Resolution problem of galactic binaries in LISA data

[1]Shao-Dong Zhao[a,b], Xue-Hao Zhang[a,b,c], Soumya D. Mohanty[c,d], Yu-Xiao Liu[a,b]

a Institute of Theoretical Physics & Research Center of Gravitation, Lanzhou University, Lanzhou 730000, China

b Lanzhou Center for Theoretical Physics, Lanzhou University, Lanzhou 730000, China

c Morningside Center of Mathematics, Academy of Mathematics and System Science, Chinese Academy of Sciences, 55, Zhong Guan Cun Donglu, Beijing, China, 100190

d Dept. of Physics and Astronomy, University of Texas Rio Grande Valley, One West University Blvd., Brownsville, Texas 78520

1 Introduction

Gravitational wave (GW) astronomy is now well-established in the frequency band $\sim [10, 1000]$Hz using ground-based interferometric detectors. Starting with the discovery in 2015 of GW150914 [1], a binary black hole (BBH) merger signal, by the twin LIGO detectors, the number of confirmed GW signals has grown to more than 90 over four observing runs of the LIGO and Virgo detectors. While the majority of signals are from BBH mergers, there is also a double-neutron star and two neutron star-black hole signals [2] in the haul so far.

The $\sim [10^{-4}, 1]$ Hz band is the target for the space-borne detectors like LISA (Laser Interferometer Space Antenna) [3], Taiji[8] and TianQin[7], all scheduled for launch in the 2030s. Different from ground-based detectors, target signals in space-borne ones are continuous GW signals, mainly emitted by inspiral GW sources. With different components in the inspiral systems, the GW sources can be categorized into extreme-mass ratio inspiral(EMRI), massive black hole binary(MBHB), binary black hole(BBH), and compact galactic binaries (GB). Besides, there are stochastic gravitational wave background(SGWB) originating from both cosmological phases and astrophysical sources.

The galactic compact binaries are the simplest signals in the LISA band, from now on we simply call frequency bands $\sim [10^{-4}, 1]$ Hz LISA band. That's because the components of these systems are stellar-mass compact objects like Neutron stars and white dwarfs, which is easy to make approximations for the waveform. But the systems are numerous, around 30 million systems are predicted by the current astrophysical model.

To provide a benchmark for the development and comparison of data analysis methods, the LISA community has organized several mock LISA data challenges (MLDCs). The most recent in this series are the LISA data challenges (LDCs) [4]. LDC1-4 provides single realizations of TDI combinations, containing Gaussian stationary instrumental noise added to the GW signals from 30×10^6 GBs. A catalog of the parameters of the GBs is also provided, allowing the performance of a multisource resolution method to be quantified rigorously.

As in the earlier paper [10], referred to as P1 from here on, we impose an algorithm for GB resolution problem named `GBSIEVER` (Galactic Binary Separation by Iterative Extraction and Val-

[1]120220909001@lzu.edu.cn

idation using Extended Range). It is convenient to define the following types of GB sources when discussing our results. (i) *True*: sources in the LDC1-4 catalog. (ii) *Reported*: the final list of estimated sources returned by `GBSIEVER`. (iii) *Identified*: The initial list of sources produced by the single source search step in `GBSIEVER` before various cuts are imposed to construct the set of reported sources. (iv) *Confirmed*: The reported sources that match true sources as determined by a prescribed metric for association. The fraction of confirmed sources in the set of reported ones is called the *detection rate*.

2 Data description

Time Delay Interferometry (TDI) [9] is a technique that will be employed in space-borne gravitational wave observatories to minimize the impact of noise in laser measurements. In the LISA case, involves combining data from different times and positions of spacecraft to effectively compress noise that is not related to the GW signal. Many combinations of TDI are proposed in order to cancel different noises, like Michelson, Sagnac, Relay, etc. We use the so-called A and E combinations in the Michelson combination, which have mutually independent instrumental noise, corresponding to the first generation of TDIs used in LDC1. (The T combination is dropped because the GW signal in it is highly attenuated.)

For a GW signal, a theoretical waveform is needed before we proceed to TDI combination of the signal. We start with the two polarizations in the transverse traceless (TT) gauge of a plane GW incident at the Solar System barycenter (SSB) origin. For a compact galactic binary, the polarizations are well-modeled in the SSB frame as linear chirps,

$$h_+(t) = \mathcal{A}\left(1+\cos^2\iota\right)\cos\Phi(t)\,, \tag{1}$$
$$h_\times(t) = -2\mathcal{A}\cos\iota\sin\Phi(t)\,, \tag{2}$$
$$\Phi(t) = \phi_0 + 2\pi f t + \pi\dot{f}t^2\,, \tag{3}$$

where $\mathcal{A}$ is the amplitude of the wave, ι is the inclination angle between the GB orbital angular momentum and the line of sight from the SSB origin to the GB, ϕ_0 is the initial phase, f is the frequency at the start of observations, and $\dot{f}$ is the secular frequency drift.

Then we need to project the GW waveform on the constellation arms, take the example of link between spacecraft 1 and spacecraft 2 $\vec{\mathbf{n}}_{12}$:

$$H_{12}(t) = h_+(t)\times\xi_+(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{n}}_{12}) + h_\times(t)\times\xi_\times(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{n}}_{12}), \tag{4}$$

where $\xi_\times$ and ξ_+ are the *antenna pattern functions* given by

$$\xi_+(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{n}}_{12}) = (\hat{\mathbf{u}}\cdot\hat{\mathbf{n}}_{12})^2 - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}}_{12})^2\,, \tag{5}$$
$$\xi_\times(\hat{\mathbf{u}},\hat{\mathbf{v}},\hat{\mathbf{n}}_{12}) = 2(\hat{\mathbf{u}}\cdot\hat{\mathbf{n}}_{12})(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}}_{12}). \tag{6}$$

After appropriate approximations, the relative frequency shift y_{12}^{slr} is

$$y_{12}(t_1) \approx \frac{1}{2\left(1-\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_{12}(t_1)\right)}\left[H_{12}\left(t_1-\frac{L_{12}(t_1)}{c}-\frac{\hat{\mathbf{k}}\cdot\mathbf{x}_2(t_1)}{c}\right)-H_{12}\left(t_1-\frac{\hat{\mathbf{k}}\cdot\mathbf{x}_1(t_1)}{c}\right)\right], \tag{7}$$

Then we can get to the first generation TDI X channel data s_1^X by multiplying time delay operators $\mathbf{D}_{ij}x(t) = x(t-L_{ij}(t))$ to the relative frequency shift like:

$$s_1^X = y_{13} + \mathbf{D}_{13}y_{31} + \mathbf{D}_{131}y_{12} + \mathbf{D}_{1312}y_{21} - \left[y_{12} + \mathbf{D}_{12}y_{21} + \mathbf{D}_{121}y_{13} + \mathbf{D}_{1213}y_{31}\right], \tag{8}$$

then Y and Z channel data can be obtained by permute the subscript for the spacecraft.

The TDI time series $\overline{y}^I_D$ for combination channel I is given by

$$\overline{y}^I \;\; = \;\; \sum_{k=1}^{N_s} \overline{s}^I(\theta_k) + \overline{n}^I \;, \tag{9}$$

where $\overline{x} \in \mathbb{R}^N$ denotes a row vector, $\overline{s}^I_D(\theta)$ denotes a single GB signal corresponding to source parameters θ, $\overline{n}^I_D$ is a realization of the instrumental noise, and N_s is the number of GBs, in LDC1-4 the number here is 29857650. For a GB, θ consists of $\{\mathcal{A}, \phi_0, \iota, \psi, \lambda, \beta, f, \dot{f}\}$, where $\mathcal{A}$, ϕ_0, ι, f, and $\dot{f}$ were defined following Eq. 1 to Eq. 3), ψ is the polarization angle defining the orientation of the binary orbit projected on the sky, and the longitude and the latitude of the source in the SSB frame are denoted by λ and β, respectively. In the rest of the paper, θ will serve as a stand in for a GB source itself where convenient. Following LDC1-4 we set the sampling frequency for the uniformly sampled time series $\overline{y}^I_D$, for all I and D, to be $f_s = 1/15$ Hz, and the number of samples to $N = 4194304$ corresponding to an observation period $T_{\rm obs} \approx 2$ yr.

3 GBSIEVER

`GBSIEVER` utilizes an iterative process where the parameters for a single source are determined at a time, and the identified signal is then removed from the data. During the parameter estimation phase, maximum likelihood estimation (MLE) is employed. Here, the log-likelihood function is formulated based on the premise that the data comprises a single source superimposed on Gaussian, stationary noise. The maximization of the log-likelihood in the case of a GB can be carried out analytically over a subset of the parameters, leaving behind a function, commonly called the $\mathcal{F}$-statistic in the GW literature, which is numerically maximized. In `GBSIEVER`, the latter maximization is carried out using particle swarm optimization (PSO)[6].

The principal algorithmic components of `GBSIEVER` and their corresponding settings are reviewed in this section. The description is self-contained but brief since further details are available in P1. We do not provide an extensive review of PSO here, which is used for the global optimization task in Eq. 15, since it is a widely known algorithm.

3.1 Single sources estimation

In the estimation of compact GB signals, $\mathcal{F}$-statistic is a commonly used simplification, we can write GW signal as

$$\overline{s}^I(\theta) = \bar{a}\mathbf{X}^I(\kappa) \tag{10}$$

where $\bar{a} = (a_1, a_2, a_3, a_4) \in \mathbb{R}^4$ is the *extrinsic* parameters obtained by reparameterizing $\mathcal{A}$, ϕ_0, ψ, and ι. $\mathbf{X}^I(\kappa)$ is *template* waveforms that depend on the *intrinsic* parameter set $\kappa = \{\lambda, \beta, f, \dot{f}\}$.

Then the estimator, $\hat{\theta}_M$, of the parameters θ of a single sources in iteration $M \geq 1$ is given by,

$$\hat{\theta}_M = \underset{\theta}{\rm argmin} \sum_{I \in \mathcal{I}} \left(\|\overline{y}^I_M - \overline{s}^I(\theta)\|^I \right)^2 , \tag{11}$$

where $\overline{y}^I_M = \overline{y}^I_{M-1} - \overline{s}^I(\theta_{M-1})$ and $\overline{y}^I_1 = \overline{y}^I$. Here $\|\cdot\|$ denotes the norm induce by the noise weighted inner product,

$$\langle \overline{x}, \overline{z} \rangle^I = \frac{1}{N f_s} (\widetilde{x}./\overline{S}^I)\widetilde{z}^\dagger, \tag{12}$$

where f_s the sampling frequency, $\tilde{x}^T = \mathbf{F}\bar{x}^T$ is the discrete Fourier transform of time domain data $\overline{x}$.

Then we define two quantities to get $\mathcal{F}$-statistic,

$$\mathbf{W}(i,j) = \sum_{I\in\mathcal{I}} \mathbf{W}^I(i,j) = \sum_{I\in\mathcal{I}} \langle \overline{X}^T(i), \overline{X}(j)\rangle^I, \mathbf{U}(i) = \sum_{I\in\mathcal{I}} \mathbf{U}^I(i) = \sum_{I\in\mathcal{I}} \langle \overline{y}^T, \overline{X}(i)\rangle^I, \tag{13}$$

then the estimator of $\bar{a}$ in Eq.10 analytically becomes

$$\hat{a}^T = \mathbf{W}^{-1}\mathbf{U} \tag{14}$$

and the MLE estimator of κ is given using $\mathcal{F}$-statistic as

$$\hat{\kappa} = \underset{\kappa}{\operatorname{argmax}}\ \mathcal{F}(\kappa) = \underset{\kappa}{\operatorname{argmax}}\ \mathbf{U}^T\mathbf{W}^{-1}\mathbf{U} \tag{15}$$

To represent the signal strength with respect to the instrumental noise, we use the signal-to-noise ratio defined below.

$$\mathrm{SNR}^2 = \sum_{I\in\mathcal{I}} \left(\|\overline{s}^I(\theta)\|^I\right)^2, \tag{16}$$

In the postprocessing of output of the multisource resolution, a metric is required to quantify the degree of association between a given pair of sources. We choose to use the correlation coefficient, $R(\theta,\theta')$, between the TDI signals corresponding to a given pair of sources as

$$R(\theta,\theta') = \frac{\sum_{I\in\mathcal{I}}\langle \overline{s}^I(\theta), \overline{s}^I(\theta')\rangle^I}{[\langle \overline{s}^I(\theta), \overline{s}^I(\theta)\rangle^I \langle \overline{s}^I(\theta'), \overline{s}^I(\theta')\rangle^I]^{1/2}} \tag{17}$$

3.2 Narrow frequency band search and edge effects

Single source searches in `GBSIEVER` are conducted simultaneously across narrow frequency bands, each with a width of 0.02 mHz. However, only sources detected within the central 0.01 mHz, we called the acceptance zone, are accepted into the set of identified sources. This limitation is implemented to prevent the high incidence of false candidates near the band edges, which can occur when the power of strong sources spreads across the search bands. The boundaries for each frequency band are defined by applying a Tukey window to the DFT of a TDI time series. The central flat portion of this window is slightly broader, at 0.015 mHz, than the acceptance zone. Sources that are excluded in one search band are not lost; they are captured in adjacent bands because the bands overlap, ensuring that the acceptance zones are contiguous.

3.3 Undersampling

After the application of the Tukey window to a specific search band, an inverse DFT is used to revert the bandpassed TDI data back to the time domain. The next critical step involves *undersampling*[5] of the time series. This technique significantly reduces the number of samples while retaining all the necessary information. Undersampling makes use of the aliasing error, which occurs when sampling below the Nyquist rate, to transfer the informational content of bandpassed data to lower frequencies. Given that the $\mathcal{F}$-statistic in each search band is calculated assuming white noise, this approach allows for the inner product in Eq.12 to be efficiently computed in the time domain.

3.4 Cross-validation

This is another key feature of `GBSIEVER`, the entire single sources search is rerun on the same data with all settings hold the same but the search range used for the frequency derivative is extended 10 times. Then we get two sets of identified sources from these two runs. We call the narrow frequency derivative search the *primary search* and the wider one *secondary search*. Finally, for

each identified source $\hat{\theta}_{1,i}$ in the primary search we compute correlation with respect to the sources in adjacent frequency region in the secondary search. We categorize the identified source $\hat{\theta}_{1,i}$ into reported source only if $R_{\rm ee}(\hat{\theta}_{1,i})$ is higher than a preset threshold. This cross-validation is extremely effective in eliminating spurious identified sources. The preset $R_{\rm ee}$ threshold is related to frequency and SNR of identified source as following,

$$R_{\rm ee} = \begin{cases} 0.9, & \nu \in [0,3]\text{mHz,SNR} \leq 25 \\ 0.5, & \nu \in [0,3]\text{mHz,SNR} > 25 \end{cases} \quad R_{\rm ee} = \begin{cases} 0.9, & \nu \in [3,4]\text{mHz,SNR} \leq 20 \\ 0.5, & \nu \in [3,4]\text{mHz,SNR} > 20 \end{cases} \tag{18}$$

For $f > 4$mHz, cross-validation is not required, that's because the number of spurious sources decreases rapidly.

4 results

Until now, we have successfully implemented `GBSIEVER` in various scenarios, including the single LISA detector, the combined LISA-Taiji network detector, and configurations that incorporate astrophysical priors. In each of these applications, the algorithm has consistently performed at the forefront of the field, demonstrating state-of-the-art effectiveness and reliability.

4.1 Single LISA detector

In table1 we show the number of identified, reported, and confirmed sources in each block which is divided according to SNR and frequency of the sources. And also the detection rate in each block and overall information. It can be seen in the low-SNR low-frequency region(block 1 and block 3), due to not only the high overlapping between the signals but also the faintness of the signals, the detection rate is low. The other results like subtraction residual and parameter estimation will be shown when comparing the single LISA detector results with improved settings.

	$\nu = [0,3]$		$\nu = [3,4]$		$\nu = [4,15]$
	SNR $[0,25]$	SNR $[25,\infty]$	SNR $[0,20]$	SNR $[20,\infty]$	SNR $[10,\infty]$
$R_{\rm ee}$	0.9	0.5	0.9	0.5	-1
Identified	23231	2106	3696	1526	4279
Reported	2767	2073	1622	1510	4279
Confirmed	1760	1892	1303	1394	4039
Detection rate	63.61%	91.27%	80.33%	92.32%	94.39%

Table 1: Performance of `GBSIEVER` for the single-detector LDC1-4 data. The total number of reported sources is 12251, and the total number of confirmed sources is 10388, and the total detection rate is 84.79%.

4.2 LISA-Taiji network

We extended our original `GBSIEVER` from the single LISA detector to network detectors by extending the definition of $\mathcal{F}$-statistic, SNR, and also generated the TDI data of 30 million GBs for the Taiji detector[11]. By extending the single LISA detector to the LISA-Taiji network detector case, we get the total number of confirmed GBs increase from 10388 to 18151, a significant improvement of 74.73%.

	$\nu = [0, 3]$		$\nu = [3, 4]$		$\nu = [4, 15]$
	SNR $[0, 25]$	SNR $[25, \infty]$	SNR $[0, 20]$	SNR $[20, \infty]$	SNR $[10, \infty]$
$R_{\rm ee}$	0.9	0.5	0.9	0.5	-1
Identified	33144	3941	3461	2528	4687
Reported	8420	3920	2440	2526	4687
Confirmed	5561	3605	1876	2406	4536
Detection rate	66.05%	91.96%	83.73%	95.25%	96.78%

Table 2: Performance of `GBSIEVER` for the LISA-Taiji network using LDC1-4 and Taiji-mod data. The total number of reported sources is 21993 and the total number of confirmed sources is 18151, and the total detection rate is 82.53%.

4.3 astrophysical prior

In the LDCs GB catalog, there are two kinds of sources, ones with mass transfer between two components of the system called semi-detached, and another set of sources without mass transfer called detached. These two kinds of sources are both well-constrained in frequency derivative when we have clear information on frequency.

We here use a Tukey window $w(\dot{f})$ in frequency derivative with window position determined by frequency, then the $\mathcal{F}$-statistic can be modified as

$$\mathcal{F}_{prior} = \mathcal{F}(\kappa) * w(\dot{f}) \tag{19}$$

we can see the effect of this Tukey window on the single parameter fitness in Fig.1, where the big blue dot shows the location and fitness value of the original fitness function and the yellow star shows the true location and fitness value, then the red diamond is the location of max windowed fitness.

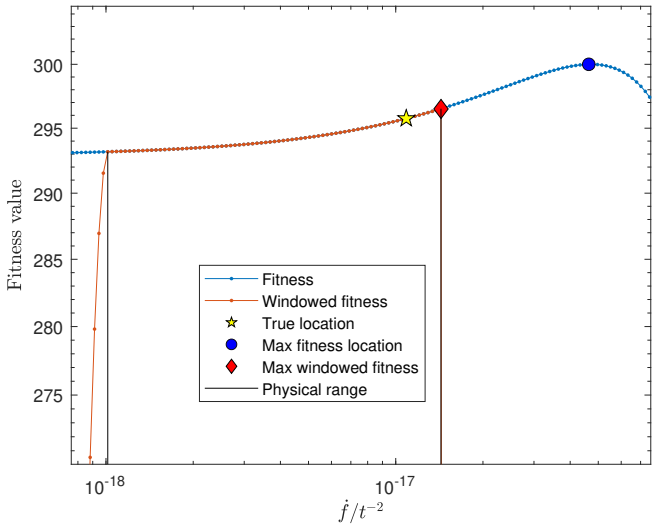

Figure 1: Illustration for the effect of frequency derivative prior on fitness function when fixing other parameters but the derivative. Here's the case with ($f \in [2.30, 2.32]$mHz). The data is a superposition of instrumental noise, all weaker signals, and the signal itself.

In the results of the test on the single LISA case with astrophysical prior implemented, the number of total confirmed increased from 10388 to 10596 with an improvement in detection rate from 84.79% to 87.46%, if we lose the $R_{\rm ee}$ threshold to get the same detection rate(84.79%),

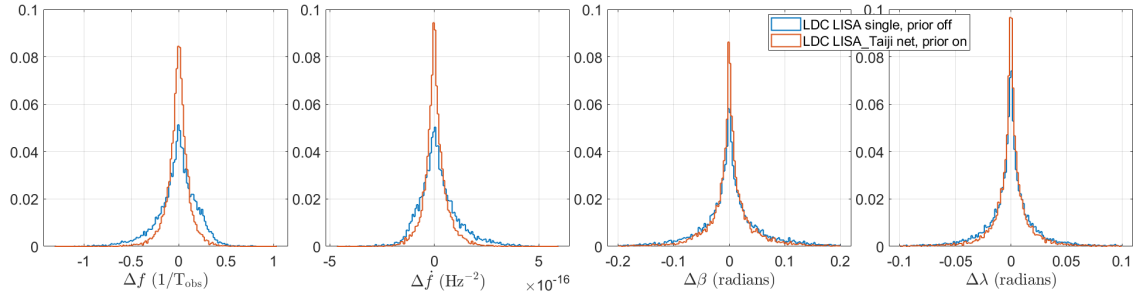

Figure 2: Intrinsic parameters estimation errors. The blue histogram is the estimation error of a single LISA detector using GBSIEVER and the red histogram is the one from a LISA-Taiji network detector using GBSIEVER with prior on frequency derivative

we can get 767 more confirmed sources. As for the LISA-Taiji network case with astrophysical prior implemented, the number of total confirmed increased from 18151 to 19142 with a great improvement in detection rate from 82.53% to 89.04%, if we lose the $R_{\rm ee}$ threshold to get the same detection rate(82.53%), we can get 2703 more confirmed sources. The improvement is much more significant in network detection than in single-detector experiments.

	$\nu = [0, 3]$		$\nu = [3, 4]$		$\nu = [4, 15]$
	SNR $\in [0, 25]$	SNR $\in [25, \infty]$	SNR $\in [0, 20]$	SNR $\in [20, \infty]$	SNR $\in [10, \infty]$
$R_{\rm ee}$	0.9	0.5	0.9	0.5	-1
Identified	25048	3919	3448	2527	4741
Reported	7914	3898	2423	2525	4738
Detection rate	79.04%	95.07%	87.66%	96.36%	97.59%

Table 3: Performance of `GBSIEVER` for the LISA-Taiji network using LDC1-4 and Taiji-mod data. The total number of reported sources is 21492, and the total number of confirmed sources is 19142, and the total detection rate is 89.04%.

Since our window fitness brings constraint on the frequency derivative, the improvement of parameter estimation performance is mainly in frequency derivative as expected, with a slight improvement in f as well. For common confirmed sources from GBSIEVER and GBSIEVER with prior on frequency derivative, we can compare the error of the estimated parameter. In Fig.2 we show the improvement in parameter estimation of 4 intrinsic parameters.

5 Conclusions

The algorithm `GBSIEVER` is one of the state-of-the-art data analysis pipelines for space-borne GW detectors aiming galactic binary. We successfully implement the pipeline with single LISA detector and LISA-Taiji network detector, also including the astrophysical prior for parameter estimation. The results of the analysis are all at the leading level in the context of multisource GB resolution problem.

Acknowledgments

This work was supported by National Key Research and Development Program of China (Grant No. 2021YFC2203003), the National Natural Science Foundation of China (Grants No. ***)

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102. URL `https://link.aps.org/doi/10.1103/PhysRevLett.116.061102`.

[2] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. *The Astrophysical Journal Letters*, 915(1):L5, jun 2021. doi: 10.3847/2041-8213/ac082e. URL `https://doi.org/10.3847/2041-8213/ac082e`.

[3] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binetruy, Michael Born, Daniele Bortoluzzi, et al. Laser Interferometer Space Antenna. *arXiv preprint arXiv:1702.00786*, 2017.

[4] Quentin Baghi. The LISA Data Challenges. *arXiv preprint arXiv:2204.12142*, 2022.

[5] David L Donoho and Jared Tanner. Precise undersampling theorems. *Proceedings of the IEEE*, 98(6):913–924, 2010.

[6] Russell Eberhart and James Kennedy. Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, volume 4, pages 1942–1948. Citeseer, 1995.

[7] Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, Jianwei Mei, Vadim Milyukov, Mikhail Sazhin, et al. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, 2016.

[8] Wen-Hong Ruan, Zong-Kuan Guo, Rong-Gen Cai, and Yuan-Zhong Zhang. Taiji program: Gravitational-wave sources. *International Journal of Modern Physics A*, 35(17):2050075, 2020.

[9] Massimo Tinto and Sanjeev V Dhurandhar. Time-delay interferometry. *Living Reviews in Relativity*, 24(1):1–73, 2021.

[10] Xue-Hao Zhang, Soumya D. Mohanty, Xiao-Bo Zou, and Yu-Xiao Liu. Resolving Galactic binaries in LISA data using particle swarm optimization and cross-validation. *Phys. Rev. D*, 104:024023, Jul 2021. doi: 10.1103/PhysRevD.104.024023. URL `https://link.aps.org/doi/10.1103/PhysRevD.104.024023`.

[11] Xue-Hao Zhang, Shao-Dong Zhao, Soumya D Mohanty, and Yu-Xiao Liu. Resolving galactic binaries using a network of space-borne gravitational wave detectors. *Physical Review D*, 106 (10):102004, 2022.