinese Academy of Science (XDB0550300), the National Natural Science Foundation of China (12325301 and 12273035), the Science Research Grants from the China Manned Space Project (CMS-CSST-2021-B01), the 111

Project for "Observational and Theoretical Research on Dark Matter and Dark Energy" (B23042) and Cyrus Chun Ying Tang Foundations. R. N. is supported in part by the National Key Research and Development Program of China (2022YFC2807303).

References

- [1] B.P. Abbott, R. Abbott, T.D. Abbott, 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 (3) (2019) 031040.
- [2] R. Abbott, T.D. Abbott, S. Abraham, et al., GWTC-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing, *Phys. Rev. X* 11 (2021) 021053.
- [3] R. Abbott, T.D. Abbott, F. Acernese, et al., GWTC-2.1: Deep extended catalog of compact binary coalescences observed by ligo and virgo during the first half of the third observing run, *Phys. Rev. D* 109 (2024) 022001.
- [4] R. Abbott, T.D. Abbott, F. Acernese, et al., GWTC-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run, *Phys. Rev. X* 13 (2023) 041039.
- [5] P. Amaro-Seoane, H. Audley, S. Babak, et al. Laser interferometer space antenna. arXiv:1702.00786.
- [6] W.-H. Ruan, Z.-K. Guo, R.-G. Cai, et al., Taiji program: Gravitational-wave sources, *Int. J. Mod. Phys. A* 35 (17) (2020) 2050075, doi:10.1142/s0217751x2050075x.
- [7] J. Luo, L.-S. Chen, H.-Z. Duan, et al., TianQin: A space-borne gravitational wave detector, *Class. Quant. Grav.* 33 (3) (2016) 035010, doi:10.1088/0264-9381/33/3/035010.
- [8] P. Repton, V. Raymond, Parameter estimation bias from overlapping binary black hole events in second generation interferometers, *Phys. Rev. D* 104 (8) (2021) 084039, doi:10.1103/physrevd.104.084039.
- [9] A. Samajdar, J. Janquart, C.V.D. Broeck, et al., Biases in parameter estimation from overlapping gravitational-wave signals in the third-generation detector era, *Phys. Rev. D* 104 (4) (2021) 044003, doi:10.1103/physrevd.104.044003.
- [10] E. Pizzati, S. Sachdev, A. Gupta, et al., Toward inference of overlapping gravitational-wave signals, *Phys. Rev. D* 105 (10) (2022) 104016, doi:10.1103/physrevd.105.104016.
- [11] J. Janquart, T. Baka, A. Samajdar, et al. Parameter estimation methods for analyzing overlapping gravitational wave signals in the third-generation detector era. arXiv:2211.01304.
- [12] J. Langendorff, A. Kolmus, J. Janquart, et al., Normalizing flows as an avenue to study overlapping gravitational wave signals, *Phys. Rev. Lett.* 130 (2023) 171402.
- [13] Y. Dang, Z. Wang, D. Liang, et al., Impact of overlapping signals on parameterized post-newtonian coefficients in tests of gravity, *Astrophys. J.* 964 (2) (2023) 194.
- [14] Y. Himemoto, A. Nishizawa, A. Taruya, Impacts of overlapping gravitational-wave signals on the parameter estimation: Toward the search for cosmological backgrounds, *Phys. Rev. D* 104 (4) (2021) 044010.
- [15] H. Zhong, R. Ormiston, V. Mandic, Detecting cosmological gravitational wave background after removal of compact binary coalescences in future gravitational wave detectors, *Phys. Rev. D* 107 (6) (2023) 064048.
- [16] F. Badaracco, B. Banerjee, M. Branchesi, et al., Blind source separation in 3rd generation gravitational-wave detectors, *New Astron. Rev.* 99 (2024) 101707.
- [17] P. Amaro-Seoane, J. Andrews, M.A. Sedda, et al., Astrophysics with the laser interferometer space antenna, *Living Rev. Relativ.* 26 (2023) 2.
- [18] K. Belczynski, M. Benacquista, T. Bulik, Double compact objects as low-frequency gravitational wave sources, *Astrophys. J.* 725 (1) (2010) 816–823.
- [19] A.J. Ruitter, K. Belczynski, M. Benacquista, et al., The lisa gravitational wave foreground: A study of double white dwarfs, *Astrophys. J.* 717 (2) (2010) 1006–1021.
- [20] V. Korol, E.M. Rossi, P.J. Groot, et al., Prospects for detection of detached double white dwarf binaries with gaia, lsst and lisa, *Mon. Not. R. Astron. Soc.* 470 (2) (2017) 1894–1910.
- [21] G. Nelemans, L.R. Yungelson, S.F. Portegies Zwart, The gravitational wave signal from the galactic disk population of binaries containing two compact objects, *Astron. Astrophys.* 375 (3) (2001) 890–898.
- [22] G. Nelemans, The galactic gravitational wave foreground, *Class. Quant. Grav.* 26 (9) (2009) 094030.
- [23] T.R. Marsh, Double white dwarfs and lisa, *Class. Quant. Grav.* 28 (9) (2011) 094019.
- [24] P. Amaro-Seoane, S. Aoudia, S. Babak, et al. eLISA: Astrophysics and cosmology in the millihertz regime, arXiv:1201.3621.
- [25] P. Amaro-Seoane, S. Aoudia, S. Babak, et al., Low-frequency gravitational-wave science with elisa/ngo, *Class. Quant. Grav.* 29 (12) (2012) 124016.
- [26] V. Korol, E.M. Rossi, E. Barausse, A multimessenger study of the milky way's stellar disc and bulge with lisa,gaia, and lsst, *Mon. Not. R. Astron. Soc.* 483 (4) (2018) 5518–5533.
- [27] M.J.C. Wilhelm, V. Korol, E.M. Rossi, et al., The milky way's bar structural properties from gravitational waves, *Mon. Not. R. Astron. Soc.* 500 (4) (2020) 4958–4971.
- [28] K.A. Arnaud, S. Babak, J.G. Baker, et al., An overview of the mock lisa data challenges, in: AIP Conference Proceedings, AIP, 2006, doi:10.1063/1.2405108.
- [29] K.A. Arnaud, S. Babak, J.G. Baker, et al., A how-to-for the mock lisa data challenges, AIP Conf. Proc. 873 (2006) 625–632, doi:10.1063/1.2405109.
- [30] K.A. Arnaud, G. Auger, S. Babak, et al., Report on the first round of the mock lisa data challenges, *Class. Quant. Grav.* 24 (19) (2007) S529–S539.
- [31] K.A. Arnaud, S. Babak, J.G. Baker, et al., An overview of the second round of the mock lisa data challenges, *Class. Quant. Grav.* 24 (19) (2007) S551–S564.
- [32] S. Babak, J.G. Baker, M.J. Benacquista, et al., Report on the second mock lisa data challenge, *Class. Quant. Grav.* 25 (11) (2008) 114037.

- [33] S. Babak, J.G. Baker, M.J. Benacquista, et al., The mock lisa data challenges: From challenge 3 to challenge 4, *Class. Quant. Grav.* 27 (8) (2010) 084009.
- [34] S. Babak, J.G. Baker, M.J. Benacquista, et al., The mock lisa data challenges: from challenge 1b to challenge 3, *Class. Quant. Grav.* 25 (18) (2008) 184026, doi:10.1088/0264-9381/25/18/184026.
- [35] Q. Baghi, The lisa data challenges. arXiv:2204.12142.
- [36] Z. Ren, T. Zhao, Z. Cao, et al. Taiji data challenge for exploring gravitational wave universe (2023). arXiv:2301.02967.
- [37] X.-H. Zhang, S.D. Mohanty, X.-B. Zou, et al., Resolving galactic binaries in LISA data using particle swarm optimization and cross-validation, *Phys. Rev. D* 104 (2) (2021) 024023.
- [38] X.-H. Zhang, S.-D. Zhao, S.D. Mohanty, et al., Resolving galactic binaries using a network of space-borne gravitational wave detectors, *Phys. Rev. D* 106 (2022) 102004.
- [39] P. Gao, X.-L. Fan, Z.-J. Cao, et al., Fast resolving galactic binaries in lisa data and its ability to study the milky way, *Phys. Rev. D* 107 (12) (2023) 123029.
- [40] P. Gao, X. Fan, Z. Cao, Simultaneously search for multi-target galactic binary gravitational waves in reduced parameter space with Impso-cv (2024). arXiv:2401.09300 doi:10.48550/ARXIV.2401.09300.
- [41] Y. Lu, E.-K. Li, Y.-M. Hu, et al. An implementation of galactic white dwarf binary data analysis for mlde-3, arXiv:2205.02384.
- [42] R. Umstätter, N. Christensen, M. Hendry, et al., Lisa source confusion: identification and characterization of signals, *Class. Quant. Grav.* 22 (18) (2005) S901–S911.
- [43] R. Umstätter, N. Christensen, M. Hendry, et al., Bayesian modeling of source confusion in lisa data, *Phys. Rev. D* 72 (2) (2005) 022001.
- [44] T.B. Littenberg, N.J. Cornish, K. Lackeos, et al., Global analysis of the gravitational wave signal from galactic binaries, *Phys. Rev. D* 101 (12) (2020) 123021.
- [45] T.B. Littenberg, N.J. Cornish, Prototype global analysis of lisa data with multiple source types, arXiv:2301.03673.
- [46] K. Lackeos, T.B. Littenberg, N.J. Cornish, et al., The lisa data challenge radler analysis and time-dependent ultra-compact binary catalogues, *Astron. Astrophys.* 678 (2023) A123.
- [47] T.B. Littenberg, A.K. Lali, Have any lisa verification binaries been found? (2024). arXiv:2404.03046 doi:10.48550/ARXIV.2404.03046.
- [48] N. Karnesis, M.L. Katz, N. Korsakova, et al. Eryn : A Multi-purpose sampler for bayesian inference, arXiv:2303.02164.
- [49] M.L. Katz, N. Karnesis, N. Korsakova, et al. An efficient gpu-accelerated multi-source global fit pipeline for lisa data analysis, arXiv:2405.04690.
- [50] S.H. Strub, L. Ferraioli, C. Schmelzbach, et al. Global analysis of lisa data with galactic binaries and massive black hole binaries, arXiv:2403.15318.
- [51] S.H. Strub, L. Ferraioli, C. Schmelzbach, et al. Bayesian parameter-estimation of galactic binaries in lisa data with gaussian process regression, arXiv:2204.04467.
- [52] S.H. Strub, L. Ferraioli, C. Schmelzbach, et al., Accelerating global parameter estimation of gravitational waves from galactic binaries using a genetic algorithm and gpus, *Phys. Rev. D* 108 (10) (2023) 103018.
- [53] M. Dax, S.R. Green, J. Gair, et al., Real-time gravitational-wave science with neural posterior estimation, *Phys. Rev. Lett.* 127 (24) (2021) 241103.
- [54] J. Wildberger, M. Dax, S.R. Green, et al., Adapting to noise distribution shifts in flow-based gravitational-wave inference, *Phys. Rev. D* 107 (8) (2023) 084046.
- [55] M. Dax, S.R. Green, J. Gair, et al., Neural importance sampling for rapid and reliable gravitational-wave inference, *Phys. Rev. Lett.* 130 (17) (2023) 171403.
- [56] M. Dax, S.R. Green, J. Gair, et al., Group equivariant neural posterior estimation, arXiv:2111.13139.
- [57] S.R. Green, J. Gair, Complete parameter inference for gw150914 using deep learning, arXiv:2008.03312.
- [58] S.R. Green, C. Simpson, J. Gair, Gravitational-wave parameter estimation with autoregressive neural network flows, *Phys. Rev. D* 102 (10) (2020) 104057.
- [59] N. Korsakova, S. Babak, M.L. Katz, et al. Neural density estimation for galactic binaries in lisa data analysis (2024). arXiv:2402.13701 doi:10.48550/ARXIV.2402.13701.
- [60] N. Seto, Annual modulation of the galactic binary confusion noise background and lisa data analysis, *Phys. Rev. D* 69 (12) (2004) 123005.
- [61] T. Robson, N. Cornish, Impact of galactic foreground characterization on a global analysis for the LISA gravitational wave observatory, *Class. Quant. Grav.* 34 (24) (2017) 244002.
- [62] J. Wu, J. Li, Subtraction of the confusion foreground and parameter uncertainty of resolvable galactic binaries on the networks of space-based gravitational-wave detectors, *Phys. Rev. D* 108 (12) (2023) 124047.
- [63] C. Liu, W.-H. Ruan, Z.-K. Guo, Confusion noise from galactic binaries for Taiji, *Phys. Rev. D* 107 (2023) 064021.
- [64] R. Flauger, N. Karnesis, G. Nardini, et al., Improved reconstruction of a stochastic gravitational wave background with LISA, *J. Cosmol. Astropart. Phys.* 01 (2021) 059–059–059.
- [65] M.R. Adams, N.J. Cornish, Discriminating between a stochastic gravitational wave background and instrument noise, *Phys. Rev. D* 82 (2) (2010) 022002.
- [66] M.R. Adams, N.J. Cornish, Detecting a stochastic gravitational wave background in the presence of a galactic foreground and instrument noise, *Phys. Rev. D* 89 (2) (2014) 022001.
- [67] S. Banagiri, A. Criswell, T. Kuan, et al., Mapping the gravitational-wave sky with lisa: A bayesian spherical harmonic approach, *Mon. Not. R. Astron. Soc.* 507 (4) (2021) 5451–5462.
- [68] G. Boileau, A. Lamberts, N. Christensen, et al., Spectral separation of the stochastic gravitational-wave background for lisa in the context of a modulated galactic foreground, *Mon. Not. R. Astron. Soc.* 508 (1) (2021) 803–826.
- [69] D. Poletti, Measuring the primordial gravitational wave background in the presence of other stochastic signals, *J. Cosmol. Astropart. Phys.* 2021 (05) (2021) 052.
- [70] K. Breivik, C.M.F. Mingarelli, S.L. Larson, Constraining galactic structure with the lisa white dwarf foreground, *Astrophys. J.* 901 (1) (2020) 4.
- [71] M. Benacquista, K. Holley-Bockelmann, Consequences of disk scale height on lisa confusion noise from close white dwarf binaries, *Astrophys. J.* 645 (1) (2006) 589–596.
- [72] M. Georgousi, N. Karnesis, V. Korol, et al., Gravitational waves from double white dwarfs as probes of the milky way, *Mon. Not. R. Astron. Soc.* 519 (2) (2022) 2552–2566.
- [73] T. Kupfer, V. Korol, T.B. Littenberg, et al., Lisa galactic binaries with astrometry from gaia dr3, *Astrophys. J.* 963 (2) (2024) 100.
- [74] T. Kupfer, V. Korol, S. Shah, et al., Lisa verification binaries with updated distances from gaia data release 2, *Mon. Not. R. Astron. Soc.* 480 (1) (2018) 302–309.
- [75] S. Nissanke, M. Vallisneri, G. Nelemans, et al., Gravitational-wave emission from compact galactic binaries, *Astrophys. J.* 758 (2) (2012) 131.
- [76] M. Benacquista, LISA and the Galactic Population of Compact Binaries, Springer Nature Singapore, 2022, pp. 885–908.
- [77] M.L. Katz, C. Danielski, N. Karnesis, et al., Bayesian characterisation of circumbinary exoplanets with lisa, *Mon. Not. R. Astron. Soc.* 517 (1) (2022) 697–711.
- [78] N.J. Cornish, T.B. Littenberg, Tests of bayesian model selection techniques for gravitational wave astronomy, *Phys. Rev. D* 76 (8) (2007) 083006.
- [79] Z. Wang, Z. Cao, X.-F. Zhang, Measuring mass transfer of am cvn binaries with a space-based gravitational wave detector, *Mon. Not. R. Astron. Soc.* 525 (1) (2023) 270–278.
- [80] N.J. Cornish, L.J. Rubbo, LISA response function, *Phys. Rev. D* 67 (2) (2003) 022001.
- [81] L.J. Rubbo, N.J. Cornish, O. Poujade, Forward modeling of space-borne gravitational wave detectors, *Phys. Rev. D* 69 (8) (2004) 082003.
- [82] A. Królak, M. Tinto, M. Vallisneri, Optimal filtering of the LISA data, *Phys. Rev. D* 70 (2) (2004) 022003.
- [83] S. Marsat, J.G. Baker, Fourier-domain modulations and delays of gravitational-wave signals (2018). 1806.10734.
- [84] M.L. Katz, mikekatz04/gbgpu: First official public release!, 2022, doi:10.5281/ZENODO.6500434.
- [85] M. Tinto, S.V. Dhurandhar, Time-delay interferometry, *Living Rev. Relativ.* 24 (1) (2020), doi:10.1007/s41114-020-00029-6.
- [86] M. Vallisneri, Synthetic LISA: Simulating time delay interferometry in a model LISA, *Phys. Rev. D* 71 (2) (2005) 022001.
- [87] S. Babak, M. Hewitson, A. Petiteau, Lisa sensitivity and snr calculations. arXiv:2108.01167.
- [88] X.-L. Zhao, P.-P. Wang, C.-G. Shao, The evaluation for plasma noise in arbitrary time-delay interferometry combinations, *Class. Quant. Grav.* 41 (10) (2024) 105002, doi:10.1088/1361-6382/ad387c.
- [89] P.-P. Wang, W.-L. Qian, Z.-Q. Wu, et al., Sensitivity functions for geometric time-delay interferometry combinations, *Phys. Rev. D* 108 (4) (2023) 044075.
- [90] P.-P. Wang, W.-L. Qian, Y.-J. Tan, et al., Geometric approach for the modified second generation time delay interferometry, *Phys. Rev. D* 106 (2) (2022) 024003.
- [91] N.J. Cornish, K. Shuman, Black hole hunting with LISA, *Phys. Rev. D* 101 (12) (2020) 124008.
- [92] S. Marsat, J.G. Baker, T.D. Canton, Exploring the bayesian parameter estimation of binary black holes with LISA, *Phys. Rev. D* 103 (8) (2021) 083011.
- [93] M. Katz, CChapmanbird, L. Speri, et al. mikekatz04/lisaanalysisools: First main release., 2024, doi:10.5281/zenodo.10930980.
- [94] M.L. Katz, J.-B. Bayle, A.J. Chua, et al., Assessing the data-analysis impact of LISA orbit approximations using a GPU-accelerated response model, *Phys. Rev. D* 106 (10) (2022) 103001.
- [95] M.L. Katz, mikekatz04/gbgpu: First official public release!, 2022, doi:10.5281/zenodo.6500434.
- [96] L.S.S. Team, LISA Science Requirements Document, Technical Report, 2018. <https://www.cosmos.esa.int/documents/678316/1700384/SciRD.pdf/25831fb3c01-e215-5916-4ac6e4b306fb?t=1526479841000>
- [97] N.J. Cornish, S.L. Larson, Lisa data analysis: Source identification and subtraction, *Phys. Rev. D* 67 (10) (2003) 103001.
- [98] N.J. Cornish, J. Crowder, LISA data analysis using markov chain monte carlo methods, *Phys. Rev. D* 72 (4) (2005) 043005.
- [99] J. Crowder, N. Cornish, A solution to the galactic foreground problem for lisa, *Phys. Rev. D* 75 (2007) 043008.
- [100] J. Crowder, N.J. Cornish, Extracting galactic binary signals from the first round of mock lisa data challenges, *Class. Quant. Grav.* 24 (19) (2007) S575–S585.
- [101] T.B. Littenberg, Detection pipeline for galactic binaries in lisa data, *Phys. Rev. D* 84 (6) (2011) 063009.
- [102] L.J. Rubbo, N.J. Cornish, R.W. Hellings, Slice and dice: Identifying and removing bright galactic binaries from lisa data, in: AIP Conference Proceedings, AIP, 2006, doi:10.1063/1.2405089.
- [103] J.T. Whelan, R. Prix, D. Khurana, Searching for galactic white-dwarf binaries in mock lisa data using an f statistic template bank, *Class. Quant. Grav.* 27 (5) (2010) 055010.
- [104] J.T. Whelan, R. Prix, D. Khurana, Improved search for galactic white-dwarf binaries in mock lisa data challenge 1b using an f statistic template bank, *Class. Quant. Grav.* 25 (18) (2008) 184029.
- [105] R. Prix, J.T. Whelan, F-statistic search for white-dwarf binaries in the first mock lisa data challenge, *Class. Quant. Grav.* 24 (19) (2007) S565–S574.
- [106] A. Blaut, S. Babak, A. Królak, Mock lisa data challenge for the galactic white dwarf binaries, *Phys. Rev. D* 81 (6) (2010) 063008.
- [107] Y. Bouffanais, E.K. Porter, Detecting compact galactic binaries using a hybrid swarm-based algorithm, *Phys. Rev. D* 93 (6) (2016) 064020.

- [108] S.D. Mohanty, R.K. Nayak, Tomographic approach to resolving the distribution of lisa galactic binaries, *Phys. Rev. D* 73 (8) (2006) 083006.
- [109] K.R. Nayak, S.D. Mohanty, K. Hayama, The tomographic method for lisa binaries: Application to mlcd data, *Class. Quant. Grav.* 24 (19) (2007) S587–S594.
- [110] J. Crowder, N.J. Cornish, J.L. Reddinger, Lisa data analysis using genetic algorithms, *Phys. Rev. D* 73 (6) (2006) 063011.
- [111] A. Stroer, J. Veitch, C. Röver, E. Bloomer, et al., Inference on white dwarf binary systems using the first round mock lisa data challenges data sets, *Class. Quant. Grav.* 24 (19) (2007) S541–S549.
- [112] M. Trias, A. Vecchio, J. Veitch, Delayed rejection schemes for efficient markov-chain monte-carlo sampling of multimodal distributions, 2009, doi:10.48550/ARXIV.0904.2207.
- [113] M. Trias, A. Vecchio, J. Veitch, Markov chain monte carlo searches for galactic binaries in mock lisa data challenge 1b data sets, *Class. Quant. Grav.* 25 (18) (2008) 184028.
- [114] M. Vallisneri, A lisa data-analysis primer, *Class. Quant. Grav.* 26 (9) (2009) 094024.
- [115] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, in: ICNN-95, IEEE, 1995, doi:10.1109/icnn.1995.488968.
- [116] M.E. Normandin, S.D. Mohanty, T.S. Weerathunga, Particle swarm optimization based search for gravitational waves from compact binary coalescences: performance improvements, *Phys. Rev. D* 98 (4) (2018) 044029.
- [117] C. Messenger, R. Prix, M.A. Papa, Random template banks and relaxed lattice coverings, *Phys. Rev. D* 79 (10) (2009) 104017.
- [118] S. Maneewongvatana, D.M. Mount, Analysis of approximate nearest neighbor searching with clustered point sets (1999). arXiv:cs/9901013 doi:10.48550/ARXIV.CS/9901013.
- [119] J.L. Bentley, Multidimensional binary search trees used for associative searching, *Commun. ACM* 18 (9) (1975) 509–517.
- [120] W.K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* 57 (1) (1970) 97–109.
- [121] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, et al., Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (6) (1953) 1087–1092.
- [122] P.J. GREEN, Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* 82 (4) (1995) 711–732.
- [123] M. Sambridge, K. Gallagher, A. Jackson, et al., Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.* 167 (2) (2006) 528–542.
- [124] W.D. Voudsen, W.M. Farr, I. Mandel, Dynamic temperature selection for parallel tempering in markov chain monte carlo simulations, *Mon. Not. R. Astron. Soc.* 455 (2) (2015) 1919–1937.
- [125] K. Hukushima, K. Nemoto, Exchange monte carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* 65 (6) (1996) 1604–1608.
- [126] R.H. Swendsen, J.-S. Wang, Replica monte carlo simulation of spin-glasses, *Phys. Rev. Lett.* 57 (21) (1986) 2607–2609.
- [127] D. Foreman-Mackey, D.W. Hogg, D. Lang, et al., emcee: The mcmc hammer, *Publ. Astron. Soc. Pac.* 125 (925) (2013) 306–312.
- [128] P.J. Green, A. Mira, Delayed rejection in reversible jump metropolis-hastings, *Biometrika* 88 (4) (2001) 1035–1053.
- [129] M. Trias, A. Vecchio, J. Veitch, Delayed rejection schemes for efficient markov-chain monte-carlo sampling of multimodal distributions. arXiv:0904.2207.
- [130] L. Martino, A review of multiple try mcmc algorithms for signal processing, *Digit. Signal Process.* 75 (2018) 134–152.
- [131] J.S. Liu, F. Liang, W.H. Wong, The multiple-try method and local optimization in metropolis sampling, *J. Am. Stat. Assoc.* 95 (449) (2000) 121–134.
- [132] M. Bédard, R. Douc, E. Moulines, Scaling analysis of multiple-try mcmc methods, *Stochastic Process. Appl.* 122 (3) (2012) 758–786.
- [133] L. Martino, V.P. Del Olmo, J. Read, A multi-point metropolis scheme with generic weight functions, *Stat. Prob. Lett.* 82 (7) (2012) 1445–1453.
- [134] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.* 11 (4) (1997) 341–359.
- [135] J. Qiang, C. Mitchell, A unified differential evolution algorithm for global optimization, *IEEE Trans. Evol. Comput.* (2014). <https://www.osti.gov/biblio/1163659>
- [136] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, et al., Numerical Optimization, Springer Berlin Heidelberg, 2006, doi:10.1007/978-3-540-35447-5.
- [137] P. Virtanen, R. Gommers, T.E. Oliphant, et al., SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* 17 (3) (2020) 261–272.
- [138] the LVK Collaboration, A guide to LIGO–virgo detector noise and extraction of transient gravitational-wave signals, *Class. Quant. Grav.* 37 (5) (2020) 055002.
- [139] E. Thrane, C. Talbot, An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, *Publ. Astron. Soc. Aust.* 36 (2019).
- [140] C. Cutler, É.E. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D* 49 (6) (1994) 2658–2697.
- [141] L.S. Finn, Detection, measurement, and gravitational radiation, *Phys. Rev. D* 46 (12) (1992) 5236–5249.
- [142] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005, doi:10.7551/mitpress/3206.001.0001.
- [143] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [144] A. McLeod, D. Beveridge, L. Wen, et al. A binary neutron star merger search pipeline powered by deep learning (2024). arXiv:2409.06266 doi:10.48550/ARXIV.2409.06266.
- [145] Y. Sakai, Y. Itoh, P. Jung, et al., Unsupervised learning architecture for classifying the transient noise of interferometric gravitational-wave detectors, *Sci. Rep.* 12 (2022) 9935.
- [146] M.L. Chan, I.S. Heng, C. Messenger, Detection and classification of supernova gravitational wave signals: A deep learning approach, *Phys. Rev. D* 102 (4) (2020) 043022.
- [147] H. Wang, Z. Ren, Y. Zhou, et al. Intelligent noise suppression for gravitational wave observational data (2023). doi:10.21203/rs.3.rs-2452860/v1.
- [148] J. McGinn, C. Messenger, M.J. Williams, et al., Generalised gravitational wave burst generation with generative adversarial networks, *Class. Quant. Grav.* 38 (15) (2021) 155005.
- [149] C.-H. Liao, F.-L. Lin, Deep generative models of gravitational waveforms via conditional autoencoder, *Phys. Rev. D* 103 (12) (2021) 124051.
- [150] E. Cuoco, J. Powell, M. Cavaglia, et al., Enhancing gravitational-wave science with machine learning, *Mach. Learn. Sci. Technol.* 2 (1) (2020) 011002.
- [151] T. Zhao, R. Shi, Y. Zhou, et al. Dawning of a new era in gravitational wave data analysis: Unveiling cosmic mysteries via artificial intelligence – a systematic review, arXiv:2311.15585.
- [152] M. Du, B. Liang, H. Wang, et al. Advancing space-based gravitational wave astronomy: rapid detection and parameter estimation using normalizing flows (2023). arXiv:2308.05510 doi:10.48550/ARXIV.2308.05510.
- [153] T. Zhao, Y. Zhou, R. Shi, et al., Dilated convolutional neural network for detecting extreme-mass-ratio inspirals, *Phys. Rev. D* 109 (8) (2024) 084054.
- [154] Q. Yun, W.-B. Han, Y.-Y. Guo, et al. Detecting extreme-mass-ratio inspirals for space-borne detectors with deep learning, 2023, doi:10.48550/ARXIV.2309.06694.
- [155] X.-T. Zhang, C. Messenger, N. Korsakova, et al., Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks, *Phys. Rev. D* 105 (12) (2022) 123027.
- [156] W.-H. Ruan, H. Wang, C. Liu, et al., Rapid search for massive black hole binary coalescences using deep learning, *Phys. Lett. B* 841 (2023) 137904.
- [157] G. Papamakarios, E. Nalisnick, D.J. Rezende, et al., Normalizing flows for probabilistic modeling and inference, *J. Mach. Learn. Res.* 22 (57) (2021) 1–64.
- [158] I. Kobyzev, S.J.D. Prince, M.A. Brubaker, Normalizing flows: An introduction and review of current methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (11) (2021) 3964–3979.
- [159] C. Winkler, D. Worrall, E. Hoogeboom, et al. Learning likelihoods with conditional normalizing flows, arXiv:1912.00042.
- [160] G. Papamakarios, Neural density estimation and likelihood-free inference, arXiv:1910.13203.
- [161] S. Yu, C.S. Jeffery, The gravitational wave signal from diverse populations of double white dwarf binaries in the galaxy, *Astron. Astrophys.* 521 (2010) A85.
- [162] A. Lamberts, S. Blunt, T.B. Littenberg, et al., Predicting the lisa white dwarf binary population in the milky way with cosmological simulations, *Mon. Not. R. Astron. Soc.* 490 (4) (2019) 5888–5903.
- [163] V. Korol, S. Toonen, A. Klein, et al., Populations of double white dwarfs in milky way satellites and their detectability with lisa, *Astron. Astrophys.* 638 (2020) A153.
- [164] M. Falxa, S. Babak, M. Le Jeune, Adaptive kernel density estimation proposal in gravitational wave data analysis, *Phys. Rev. D* 107 (2) (2023) 022008.
- [165] M.C. Dignan, N.J. Cornish, Lisa gravitational wave sources in a time-varying galactic stochastic background 940 (2022) 10.
- [166] M.C. Edwards, P. Maturana-Russel, R. Meyer, et al., Identifying and addressing nonstationary LISA noise, *Phys. Rev. D* 102 (8) (2020) 084062.
- [167] N. Bartolo, C. Caprini, V. Domcke, et al., Science with the space-based interferometer LISA. IV: probing inflation with gravitational waves, *J. Cosmol. Astropart. Phys.* 12 (2016) 026.
- [168] B. Abbott, R. Abbott, T. Abbott, et al., Search for the isotropic stochastic background using data from advanced ligo's second observing run, *Phys. Rev. D* 100 (6) (2019) 061101.
- [169] P.A.V.M. Babak Stas, L.J. Maude, LISA Data Challenge: Sangria, Technical Report, 2020. <https://lisa-ldc.lal.in2p3.fr/static/data/pdf/LDC-manual-Sangria.pdf>
- [170] N. Cornish, T. Robson, Galactic binary science with the new lisa design, *J. Phys. Conf. Ser.* 840 (2017) 012024.
- [171] T. Robson, N.J. Cornish, C. Liu, The construction and use of lisa sensitivity curves, *Class. Quant. Grav.* 36 (10) (2019) 105011.
- [172] K. Schmitz, Lisa sensitivity to gravitational waves from sound waves, *Symmetry* 12 (9) (2020) 1477.
- [173] N.J. Cornish, Making maps with lisa, *Class. Quant. Grav.* 19 (7) (2002) 1279–1283.
- [174] N.J. Cornish, Mapping the gravitational-wave background, *Class. Quant. Grav.* 18 (20) (2001) 4277–4291.
- [175] M.R. Adams, N.J. Cornish, T.B. Littenberg, Astrophysical model selection in gravitational wave astronomy, *Phys. Rev. D* 86 (12) (2012) 124032.
- [176] C. Badger, K. Martinovic, A. Torres-Forné, et al., Dictionary learning: A novel approach to detecting binary black holes in the presence of galactic noise with lisa, *Phys. Rev. Lett.* 130 (9) (2023) 091401.
- [177] M. Sadeghi, M. Babaie-Zadeh, C. Jutten, Dictionary learning for sparse representation: A novel approach, *IEEE Signal Process. Lett.* 20 (12) (2013) 1195–1198.
- [178] V. Belokurov, Galactic archaeology: The dwarfs that survived and perished, *New Astron. Rev.* 57 (3–4) (2013) 100–121.
- [179] V. Korol, V. Belokurov, C.J. Moore, et al., Weighing milky way satellites with lisa, *Mon. Notices R. Astron. Soc. Lett.* 502 (1) (2021) L55–L60.
- [180] A. Taruya, H. Kudoh, Probing anisotropies of gravitational-wave backgrounds with a space-based interferometer. ii. perturbative reconstruction of a low-frequency skymap, *Phys. Rev. D* 72 (10) (2005) 104015.
- [181] A. Spadaro, R. Busicchio, D. Vetrugno, et al., Glitch systematics on the observation of massive black-hole binaries with lisa, *Phys. Rev. D* 108 (12) (2023) 123029.
- [182] Q. Baghi, N. Korsakova, J. Slutsky, et al., Detection and characterization of instrumental transients in lisa pathfinder and their projection to lisa, *Phys. Rev. D* 105 (4) (2022) 042002.

- [183] L. Wang, H.-Y. Chen, X. Lyu, et al. Window and inpainting: Dealing with data gaps for tianqin. arXiv:2405.14274.
- [184] Q. Baghi, J.I. Thorpe, J. Slutsky, et al., Gravitational-wave parameter estimation with gaps in LISA: A bayesian data augmentation method, *Phys. Rev. D* 100 (2) (2019) 022003.

Author profile

Prof. Wen Zhao is a professor in department of astronomy, University of Science and Technology of China. His major is gravitation-wave astronomy and cosmology, including the theory and detection of gravitational wave, the detection of electromagnetic counterparts of gravitational-wave events, primordial gravitational waves, the physics of the early Universe and so on.

Rui Niu is a postdoc in department of astronomy, University of Science and Technology of China. His major is the gravitational-wave astronomy, including the tests of gravitational theory with gravitational wave, the data analysis of ground-based and space-based gravitational-wave detection and so on.