



# Taiji data challenge for exploring gravitational wave universe

Zhixiang Ren<sup>1,2,\*</sup>, Tianyu Zhao<sup>1,3,2,\*</sup>, Zhoujian Cao<sup>1,3,4,†</sup>, Zong-Kuan Guo<sup>1,5,6,4</sup>,  
Wen-Biao Han<sup>1,4,7,8,9,10</sup>, Hong-Bo Jin<sup>1,11,8</sup>, Yue-Liang Wu<sup>1,5,4,9,‡</sup>

1 Taiji Laboratory for Gravitational Wave Universe, University of Chinese Academy of Sciences (UCAS), Beijing 100049, China

2 Peng Cheng Laboratory, Shenzhen 518055, China

3 Department of Astronomy, Beijing Normal University, Beijing 100875, China

4 School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China

5 CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China

6 School of Physical Sciences, UCAS, Beijing 100049, China

7 Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

8 School of Astronomy and Space Science, UCAS, Beijing 100049, China

9 International Centre for Theoretical Physics Asia-Pacific (ICTP-AP, UNESCO), UCAS, Beijing 100190, China

10 Shanghai Frontiers Science Center for Gravitational Wave Detection, Shanghai 200240, China

11 Key Laboratory of Computational Astrophysics, National Astronomical Observatories, Beijing 100101, China

\*These two authors contributed equally to this work.

Corresponding authors. E-mail: †zjcao@amt.ac.cn, ‡ylwu@itp.ac.cn

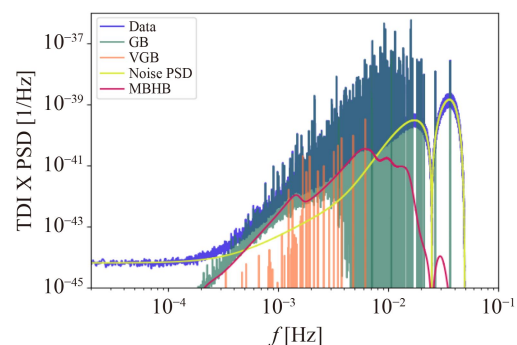
Received April 1, 2023; accepted June 6, 2023

© Higher Education Press 2023

## ABSTRACT

The direct observation of gravitational waves (GWs) opens a new window for exploring new physics from quanta to cosmos and provides a new tool for probing the evolution of universe. GWs detection in space covers a broad spectrum ranging over more than four orders of magnitude and enables us to study rich physical and astronomical phenomena. Taiji is a proposed space-based gravitational wave (GW) detection mission that will be launched in the 2030s. Taiji will be exposed to numerous overlapping and persistent GW signals buried in the foreground and background, posing various data analysis challenges. In order to empower potential scientific discoveries, the Mock Laser Interferometer Space Antenna (LISA) data challenge and the LISA data challenge (LDC) were developed. While LDC provides a baseline framework, the first LDC needs to be updated with more realistic simulations and adjusted detector responses for Taiji's constellation. In this paper, we review the scientific objectives and the roadmap for Taiji, as well as the technical difficulties in data analysis and the data generation strategy, and present the associated data challenges. In contrast to LDC, we utilize second-order Keplerian orbit and second-generation time delay interferometry techniques. Additionally, we employ a new model for the extreme-mass-ratio inspiral waveform and stochastic GW background spectrum, which enables us to test general relativity and measure the non-Gaussianity of curvature perturbations. Furthermore, we present a comprehensive showcase of parameter estimation using a toy dataset. This showcase not only demonstrates the scientific potential of the Taiji data challenge (TDC) but also serves to validate the effectiveness of the pipeline. As the first data challenge for Taiji, we aim to build an open ground for data analysis related to Taiji sources and sciences. More details can be found on the official website ([taiji-tdc.ictp-ap.org](http://taiji-tdc.ictp-ap.org)).

**Keywords** gravitational wave, universe evolution, Taiji, data challenge



## 1 Introduction

Ground-based gravitational wave (GW) detectors have made remarkable achievements since the first detection of GW150914 [1], nearly 100 GW events have been observed so far [2]. The discovery of GWs not only provides a more fundamental test on Einstein's theory of general relativity (GR) but also leads us to study deeply on the nature of gravity and spacetime. In particular, GWs detection in space enables us to probe the formation and structure of black holes from the early Universe through the peak of star formation era and to understand how does the intermediate mass seed black hole formed in the early universe and how the seed black hole grows into the large or extreme-large black hole [3]. It will also be helpful to explore whether the dark matter could form into the black hole or primordial black hole could be the candidate of dark matter.

Due to the limitation by seismic noise and arm length, ground-based detectors are not sensitive at frequencies lower than 10 Hz [4]. By moving the detector to space, we can eliminate the seismic and gravity-gradient noise and open up the frequency band of the GW within  $[10^{-4}, 0.1]$  Hz. The first planned space-based GW detection project is the laser interferometer space antenna (LISA) [5], proposed by the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA), which will be launched in around 2037<sup>1)</sup>. Furthermore, LISA has laid a solid foundation for space-borne GW detection, including scientific objectives, data analysis algorithms, and instrumental techniques.

During the early 2000s, Chinese scientists were also interested in space-borne GW detection programs and considered developing independent missions. The feasibility of space-based GW detection was studied by the Chinese Academy of Sciences (CAS) in 2008, followed by the initial mission proposal presented in 2011 [6]. A half-million-kilometer arm length and a noise budget that is a hundred times more sensitive than LISA at 0.1 Hz made this instrument particularly sensitive to intermediate mass black hole mergers [7]. In 2012, the first mission descope attempt was reported to prefer a constellation of three satellites separated each by about 3 million kilometers, coordinating both technical and scientific goals, and the proposed three-step road map was first presented in which the launch time for the Chinese space GW detector was expected to be in the 2030s [7, 8]. The CAS built prototypes that were used to conduct proof-of-principle experiments [9–12]. Similar to LISA, Taiji consists of three spacecraft (SC), and each SC follows a heliocentric orbit. The three SC form an approximately 3 million-kilometer-long equilateral triangle. There is a 20-degree trail between the center of mass of the constellation and the Earth. The constellation

is roughly 1AU distant from the Sun. This caused a slight eccentricity in the orbits of the three SC. With the stable constellation, Taiji's sensitive frequency band is  $[10^{-4}, 0.1]$  Hz [7]. A three-step road map of the Taiji program is: i) launch a satellite Taiji-1 for individual technique validation, ii) launch two satellites Taiji-2 for almost all techniques test, and iii) launch complete Taiji three satellites [13, 14]. The Taiji program, with its three-step road map, has received priority support from the CAS's strategic priority research program since 2016 for its pre-experimental study. Some key technologies have been developed further in the studies [13, 15–20]. Currently, the first phase of the Taiji-1 on-orbit experiment has been successfully completed [21, 22].

In this paper, we present the first Taiji data challenge (TDC). Specifically, we review the scientific objectives and technical challenges during the data processing procedure. Then, we visualize all TDC datasets and explain their scientific potential and technical obstacles. We implement time-domain time delay interferometry (TDI) 2.0 for detector response calculation, which makes our data more realistic compared to LDC1's original datasets. As the main component of our data challenge, we generate datasets that contain only one source and also datasets that contain multiple sources. We further generate an extreme-mass-ratio inspiral (EMRI) waveform constructed under the general parameterized metric, which allows us to test GR and investigate the non-Gaussianity of curvature perturbations by observing the GW background it induces.

The paper is organized as follows. In Section 2, we introduce the scientific objectives and technical challenge of Taiji mission. In Section 3, we demonstrate how to simulate Taiji instrumental noise, detector response, and GW waveform respectively. In Section 4, we show the properties of each dataset. In Section 5, we briefly introduce the application. The last section contains our discussions about future work and features of our TDC website.

## 2 Scientific objective and technical challenge of Taiji mission

Taiji detectors will receive a large amount of GW signals from different sources in the target frequency band of  $[10^{-4}, 0.1]$  Hz [13]. The observatory will measure signals from a wide range of different sources relevant to the astrophysical mechanism of the formation of black holes and galaxies, the test of GR, and the study of cosmology, including massive black holes binaries (MBHBs) at all redshifts; EMRIs; the inspiral of stellar-origin binary black holes (SOBBHs); galactic binaries (GB); verification galactic binaries (VGBs), and various stochastic GW background (SGWB) sources [7].

The mission's primary objective is to determine how

<sup>1)</sup> URL: [sci.esa.int/web/lisa](http://sci.esa.int/web/lisa)

**Table 1** Scientific objectives of Taiji mission.

No.	Scientific objective
1	Dynamical evolution and population of MBHBs, study the birth and growth of MBHs, and the astrophysical environment of the host galaxy
2	Precisely estimate parameters of MBH, reveals the physical nature of BHs, probe dynamics of galactic nuclei, and the astrophysical environment at the galaxy center by EMRI
3	Formation evolution and population of Galaxy binaries
4	Study SOBHB formation, environment, population, and joint observation with LIGO
5	Test GR, study the properties of GW propagation
6	GW cosmology, i.e., measurement of the Hubble constant and cosmology constant
7	PSD shape and upper limit of SGWB signals
8	Detect unmodeled signals
9	New physics and cosmology beyond GR

and when massive black holes have formed and grown over cosmic time. For the purpose of reconstructing their evolution, the study will explore almost all of the mass-redshift parameter space that is relevant [21]. We will investigate the distribution of MBHB's spins and masses in order to distinguish between different formation mechanisms based on the GW signal from MBHBs [23]. Additionally, the mission will analyze the signals from GBs and provide information on stellar evolution. Tens of millions of binaries in our galaxy produce stochastic foregrounds, which we can estimate by observing GBs [24]. It is possible to constrain the evolutionary pathways of compact binaries by examining the characteristics of the population, such as the number density of sources as a function of frequency. It is expected that Taiji will provide exceptionally strong tests of GR predictions by observing highly relativistic MBHBs and EMRIs [5]. During the merger phase of a binary black hole system, BHs travel at nearly the speed of light and interact strongly with each other, which allows the study of the full nonlinear dynamics of gravity [25]. By observing the signal of the EMRI system, the mass, spin, and quadrupole moment of the central massive black hole will be measured making testing its level of Kerr-ness and proving the no-hair hypothesis possible [26]. Finally, a space-based GW detector will explore new physics and cosmology and seek out previously unknown GW sources. The scientific objectives of the Taiji mission are listed in Table 1.

GWs signals are expected to be many orders of magnitude below the noise level caused by laser frequency fluctuations [5]. A post-processing technique called TDI has been proposed to reduce laser noise to an acceptable level [27]. The idea is to construct virtual equal-arm interference by combining the data streams in the appropriate way. The second-generation TDI is applicable to rotating and flexing detectors with linearly varying arm lengths. The actual Taiji orbits are affected by the gravitational field of the planets, which leads to non-periodic orbits. In science analysis, it is necessary to extract scientific information from the various TDI data streams; in a similar manner to the procedure for the

electromagnetic case, science analysis is carried out by deconvolution of the response function from the data stream in order to obtain the incident GW field observed by Taiji.

The main task of the data analysis is to quickly identify the signals and estimate their physical parameters precisely. Unlike current ground-based detectors, Taiji will be sensitive to continuous sources, primarily from GBs. It is expected that GBs will be the most numerous of the sources, the unresolved component of the galactic population produces an effectively stochastic "confusion noise". Due to the non-Gaussianity of GB foreground noise, detecting other signals buried in the foreground noise is a challenging task. What's more, analysis of the foreground itself is another challenge because of the enormous number of signals. Because of the long duration and complex morphology of their signals, detecting and characterizing the physics of SOBHBs with LISA is a highly nontrivial challenge. Data gaps will appear in the Taiji data stream due to maintenance, telescope realignment, and communication with Earth by a high-gain antenna. A series of significant challenges arise when dealing with gapped data, namely the information loss, spectral leakage, and noise correlations [28]. Another challenge for GW signal detection is the interference of instrumental glitches, which can be separated relatively natural from signals by ground observations with multiple interferometers. In contrast, this coherent searching approach cannot be directly applied in space-based detection, so novel glitch classification algorithms need to be developed [29]. All the technical challenges are summarized in Table 2.

### 3 Method

When analyzing real detector data, it is necessary to model the signal and noise; therefore, for simplicity in the TDC, several idealized models are utilized during data simulation. This section will introduce these models and their respective approximations. Using the existing waveform model, we first obtain the source frame GW

**Table 2** Technical challenges of Taiji mission.

No.	Technical challenges
1	Foreground GB signals separation
2	SOBBH signals separation
3	Others signal with non-Gaussian GB noise
4	Overlapping MBHB signals
5	Instrumental glitches
6	Data gap due to maintenance, telescope re-alignment, and communication issue
7	TDI technique to suppress laser frequency noise

waveform during signal simulation. Then, we determine the response of each link, which is dependent upon the orbit of the Taiji SCs. In addition, TDI is used to obtain the signal in the  $X$ ,  $Y$ , and  $Z$  channels. The instrumental noise in those 3 channels is Gaussian noise simulated by their PSD. In the final step, the signal is injected into the noise.

### 3.1 Orbital motion

We use Keplerian orbit to approximate the motion of the Taiji SC [30–32]. Given the distance between the SC, i.e., the arm length  $L$  and the distance from the SC to the sun  $a$ , the eccentricity of the orbit is defined by

$$e = \left(1 + \frac{4}{\sqrt{3}}\alpha \cos \phi + \frac{4}{3}\alpha^2\right)^{1/2} - 1, \quad (1)$$

where  $\phi = \pi/3 + 5\alpha/8$  is the angle between the SC's orbital planes and the ecliptic plane.  $\alpha = L/(2a)$  is a small parameter originally used in Ref. [30]. The specific value of  $\phi$  reduces the breathing effect of the arm length [31].

The reference position of the 3 SC is given by

$$\begin{cases} x_{\text{ref},k}(t) = a \cos \iota (\cos \psi_k(t) - e), \\ y_{\text{ref},k}(t) = a \sqrt{1 - e^2} \sin \psi_k(t), \\ z_{\text{ref},k}(t) = -a \sin \iota (\cos \psi_k(t) - e), \end{cases} \quad (2)$$

where  $\iota$  is the orbital inclination of SC1 relative to the ecliptic plane,

$$\tan \iota = \frac{2}{\sqrt{3}} \frac{\alpha \sin \phi}{1 + \frac{2}{\sqrt{3}}\alpha \cos \phi}. \quad (3)$$

Here  $\psi_k(t)$  is the eccentric anomaly of SC <sub>$k$</sub> , which is obtained by solving the Kepler equation to the desired order iteratively:

$$\begin{aligned} \psi_k(t) - e \sin \psi_k(t) &= m_k(t) = m_{0,k} + \Omega(t - t_0) \\ &= m_{0,1} - \frac{2\pi(k-1)}{3} + \Omega(t - t_0), \\ \Omega &= \frac{2\pi}{1 \text{ year}}. \end{aligned} \quad (4)$$

Then, the position of 3 SC could be expressed in terms of the reference position and periastron  $\lambda_k = \lambda_1 +$

$2\pi(k-1)/3$ :

$$\begin{cases} x_k(t) = \cos \lambda_k x_{\text{ref},k}(t) - \sin \lambda_k y_{\text{ref},k}(t), \\ y_k(t) = \sin \lambda_k x_{\text{ref},k}(t) + \cos \lambda_k y_{\text{ref},k}(t), \\ z_k(t) = z_{\text{ref},k}(t). \end{cases} \quad (5)$$

In summary, the orbits of 3 SCs are determined by 4 parameters: the semi-major axis  $a$ , the arm length  $L$ , SC<sub>1</sub>'s initial mean anomaly  $m_{0,1}$ , and SC<sub>1</sub>'s initial periastron  $\lambda_1$ .

### 3.2 Light travel time

Following Ref. [33], we could use the Taylor expansion of the position of the emitter SC to calculate the light travel time analytically. So the light travel time could be expressed as

$$\text{LTT} = \text{LTT}^{(0)} + \text{LTT}^{(1)} + \text{LTT}^{(2)} + \text{LTT}^{(\text{Sh})} \quad (6)$$

with

$$\text{LTT}^{(0)}(t) = \frac{1}{c} |\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)| \quad (7)$$

$$\text{LTT}^{(1)}(t) = \frac{\mathbf{v}_{\text{recv}}(t)}{c^2} \cdot (\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)) \quad (8)$$

$$\begin{aligned} \text{LTT}^{(2)}(t) &= \frac{|\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)|}{c^3} \\ &\cdot \left[ \mathbf{v}_{\text{recv}}^2(t) - \mathbf{a}_{\text{recv}}(t) \cdot (\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)) \right. \\ &\quad \left. + \left( \frac{\mathbf{v}_{\text{recv}}(t) \cdot (\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t))}{|\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)|} \right)^2 \right], \end{aligned} \quad (9)$$

where the subscripts “send” and “recv” represent the emitter SC and the receiver SC respectively. Finally, we consider the Shapiro delay caused by the Sun's gravitational field, which is a second-order effect [34]:

$$\begin{aligned} \text{LTT}^{(\text{Sh})} &= \\ &= \frac{R_S}{c} \cdot \ln \frac{|\mathbf{x}_{\text{send}}(t)| + |\mathbf{x}_{\text{recv}}(t)| + |\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)|}{|\mathbf{x}_{\text{send}}(t)| + |\mathbf{x}_{\text{recv}}(t)| - |\mathbf{x}_{\text{recv}}(t) - \mathbf{x}_{\text{send}}(t)|}, \end{aligned} \quad (11)$$

where  $R_S$  is the Schwarzschild radius of the Sun.

### 3.3 Detector response

For the data generation, we use a time domain detector response and the GPU accelerated code provided by [35], which enables us to generate a large amount of data directly in the time domain. First, we transform the GW waveform from the source frame to the solar-system barycenter (SSB) frame by

$$\begin{pmatrix} h_+^{\text{SSB}} \\ h_\times^{\text{SSB}} \end{pmatrix} = \begin{pmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{pmatrix} \begin{pmatrix} h_+^{\text{src}} \\ h_\times^{\text{src}} \end{pmatrix}. \quad (12)$$



Then, the length variation of one arm caused by the GW signal is

$$H(t) = h_+^{\text{SSB}}(t)\xi_+(\hat{\theta}, \hat{\phi}, \mathbf{n}(t)) + h_\times^{\text{SSB}}(t)\xi_\times(\hat{\theta}, \hat{\phi}, \mathbf{n}(t)), \tag{13}$$

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

$$\begin{cases} \xi_+(\mathbf{u}, \mathbf{v}, \mathbf{n}) = (\mathbf{u} \cdot \mathbf{n})^2 - (\mathbf{v} \cdot \mathbf{n})^2, \\ \xi_\times(\mathbf{u}, \mathbf{v}, \mathbf{n}) = 2(\mathbf{u} \cdot \mathbf{n})(\mathbf{v} \cdot \mathbf{n}), \end{cases} \tag{14}$$

$\hat{n}(t)$  is the unit vector of the arm,  $\hat{\theta}, \hat{\phi}$  are polar and azimuthal angle in the SSB frame, based on the orthonormal basis vectors  $(\hat{e}_r, \hat{e}_\theta, \hat{e}_\phi)$ .  $\mathbf{u}, \mathbf{v}$  are polarization vectors defined by  $\mathbf{u} = -\hat{e}_\phi$  and  $\mathbf{v} = -\hat{e}_\theta$ . The propagation vector is  $\mathbf{k} = -\hat{e}_r$ . Light travel time along the arm affected by the GW is

$$t_{\text{recv}} \simeq t_{\text{send}} + \text{LTT} - \frac{1}{2c} \int_0^L H(\mathbf{x}(\lambda), t(\lambda)) d\lambda, \tag{15}$$

where we use first order approximation as  $t(\lambda) = t_{\text{send}} + \lambda/c, \mathbf{x}(\lambda) = \mathbf{x}_{\text{send}}(t_{\text{send}}) + \lambda\mathbf{n}(t_{\text{send}})$ ,  $\mathbf{x}_{\text{send}}$  is the position of the emitter SC. The parameter  $\lambda$  used as the integration variable describes the photon path. With those approximations,  $H(\mathbf{x}(\lambda), t(\lambda))$  is given by

$$\begin{aligned} H(\mathbf{x}(\lambda), t(\lambda)) &= H\left(t(\lambda) - \frac{\mathbf{k} \cdot \mathbf{x}(\lambda)}{c}\right) \\ &= H\left(t_{\text{send}} - \frac{\mathbf{k} \cdot \mathbf{x}_{\text{send}}(t_{\text{send}})}{c} + \lambda \frac{1 - \mathbf{k} \cdot \mathbf{n}(t_{\text{send}})}{c}\right). \end{aligned} \tag{16}$$

Combining Eqs. (15) and (16) and differentiating the resulting expression with respect to  $t_{\text{send}}$  yields the relative frequency shift,

$$\begin{aligned} y(t_{\text{send}}) &= \frac{1}{2[1 - \mathbf{k} \cdot \mathbf{n}(t_{\text{send}})]} \left[ H\left(t_{\text{send}} - \frac{\mathbf{k} \cdot \mathbf{x}_{\text{send}}(t_{\text{send}})}{c}\right) \right. \\ &\quad \left. - H\left(t_{\text{send}} - \frac{\mathbf{k} \cdot \mathbf{x}_{\text{recv}}(t_{\text{recv}})}{c} + \text{LTT}\right) \right]. \end{aligned} \tag{17}$$

We assume that  $\mathbf{n}(t_{\text{send}}) \simeq \mathbf{n}(t_{\text{recv}})$  since the variation of arm direction due to the SC's motion during the light travel time is less than  $10^{-3}$  rad in the case of Taiji [36] and  $t_{\text{recv}} \simeq t_{\text{send}} + \text{LTT}$ . Finally, the GW strain of one arm is

$$\begin{aligned} y(t_{\text{recv}}) &= \frac{1}{2(1 - \mathbf{k} \cdot \mathbf{n}(t_{\text{recv}}))} \left[ H\left(t_{\text{recv}} - \frac{\mathbf{k} \cdot \mathbf{x}_{\text{send}}(t_{\text{recv}})}{c}\right) \right. \\ &\quad \left. - \text{LTT}(t_{\text{recv}}) - H\left(t_{\text{recv}} - \frac{\mathbf{k} \cdot \mathbf{x}_{\text{recv}}(t_{\text{recv}})}{c}\right) \right]. \end{aligned} \tag{18}$$

### 3.4 TDI combination

Due to the unequal arm length of the Taiji SCs, the

laser frequency noise could not be canceled naturally, which will be orders of magnitude above the GW signals. So the TDI technique was developed to cancel the noise by shifting the data from different SCs by subtle time delays and combining them together. The TDI technique can be divided into several generations:

- TDI 1.0 is used in the static case, which does not consider the rotation and breathing effect. In other words, the arm lengths are constants in time, i.e., the light travel time satisfies  $\text{LTT}_i = \text{LTT}_{i'} = \text{const.}$
- TDI 1.5 is used in the rigid rotating case, which does not consider the breathing effect. The arm lengths are still constants in time, i.e.,  $\text{LTT}_i = \text{const.}, \text{LTT}_{i'} = \text{const.}$  but  $\text{LTT}_i \neq \text{LTT}_{i'}$  here.
- TDI 2.0 is used in the flexing case, the arm lengths are functions of time, i.e.,  $\text{LTT}_i = \text{LTT}_i(t), \text{LTT}_{i'} = \text{LTT}_{i'}(t)$ , and  $\text{LTT}_i(t) \neq \text{LTT}_{i'}(t)$ .

Here we adopt the notation from Ref. [27], the number in the subscript represents different links, where  $i \in \{1, 2, 3\}$  denotes the clockwise link opposite to SC<sub>*i*</sub> and  $i'$  denotes the counter-clockwise link. The first generation unequal arm Michelson combination is

$$\begin{aligned} X_1(t) &= (y_{2':322'} + y_{1:22'} + y_{3:2'} + y_{1'}) \\ &\quad - (y_{3:2'3'} + y_{1':3'3} + y_{2':3} + y_{1'}). \end{aligned} \tag{19}$$

The second-generation Michelson combination is

$$\begin{aligned} X_2(t) &= y_{1'} + y_{3,2'} + y_{1,22'} + y_{2',322'} + y_{1,3'322'} \\ &\quad + y_{2',33'322'} + y_{1',3'33'322'} + y_{3,2'3'33'322'} \\ &\quad - y_{1'} - y_{2',3} - y_{1',3'3} - y_{3,2'3'3} - y_{1',22'3'3} \\ &\quad - y_{3,2'22'3'3} - y_{1,22'22'3'3} - y_{2',322'22'3'3}. \end{aligned} \tag{20}$$

The variables  $Y$  and  $Z$  could be obtained by cyclic permutation of the indices. The complete derivation and detailed explanation of these equations can be found in Ref. [27]. The colon denotes time delay when the arm length is time-independent,

$$\begin{aligned} f(t)_{:j} &:= f\left(t - \frac{L_j}{c}\right), \\ f(t)_{:jk} &= f\left(t - \frac{L_k}{c} - \frac{L_j}{c}\right) = f(t)_{:kj}, \end{aligned} \tag{21}$$

which is commute for  $j \neq k$ . Next, a comma denotes time delays with time-dependent arm lengths, here we compute chained delays as simple sums of delays rather than nested delays,

$$f(t)_{,jk} = f\left(t - \frac{L_k(t)}{c} - \frac{L_j(t)}{c}\right). \tag{22}$$

This approximation is sufficient for computing the GW response function [35].

After performing the TDI combination, we obtain the  $X, Y,$  and  $Z$  channel gravitational wave strain data. Simulating the instrumental noise is another important aspect of the data simulation process, distinct from the

strain data. In the next subsection, we will delve into the detailed process of simulating instrumental noise.

### 3.5 Instrumental noise

Despite the complex noise sources that will be involved in the Taiji detection data in practice, in our simulation, we use a noise model consisting of just two components for simplicity. The first one is the high-frequency component, which represents the overall noises that come from the optical metrology system, and the second is the low-frequency component, which represents the test mass acceleration noise. The power spectral density (PSD) is given by

$$P_{\text{oms}}(f) = 64 \times 10^{-24} \frac{1}{\text{Hz}} \left[ 1 + \left( \frac{2 \text{ mHz}}{f} \right)^4 \right] \left( \frac{2\pi f}{c} \right)^2,$$

$$P_{\text{acc}}(f) = 9 \times 10^{-30} \frac{1}{\text{Hz}} \left[ 1 + \left( \frac{0.4 \text{ mHz}}{f} \right)^2 \right] \cdot \left[ 1 + \left( \frac{f}{8 \text{ mHz}} \right)^4 \right] \left( \frac{1}{2\pi f c} \right)^2. \quad (23)$$

Following [37], the analytic model of the one-sided noise PSD is derived from the noise components in each interferometric measurement. We will explicitly show PSD for a Michelson-type TDI generator. Given the data combination (19) and (20) with the assumption that the noise of each link has the same PSD, the noise PSD of the first and the second generation TDI X channel is

$$\text{PSD}_{X_1} = 16 \sin^2(\omega L) \{P_{\text{oms}} + [3 + \cos(2\omega L)]P_{\text{acc}}\}, \quad (24)$$

$$\text{PSD}_{X_2} = 64 \sin^2(\omega L) \sin^2(2\omega L) \cdot \{P_{\text{oms}} + [3 + \cos(2\omega L)]P_{\text{acc}}\}, \quad (25)$$

where  $\omega = 2\pi f/c$ . In this paper, we use TDI 2.0 PSD for noise generation.

## 4 Dataset

### 4.1 Overall setting

The TDC dataset was designed based on the scientific objectives of Taiji. For most of the sources of Taiji, we do not have well-established data analysis algorithms because they are not detected by ground-based detectors. So at the beginning of the data challenge, we should focus on each source separately for simplicity. Each of the datasets from TDC-1 to TDC-6 contains signals from a single source. Multiple sources are included in the last dataset TDC-7, including MBHBs, VGBs, and GB foregrounds. Those datasets cover all the sources for the scientific objectives. Similar to LDC, the sampling rate

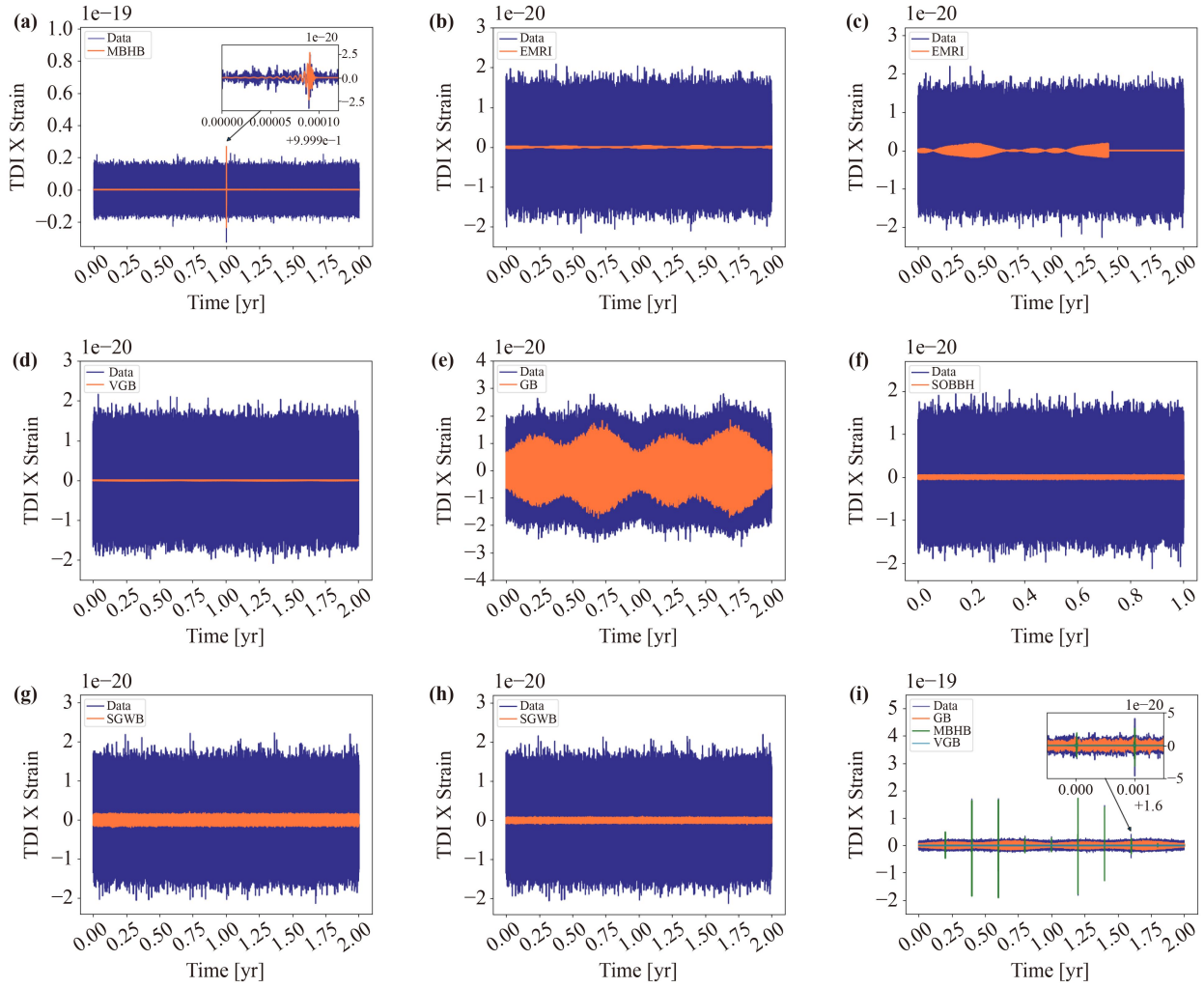
here in TDC is 0.1 Hz for all datasets, while this setting will be adjustable with Taiji's configuration accordingly. First, we generate the signal, then compute the TDI 2.0 response, and add it to the simulated noise. For the time domain signals, we use the time domain TDI response as mentioned in Sections 3.3 and 3.4. For the stochastic GW signals, we generate them directly by their PSD, which considers TDI response. Next, we will introduce the configuration and parameters used to generate the datasets.

### 4.2 Massive black hole binary

MBHBs are primary sources of Taiji, which is sensitive to MBHBs with total mass in the range of  $[10^5, 10^8]M_{\odot}$  [38]. It is expected to detect 1–20 MBHBs per year [39]. We will be able to gain valuable insight into how MBHBs are formed by measuring their spins and masses. It should be possible for us to test the hypothesis that modern galaxies are the result of the merger of smaller “seed” galaxies in the early universe, as we will be able to detect the merger of MBHB out to a high redshift. We use SEOBNRv4\_opt [39] here to generate the MBHB signal and shift it by the coalescence time  $t_c$ . Only the  $l = m = 2$  mode is included in the data. Both TDC-1 and TDC-7 contain MBHB signals. For TDC-1, there is only one MBHB signal in this dataset, the time domain and the frequency domain behavior of the TDC-1 data are shown in Fig. 1(a) and Fig. 2(a) respectively. In the time domain, we could see that the merger portion of the signal is not buried in the noise, which has a short duration of  $\sim$  hour. Then, in the frequency domain, the signal is below the noise PSD. The parameter of MBHB signals injected in those two datasets are shown in Table 3, where  $M_T$  is the total mass of two black holes,  $q$  is the mass ratio with  $q < 1$ ,  $s_1$  and  $s_2$  are the spin parameters,  $\iota$  is the inclination angle,  $\psi$  is the polarization angle,  $\lambda$  and  $\beta$  are ecliptic longitude and ecliptic latitude characterize the sky location of the source,  $t_c$  is the coalescence time, and  $\phi_c$  is coalescence phase. TDC-1 only contains signal No. 5 and all 10 signals are included in TDC-7. The time domain and the frequency domain behavior of the TDC-7 data are shown in Fig. 1(i) and Fig. 2(i) respectively. Signals No. 8 and No. 9 are close together, presenting additional challenges to data processing algorithms that are not encountered by ground-based detectors.

### 4.3 Extreme-mass-ratio inspiral

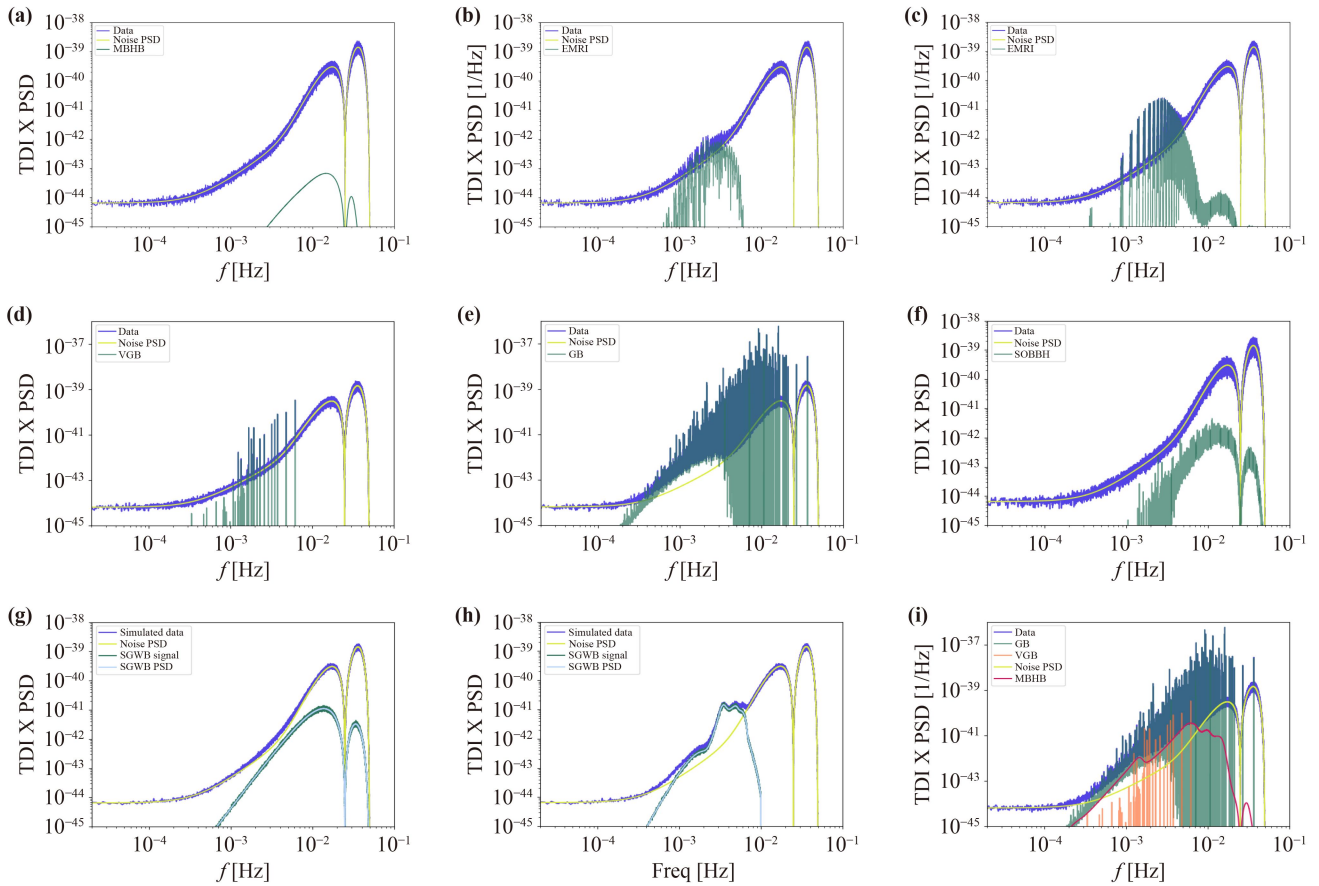
EMRIs are binary black hole systems consisting of a stellar-mass compact object (CO) and a massive black hole with mass  $M \sim 10^4 M_{\odot} - 10^7 M_{\odot}$ . Taiji will observe few tens to hundreds EMRI events over 2 years with signal-to-noise ratio (SNR) above 20 [40]. Because of the



**Fig. 1** Time-domain data of all the TDC datasets. The strain of TDI X channel data and signals are presented. (a) TDC-1, the MBHB signal is zoomed in because of its duration. (b, c) TDC-2-1 and TDC-2-2. (d) TDC-3. (e) TDC-4. (f) TDC-5. (g, h) TDC-6-1 and TDC-6-2. (i) TDC-7, signals No. 8 and No. 9 are zoomed in due to the closeness of their coalescence time.

extremely large mass ratio, the EMRI evolves slowly, inspiraling about  $10^4$ – $10^5$  cycles in Taiji’s frequency band [41]. Since EMRI signals contain rich information about the geometry surrounding the central black holes, they are one of the primary fundamental physics goals of the Taiji mission. It has been suggested that detections of EMRIs could provide a highly accurate observational test of the “Kerr-ness” of the central MBH [42]. What’s more, EMRIs sample the stellar population near the central black holes. By detecting EMRI signals, we can probe the physical properties of the central MBH and test GR [5]. Several waveform templates are developed for such a promising source of Taiji i) Teukolsky-based method [43, 44], ii) Numerical Kludge (NK) [45], iii) Analytic Kludge (AK) [46], iv) Augmented Analytic Kludge (AAK) [47–49], and v) XSPEG [26, 50]. The first 2 of them are computationally expensive, and the AK waveform does not suit Kerr space-time, so we utilize the GPU-accelerated AAK waveform [49] for the

TDC-2-1 dataset generation and the XSPEG waveform for the TDC-2-2 dataset generation. There is only one EMRI signal in each dataset. Because of the KRZ metric used in XSPEG waveform construction, TDC-2-2 could be used to test GR, which is a key scientific objective of the Taiji mission. The time domain and the frequency domain behavior of the data are shown in Fig. 1(b) and Fig. 2(c). Unlike MBHBs, the amplitude of a typical EMRI is an order of magnitude below the instrumental noise. Detection is only possible if enough SNR is accumulated over a number of wave cycles. The parameters of EMRI waveform injected in TDC-2-1 is  $(M, \mu, a_0, e_0, p_0, \iota_0, \theta_S, \phi_S, \theta_K, \phi_K, \Phi_{\varphi,0}, \Phi_{\theta,0}, \Phi_{r,0}, D_L) = (10^6 M_\odot, 30 M_\odot, 0.6, 0.6, 15, 0.7, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 0, 0, 0, 100 \text{Mpc})$ ,  $M$  and  $a$  are the mass and spin parameter of the MBH,  $p$  is the semi-latus rectum,  $e$  is eccentricity,  $\iota$  is the orbit’s inclination angle from the equatorial plane,  $\theta_S$  and  $\phi_S$  are the polar and azimuthal sky location angles.  $\theta_K$  and  $\phi_K$  are the azimuthal and polar angles describing the orientation



**Fig. 2** Frequency domain data of all the TDC datasets. The PSD of TDI X channel data and signals are compared with the noise PSD. (a) TDC-1. (b, c) TDC-2-1 and TDC-2-2. (d) TDC-3. (e) TDC-4. (f) TDC-5. (g, h) TDC-6-1 and TDC-6-2. (i) TDC-7.

**Table 3** Parameters of MBHB signals used in TDC-1 and TDC-7.

No.	$M_T(M_\odot)$	$q$	$s_1$	$s_2$	$D_L(\text{Gpc})$	$\iota$	$\psi$	$\lambda$	$\beta$	$t_c(\text{yr})$	$\phi_c$
1	$1.2 \times 10^6$	0.3	0.1	0.2	53.39	$2\pi/3$	$\pi/3$	$\pi/10$	$\pi/8$	0.2	1
2	$1.8 \times 10^6$	0.8	0.14	0.12	16.02	$\pi$	$\pi/3$	$\pi/15$	$5\pi/4$	0.4	0.05
3	$2.4 \times 10^6$	0.4	0.18	0.08	21.36	$5\pi/6$	$\pi/3$	$\pi/20$	$7\pi/4$	0.6	0.1
4	$1.5 \times 10^6$	0.7	0.14	0.8	53.39	$\pi/3$	$\pi/3$	$\pi/25$	$3\pi/4$	0.8	0.15
5	$6 \times 10^5$	0.5	0.2	0.4	53.39	$\pi/3$	$\pi/3$	$\pi/5$	$\pi/4$	1	0.5
6	$1.2 \times 10^6$	0.6	0.1	0.6	106.78	$\pi/6$	$2\pi/3$	$\pi/30$	$5\pi/4$	1.2	0.35
7	$2.4 \times 10^6$	0.1	0.22	0.48	16.02	$4\pi/15$	$5\pi/3$	$\pi/35$	$\pi/8$	1.4	0.45
8	$1.2 \times 10^6$	0.05	0.16	0.28	26.69	$13\pi/30$	$\pi/3$	$2\pi/5$	$\pi/12$	1.6	1
9	$1.8 \times 10^6$	0.5	0.02	0.08	32.03	$14\pi/30$	$\pi/3$	$3\pi/5$	$11\pi/40$	1.601	0.4
10	$1.2 \times 10^7$	1	0.06	0.08	10.68	$\pi/3$	$\pi/3$	$6\pi/5$	$\pi/4$	1.8	0.5

of the spin angular momentum vector of the MBH.  $\Phi_{\varphi,0}$ ,  $\Phi_{\theta,0}$ , and  $\Phi_{r,0}$  represent the phase of azimuthal, polar, and radial modes respectively, and  $D_L$  is the luminosity distance. The parameters of the EMRI waveform injected in TDC-2-2 is  $(M, \mu, a_0, e_0, p_0, \iota_0, \lambda, \beta, D_L) = (10^6 M_\odot, 10 M_\odot, 0.9, 0.6, 15, \pi/4, 0, \pi/4, 8.3\text{Mpc})$ , and the KRZ metric deformation parameter is  $(\delta_1, \delta_2, \delta_3) = (0.1, 0.1, 0.1)$ . Compared to the existing LDC1-2 dataset for EMRI, an

updated AAK waveform template is used in TDC-2-1, and a waveform based on the KRZ metric is used in TDC-2-2.

#### 4.4 Galactic binary

We use the simple GB model from Ref. [35], which is expressed in the source frame as



$$\begin{aligned}
 h_+^{\text{src}}(t) &= \mathcal{A}(1 + \cos^2 \iota) \cos \Phi(t), \\
 h_\times^{\text{src}}(t) &= 2\mathcal{A} \sin \iota \sin \Phi(t), \\
 \Phi(t) &= \phi_0 + 2\pi f_0 t + \pi \dot{f}_0 t^2 + \frac{\pi}{3} \ddot{f}_0 t^3, \\
 \ddot{f}_0 &= \frac{11}{3} \frac{\dot{f}_0^2}{f_0}.
 \end{aligned} \tag{26}$$

where  $\mathcal{A}$  is the overall amplitude,  $\phi_0$  is the initial phase at the start of the observation,  $\iota$  is the inclination of the BWD orbit to the line of sight from the origin of the SSB frame. The intrinsic parameter is the frequency of the signal  $f$  and its derivative  $\dot{f}$ . The comparison to LDC's GB waveform  $\ddot{f}_0$  is considered here. There are enormous GBs populated in the Milky Way [5] which form foreground noise, and disentangling those signals is a challenging task [24]. With the help of GWs, GB foreground observations provide a key method for measuring the properties of a fraction of the galactic stellar population that is routinely assumed but has never been directly observed. What's more, there are some bright GBs called LISA verified binaries [51], the latest catalog is available online<sup>2)</sup>. Following the above catalog, we generate the TDC-3 dataset, which contains 43 VGBs. The foreground GB noise is in the TDC-4 dataset, which consists of  $\sim 3e7$  GB signals. The catalog of the TDC-4 dataset comes from the population study of GBs [52, 53]. Both the signal and the noise PSD of these two datasets are shown in Figs. 2(d) and (e), associated with the corresponding time domain data shown in 1d and 1e. Compared to LDC1's dataset, we use TDI 2.0 here. That GB noise is also included in TDC-7. The Taiji mission faces a data analysis challenge because of the presence of non-Gaussian and non-stationary astrophysical foregrounds. Due to the presence of foreground signals, GW source separation procedures will be complicated. A resolvable signal's source parameter estimation will also be affected by the level of confusion noise. Therefore, it is critical to study, understand, and predict the overall shape and amplitude of the potential foreground components of the Taiji data.

#### 4.5 Stellar-origin binary black holes

Space-based detectors observe the early inspiral of SOBBH systems, whose merger will be detected by ground-based detectors. Those systems evolve slowly in the LISA and Taiji frequency bands, which enable high-precision parameter estimation [54]. The joint observation of SOBBH by Taiji and ground-based detectors could be used to probe low-frequency modifications due to deviations from GR or to environmental effects, to facilitate electromagnetic joint observations [25, 55]. By estimating coalescence time and sky location, we could forecast the signals for ground-based detectors. We generate the

TDC-5 dataset, which contains 21721 SOBBH signals, whose catalog comes from [25]. The time domain and the frequency domain behavior of the data are shown in Fig. 1(f) and Fig. 2(f). The SOBBH signals are completely below the noise in the frequency domain, which are background signals. We generate the waveform using the IMRPhenomD template, then transform it to the time domain and apply the TDI response. The main difference between the TDC-5 and LDC1-5 datasets is that we use the TDI 2.0 response in the time domain rather than the TDI 1.0 response in the frequency domain.

#### 4.6 Stochastic gravitational wave background

One of the main scientific objectives of Taiji is to probe GWs from the early universe and reveal physics of GW sources [56]. Fundamental physics and astronomy would be greatly enhanced by the detection of a cosmological background from the early universe. The only way to obtain direct information about earlier epochs than decoupling may be through GWs and possibly neutrinos. There are many possible astrophysical and cosmological sources that contribute to the stochastic background, and up to now we put only upper bounds on its amplitude [57], and on parameters characterizing its directional properties [58], by the LIGO/Virgo collaboration. For space-based detection, we just have a single detector, cross-correlation based method is not suitable now, so a new SGWB detection algorithm is needed.

SGWBs that are Gaussian, isotropic, and stationary can be fully described by their energy density spectrum [59]:

$$\Omega_{GW}(f) = \frac{f}{\rho_c} \frac{d\rho_{GW}}{df}, \tag{27}$$

where  $\rho_{GW}$  is the energy density of gravitational radiation contained in the frequency range  $f$  to  $f + df$ ,  $\rho_c = \frac{3H_0^2 c^2}{8\pi G}$  is the critical density of the universe, where  $c$  is the speed of light, and  $G$  is Newton's constant, and  $H_0$  is the Hubble constant. Typical SGWB sources include compact binary coalescence [57], cosmic phase transitions [60], reheating or preheating after inflation [61], and primordial scalar and tensor perturbations from inflation [62]. There are 2 datasets that contain different types of SGWB signals, TDC-6-1 is generated by the power-law spectrum:

$$h^2 \Omega_{GW}(f) = 10^\alpha \left( \frac{f}{f_*} \right)^\beta, \tag{28}$$

where  $h$  is the dimensionless Hubble constant,  $f_*$  is the pivot frequency,  $\alpha$  characterize its amplitude at  $f_*$  and  $\beta$  is the slope of the spectrum. The SGWB signal in TDC-6-1 is generated using the parameters  $\alpha = -12$  and

<sup>2)</sup> URL: [gitlab.in2p3.fr/LISA/lisa-verification-binaries](https://gitlab.in2p3.fr/LISA/lisa-verification-binaries)

**Table 4** Summary of Taiji data challenge datasets.

Name	Scientific objective	Technical challenge	Signals	Waveform model
TDC-1	1, 5, 6	7	1 MBHB signal	SEOBNRv4_opt
TDC-2-1	2, 5, 6	7	1 EMRI signal	PN5AAK
TDC-2-2	2, 5, 6	7	1 EMRI signal	XSPEG
TDC-3	3, 5, 6	7	43 VGB signals	Sinusoidal
TDC-4	3, 5, 6	1, 7	$3 \times 10^7$ GB signals	Sinusoidal
TDC-5	4, 5, 6	2, 7	$2 \times 10^5$ SOBBH signals	IMRPhenomD
TDC-6-1	5, 6, 7	7	SGWB	Power-law
TDC-6-2	5, 6, 7	7	SGWB	Non-Gaussian curvature perturbations
TDC-7	1, 3, 5, 6	1, 3, 7	10 MBHBs, 43 VGBs, $3 \times 10^7$ GBs	SEOBNRv4_opt, Sinusoidal

$\beta = 2/3$ .

TDC-6-2 is generated by non-Gaussian scalar perturbation-induced SGWB signals. The induced GWs are generated during the formation of primordial black holes (PBHs), providing a powerful tool to search for PBHs [63]. Typically, curvature perturbations are assumed to have Gaussian probability density functions. There is a natural expectation that non-Gaussianity will be present when a sharp peak appears in the power spectrum of curvature perturbations. GWs induced by scalar perturbations with a second-order local-type non-Gaussianity were studied in Ref. [62]. The formulae of the SGWB spectrum used to generate TDC-6-2 data is given in equation (7) in Ref. [62] with the parameters  $F_{NL} = 10, \sigma = 10^{-4}, M_{PBH} = 10^{22}g$ , where  $F_{NL}$  characterize the non-Gaussianity,  $\sigma$  is the peak width of curvature perturbation spectrum, and  $M_{PBH}$  is the mass of PBH with the assumption that PBHs can serve as all dark matter.

In summary, 9 datasets are designed for TDC, and their essential motivation (i.e., the scientific objective and the technical challenge), and configurations are presented in Table 4. The scientific objectives and the corresponding technical challenges are shown in Table 1 and Table 2.

## 5 Application

### 5.1 File format

The TDC consists of multiple datasets, with each dataset stored in a separate HDF file, which can be accessed and downloaded from the dataset page on our TDC website. Within the HDF file, all the data is organized under the attribute `TDIdata`. Specifically, there is a data array of dimensions  $(4, n\_samples)$ , where `n_samples` represents the total number of sampling points in the time domain. Each row of the array corresponds to a specific aspect of the data, namely time, TDI X strain, TDI Y strain, and TDI Z strain. This structured

arrangement ensures easy and efficient access to the desired information within the dataset.

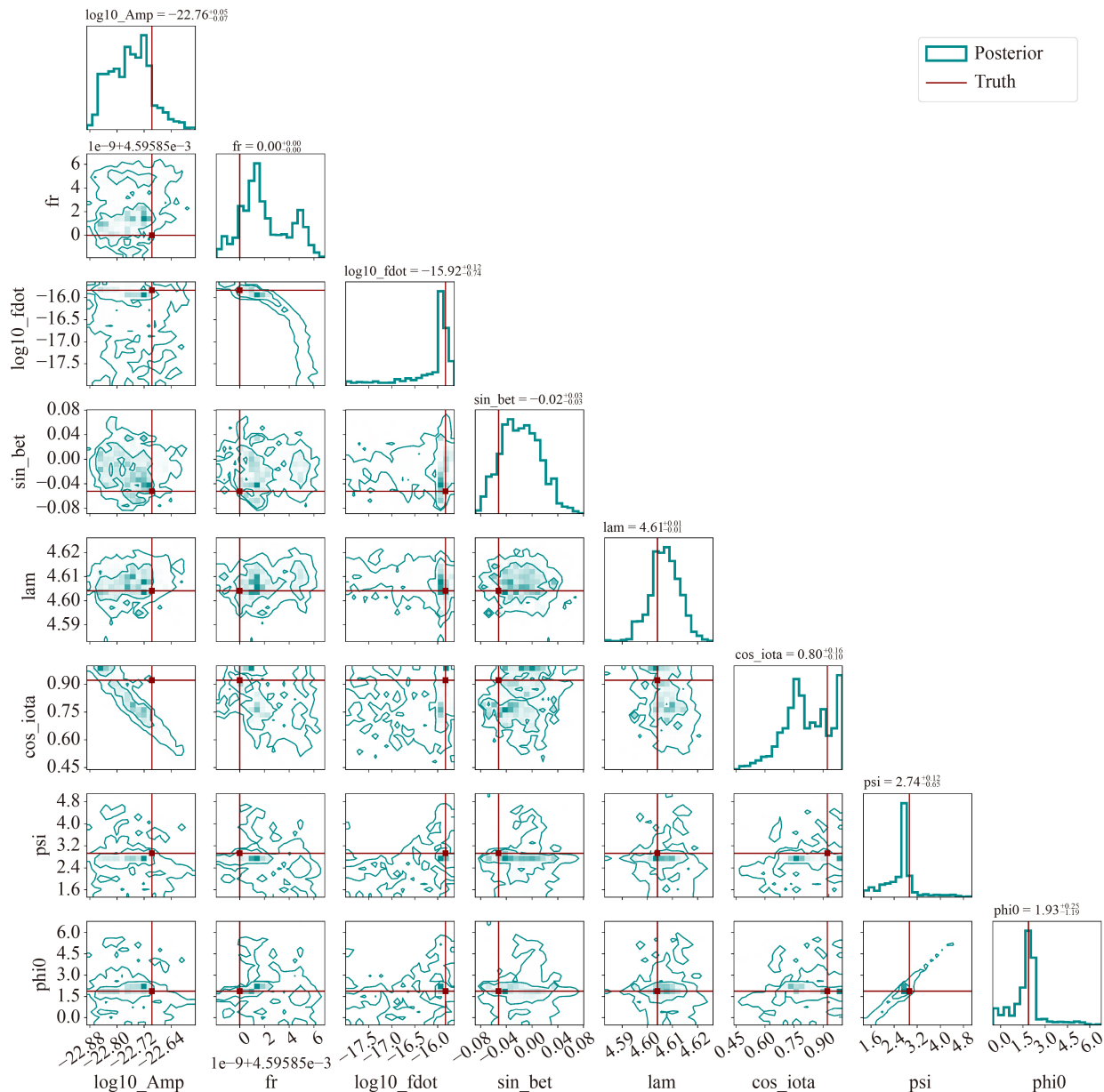
### 5.2 Toy dataset

In this subsection, we present a simple toy dataset, which provides participants with an effective means to validate their pipelines and assess the accuracy of their parameter estimation techniques. The toy dataset has been intentionally designed to be straightforward, featuring only one gravitational wave signal originating from a GB system. By simplifying the dataset to a single signal, participants can precisely evaluate their pipelines without the complexities introduced by waveform models or multiple signals. This simplified setup provides a clear and unambiguous environment for pipeline validation and establishes a solid foundation for accurate parameter estimation. Using this toy dataset, participants have the opportunity to rigorously evaluate the performance of their pipelines. They can assess their pipelines' ability to extract relevant parameters, such as the gravitational wave frequency, amplitude, and arrival time. The well-defined characteristics of the dataset allow participants to directly compare their results, identify any discrepancies, and refine their pipelines accordingly.

To facilitate the effective utilization of the toy dataset, we offer an illustrative example of running a Markov Chain Monte Carlo (MCMC) parameter estimation pipeline. The MCMC pipeline utilized in this demonstration is based on the work by Falxa *et al.* [64], and the source code can be accessed at the provided link<sup>3)</sup>. This practical demonstration serves as a comprehensive guide, showcasing how the dataset can be leveraged to obtain reliable parameter estimates.

Furthermore, we showcase the outcomes of the parameter estimation process in Fig. 3, illustrating the attained accuracy and precision through the implementation of the MCMC pipeline on the toy dataset. The visual depiction of the estimated parameters, juxtaposed with the corresponding true values, offers a comprehensive

<sup>3)</sup> URL: [gitlab.in2p3.fr/lisa-apc/m3c2](https://gitlab.in2p3.fr/lisa-apc/m3c2)



**Fig. 3** MCMC posterior corner plot for the toy dataset. The MCMC posterior analysis is conducted on our toy dataset, utilizing the following source parameter values:  $(\mathcal{A}, f, \dot{f}, \lambda, \beta, \iota, \phi_0, \psi) = (2.008192 \times 10^{-23}, 0.00459585, 1.469823 \times 10^{-16}, 4.604082, -0.052144, 0.39811, 1.8777623071795864, 2.927109)$ . The posterior distribution is represented by the dark cyan color, while the true parameter values are indicated by the red lines.

overview of the pipeline's performance and establishes a benchmark for participants to assess and compare their own findings.

By introducing the practical demonstration of the MCMC pipeline and providing visual representations of parameter estimation results, we not only equip participants with the necessary tools to effectively utilize the toy dataset but also prepare them for the real challenge posed by other TDC datasets. The experience gained from working with the toy dataset serves as a valuable cornerstone, allowing participants to build their skills and confidence before tackling the complexities of the

actual challenge datasets. The insights and techniques acquired during this process will enable participants to face TDC datasets with increased proficiency and contribute to advancements in space-based gravitational wave data analysis.

## 6 Discussion

Building a data challenge is critical for scientists to develop algorithms and pipelines for future data analysis tasks. We provide a complete workflow by constructing

detection data and waveform templates within and beyond GR. The datasets and challenges are available on the official website of TDC (URL: taiji-tdc.ictp-ap.org). Scientists are encouraged to participate in these data challenges and to submit their results. Currently, there is no deadline, and users are free to submit their results at any time. We will rank the submitted results for each challenge. As part of our website, we offer both data downloads and data customization services, which enable users to submit the data configuration file suited to their requirements, and we will provide users links to download the customized dataset. With TDC, we hope to create a community of researchers who can collaboratively contribute to the development of Taiji's data analysis pipelines and join the journey of exploring the universe and making new discoveries with Taiji Collaboration.

**Declarations** The authors declare that they have no competing interests and there are no conflicts.

**Acknowledgements** The research was supported by the Peng Cheng Laboratory Cloud Brain and by Peng Cheng Cloud-Brain. Further funding was provided by the National Key Research and Development Program of China (Grant Nos. 2021YFC2203001, 2020YFC2201501, and 2021YFC2203002), as well as the NSFC (Nos. 11920101003, 12021003, 12173071, 12147103, 12235019, and No. 12075297). Z.C. and W.H. are supported by the CAS Project for Young Scientists in Basic Research YSBR-006. Z.C. was also supported by the Interdisciplinary Research Funds of Beijing Normal University.

## References

1. B. P. Abbott, et al. (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* 116(6), 061102 (2016)
2. The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al., GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv: 2111.03606 (2021)
3. G. González, A. Viceré, and L. Wen, Gravitational wave astronomy, *Front. Phys.* 8(6), 771 (2013)
4. F. Matichard, B. Lantz, R. Mittleman, K. Mason, J. Kissel, et al., Seismic isolation of advanced LIGO: Review of strategy, instrumentation and performance, *Class. Quantum Gravity* 32(18), 185003 (2015)
5. P. Amaro-Seoane, et al., Laser interferometer space antenna, arXiv: 1702.00786 (2017)
6. X. Gong, S. Xu, S. Bai, Z. Cao, G. Chen, Y. Chen, X. He, G. Heinzl, Y. K. Lau, C. Liu, J. Luo, Z. Luo, A. P. Patón, A. Rüdiger, M. Shao, R. Spurzem, Y. Wang, P. Xu, H. C. Yeh, Y. Yuan, and Z. Zhou, A scientific case study of an advanced LISA mission, *Class. Quantum Gravity* 28(9), 094012 (2011)
7. Z. Luo, Y. Wang, Y. Wu, W. Hu, and G. Jin, The Taiji program: A concise overview, *Prog. Theor. Exp. Phys.* 2021(5), 05A108 (2021)
8. Y. L. Wu, in: Presentation to 1st eLISA Consortium Meeting (2012)
9. Y. Q. Li, Z. R. Luo, H. S. Liu, Y. H. Dong, and G. Jin, Laser interferometer used for satellite—satellite tracking: An on-ground methodological demonstration, *Chin. Phys. Lett.* 29(7), 079501 (2012)
10. H. S. Liu, Y. H. Dong, Y. Q. Li, Z. R. Luo, and G. Jin, The evaluation of phasemeter prototype performance for the space gravitational waves detection, *Rev. Sci. Instrum.* 85(2), 024503 (2014)
11. Y. H. Dong, H. S. Liu, Z. R. Luo, Y. Q. Li, and G. Jin, Methodological demonstration of laser beam pointing control for space gravitational wave detection missions, *Rev. Sci. Instrum.* 85(7), 074501 (2014)
12. Y. Q. Li, Z. R. Luo, H. S. Liu, Y. H. Dong, and G. Jin, Path-length measurement performance evaluation of polarizing laser interferometer prototype, *Appl. Phys. B* 118(2), 309 (2015)
13. W. R. Hu and Y. L. Wu, The Taiji Program in Space for gravitational wave physics and the nature of gravity, *Natl. Sci. Rev.* 4(5), 685 (2017)
14. Z. Luo, Z. Guo, G. Jin, Y. Wu, and W. Hu, A brief analysis to Taiji: Science and technology, *Results Phys.* 16, 102918 (2020)
15. Z. Luo, Q. Wang, C. Mahrtdt, A. Goerth, and G. Heinzl, Possible alternative acquisition scheme for the gravity recovery and climate experiment follow-on-type mission, *Appl. Opt.* 56(5), 1495 (2017)
16. Z. Luo, H. Liu, and G. Jin, The recent development of interferometer prototype for Chinese gravitational wave detection pathfinder mission, *Opt. Laser Technol.* 105, 146 (2018)
17. H. Liu, Y. Dong, R. Gao, Z. Luo, and G. Jin, Principle demonstration of the phase locking based on the electro-optic modulator for Taiji space gravitational wave detection pathfinder mission, *Opt. Eng.* 57(5), 054113 (2018)
18. H. Liu, Z. Luo, and G. Jin, The development of phasemeter for Taiji space gravitational wave detection, *Microgravity Sci. Technol.* 30(6), 775 (2018)
19. W. Deng, T. Yang, J. Cao, E. Zang, L. Li, L. Chen, and Z. Fang, High-efficiency 1064 nm nonplanar ring oscillator Nd:YAG laser with diode pumping at 885 nm, *Opt. Lett.* 43(7), 1562 (2018)
20. Z. Wang, W. Sha, Z. Chen, Y. S. Kang, Z. R. Luo, M. Li, and Y. P. Li, Preliminary design and analysis of telescope for space gravitational wave detection, *Chin. Opt.* 11(1), 131 (2018)
21. The Taiji Scientific Collaboration, China's first step towards probing the expanding universe and the nature of gravity using a space borne gravitational wave antenna, *Commun. Phys.* 4, 34 (2021)
22. The Taiji Scientific Collaboration, Taiji program in space for gravitational universe with the first run key technologies test in Taiji-1, *Int. J. Mod. Phys. A* 36, 2102002 (2021)
23. A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme, and



- B. Wardell, Science with the space-based interferometer eLISA: Supermassive black hole binaries, *Phys. Rev. D* 93(2), 024003 (2016)
24. X. H. Zhang, S. D. Mohanty, X. B. Zou, and Y. X. Liu, Resolving Galactic binaries in LISA data using particle swarm optimization and cross-validation, *Phys. Rev. D* 104(2), 024023 (2021)
  25. A. Sesana, Prospects for multiband gravitational-wave astronomy after GW150914, *Phys. Rev. Lett.* 116(23), 231102 (2016)
  26. S. Xin, W. B. Han, and S. C. Yang, Gravitational waves from extreme-mass-ratio inspirals using general parametrized metrics, *Phys. Rev. D* 100(8), 084055 (2019)
  27. M. Otto, Time-Delay Interferometry Simulations for the Laser Interferometer Space Antenna, Ph. D. thesis, Hannover: Gottfried Wilhelm Leibniz Universität Hannover, 2015
  28. A. Blelly, J. Bobin, and H. Moutarde, Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data, *Mon. Not. R. Astron. Soc.* 509(4), 5902 (2021)
  29. T. Robson and N. J. Cornish, Detecting gravitational wave bursts with LISA in the presence of instrumental glitches, *Phys. Rev. D* 99(2), 024019 (2019)
  30. S. V. Dhurandhar, K. R. Nayak, S. Koshti, and J. Y. Vinet, Fundamentals of the LISA stable flight formation, *Class. Quantum Gravity* 22(3), 481 (2005)
  31. K. R. Nayak, S. Koshti, S. V. Dhurandhar, and J. Y. Vinet, On the minimum flexing of LISA's arms, *Class. Quantum Gravity* 23(5), 1763 (2006)
  32. B. Wu, C. G. Huang, and C. F. Qiao, Analytical analysis on the orbits of Taiji spacecrafts, *Phys. Rev. D* 100(12), 122001 (2019)
  33. B. Chauvineau, T. Regimbau, J. Y. Vinet, and S. Pireaux, Relativistic analysis of the LISA long range optical links, *Phys. Rev. D* 72(12), 122003 (2005)
  34. A. Hees, S. Bertone, and C. Le Poncin-Lafitte, Relativistic formulation of coordinate light time, Doppler, and astrometric observables up to the second post-Minkowskian order, *Phys. Rev. D* 89(6), 064045 (2014)
  35. M. L. Katz, J. B. Bayle, A. J. K. Chua, and M. Vallisneri, Assessing the data-analysis impact of LISA orbit approximations using a GPU-accelerated response model, *Phys. Rev. D* 106(10), 103001 (2022)
  36. N. Rijnveld and J. A. C. M. Pijnenburg, in: International Conference on Space Optics — ICSO 2010, edited by N. Kadowaki, SPIE, Rhodes Island, Greece, 2017, p. 96
  37. S. Babak, M. Hewitson, and A. Petiteau, LISA sensitivity and SNR calculations, arXiv: 2108.01167 (2021)
  38. M. L. Katz, L. Z. Kelley, F. Dosopoulou, S. Berry, L. Blecha, and S. L. Larson, Probing massive black hole binary populations with LISA, *Mon. Not. R. Astron. Soc.* 491, 2301 (2019)
  39. A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, et al., Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, *Phys. Rev. D* 95(4), 044028 (2017)
  40. J. R. Gair, S. Babak, A. Sesana, P. Amaro-Seoane, E. Barausse, C. P. L. Berry, E. Berti, and C. Sopuerta, Prospects for observing extreme-mass-ratio inspirals with LISA, *J. Phys. Conf. Ser.* 840, 012021 (2017)
  41. S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals, *Phys. Rev. D* 95(10), 103012 (2017)
  42. L. Barack and C. Cutler, Using LISA extreme-mass-ratio inspiral sources to test off-Kerr deviations in the geometry of massive black holes, *Phys. Rev. D* 75(4), 042003 (2007)
  43. K. Glampedakis, Extreme mass ratio inspirals: LISA's unique probe of black hole gravity, *Class. Quantum Gravity* 22(15), S605 (2005)
  44. W. B. Han and Z. Cao, Constructing effective one-body dynamics with numerical energy flux for intermediate-mass-ratio inspirals, *Phys. Rev. D* 84(4), 044014 (2011)
  45. S. Babak, H. Fang, J. R. Gair, K. Glampedakis, and S. A. Hughes, "Kludge" gravitational waveforms for a test-body orbiting a Kerr black hole, *Phys. Rev. D* 75(2), 024005 (2007)
  46. L. Barack and C. Cutler, LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy, *Phys. Rev. D* 69(8), 082005 (2004)
  47. A. J. K. Chua and J. R. Gair, Improved analytic extreme-mass-ratio inspiral model for scoping out eLISA data analysis, *Class. Quantum Gravity* 32(23), 232002 (2015)
  48. A. J. K. Chua, C. J. Moore, and J. R. Gair, Augmented kludge waveforms for detecting extreme-mass-ratio inspirals, *Phys. Rev. D* 96(4), 044005 (2017)
  49. M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, Fast extreme-mass-ratio-inspiral waveforms: New tools for millihertz gravitational-wave data analysis, *Phys. Rev. D* 104(6), 064047 (2021)
  50. C. Zhang, W. B. Han, and S. C. Yang, Analytical effective one-body formalism for extreme-mass-ratio inspirals with eccentric orbits, *Commun. Theor. Phys.* 73(8), 085401 (2021)
  51. T. Kupfer, V. Korol, S. Shah, G. Nelemans, T. R. Marsh, G. Ramsay, P. J. Groot, D. T. H. Steeghs, and E. M. Rossi, LISA verification binaries with updated distances from Gaia Data Release 2, *Mon. Not. R. Astron. Soc.* 480(1), 302 (2018)
  52. G. Nelemans, L. R. Yungelson, and S. F. P. Zwart, The gravitational wave signal from the Galactic disk population of binaries containing two compact objects, *Astron. Astrophys.* 375(3), 890 (2001)
  53. G. Nelemans and C. A. Tout, Reconstructing the evolution of white dwarf binaries: Further evidence for an alternative algorithm for the outcome of the common-envelope phase in close binaries, *Mon. Not. R. Astron. Soc.* 356(2), 753 (2005)
  54. J. Gair, M. Hewitson, A. Petiteau, and G. Mueller, in: Handbook of Gravitational Wave Astronomy, edited by C. Bambi, S. Katsanevas, and K. D. Kokkotas, Springer Singapore, Singapore, 2021, pp 1–71
  55. A. Toubiana, S. Marsat, S. Babak, J. Baker, and T. Dal Canton, Parameter estimation of stellar-mass black hole binaries with LISA, *Phys. Rev. D* 102(12), 124037

- (2020)
56. N. Bartolo, D. Bertacca, R. Caldwell, C. R. Contaldi, G. Cusin, et al., Probing anisotropies of the stochastic gravitational wave background with LISA, *J. Cosmol. Astropart. Phys.* 11, 009 (2022)
  57. R. Abbott, et al., (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Upper limits on the isotropic gravitational-wave background from advanced LIGO and advanced Virgo's third observing run, *Phys. Rev. D* 104(2), 022004 (2021)
  58. The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, All-sky, all-frequency directional search for persistent gravitational waves from advanced LIGO's and advanced Virgo's first three observing runs, *Phys. Rev. D* 105(12), 122001 (2022)
  59. C. Caprini, D. G. Figueroa, R. Flauger, G. Nardini, M. Peloso, M. Pieroni, A. Ricciardone, and G. Tasinato, Reconstructing the spectral shape of a stochastic gravitational wave background with LISA, *J. Cosmol. Astropart. Phys.* 11, 017 (2019)
  60. M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, Gravitational waves from the sound of a first order phase transition, *Phys. Rev. Lett.* 112(4), 041301 (2014)
  61. J. Liu, Z. K. Guo, R. G. Cai, and G. Shiu, Gravitational waves from oscillons with Cuspy potentials, *Phys. Rev. Lett.* 120(3), 031301 (2018)
  62. R. G. Cai, S. Pi, and M. Sasaki, Gravitational waves induced by non-Gaussian scalar perturbations, *Phys. Rev. Lett.* 122(20), 201101 (2019)
  63. C. Yuan and Q. G. Huang, Gravitational waves induced by the local-type non-Gaussian curvature perturbations, *Phys. Lett. B* 821, 136606 (2021)
  64. M. Falxa, S. Babak, and M. Le Jeune, Adaptive kernel density estimation proposal in gravitational wave data analysis, *Phys. Rev. D* 107(2), 022008 (2023)