



# 空间引力波探测科学数据处理的挑战与人工智能技术的应用

王赫<sup>1,2\*</sup>, 杜明辉<sup>3\*</sup>, 徐鹏<sup>2,3,4,5</sup>, 周宇峰<sup>6</sup>

- 中国科学院大学国际理论物理中心(亚太地区), 北京 100190;
- 中国科学院大学引力波宇宙太极实验室(北京/杭州), 北京 100049;
- 中国科学院力学研究所国家微重力实验室引力波实验中心, 北京 100190;
- 国科大杭州高等研究院, 杭州 310024;
- 兰州大学兰州理论物理中心, 兰州 730000;
- 中国科学院理论物理前沿重点实验室, 中国科学院理论物理研究所, 北京 100190

\*联系人, 王赫, E-mail: hewang@ucas.ac.cn; 杜明辉, E-mail: duminghui@imech.ac.cn

收稿日期: 2024-03-04; 接受日期: 2024-05-31; 网络出版日期: 2024-06-20

国家重点研发计划(编号: 2021YFC2203004, 2021YFC2201903)、国家自然科学基金(编号: 12247187, 12147103)、国家天文科学数据中心(编号: NADC2023YDS-01)和中央高校基本科研业务费专项资金资助项目

**摘要** 随着空间引力波探测项目, 如LISA、太极、天琴等的不断推进, 我们即将获得一个观察宇宙的全新视角。然而, 这些项目的科学数据处理面临着前所未有的挑战, 包括大量波源混叠、非稳态噪声、数据异常等。本文旨在为研究人员提供一个相对全面的视角, 以贝叶斯统计推断框架为线索, 综述了这些挑战及其可能的解决方案, 讨论了从波源模板构建、探测器响应模拟, 到噪声和数据异常处理的过程, 并着重探讨了全局拟合与参数反演的技术策略, 尤其是似然函数的构造与计算方法, 以及如何利用多种随机采样技术提升分析效率和准确度。特别地, 文章重点介绍了人工智能技术在引力波信号建模、噪声与数据异常处理、信号识别与参数估计等方面的应用, 展示了人工智能如何为解决空间引力波探测数据分析中的复杂问题提供新的路径和工具。

**关键词** 引力波探测, 数据处理, 参数估计, 人工智能

**PACS:** 04.30.-w, 04.80.Nn, 07.05.Kf, 07.05.Mh

## 1 引言

地面引力波探测器<sup>[1-3]</sup>, 如aLIGO, aVIRGO和KAGRA等, 已经成功观测到了上百个源自双黑洞、双中子星和黑洞-中子星系统的引力波事件<sup>[4-9]</sup>, 随着正在进行的第四轮观测的开展, 预计将会捕获到更

多的此类信号。地面引力波探测的目标频段是1 Hz–1 kHz, 而要将探测的视野深入到波源更加丰富的mHz频段, 则需要将探测器送入太空, 以突破地球空间尺度和环境噪声的限制<sup>[10,11]</sup>。

欧洲航天局提出的LISA项目<sup>[12-14]</sup>, 以及中国的太极<sup>[15-18]</sup>和天琴<sup>[19-22]</sup>计划, 都旨在捕捉0.1 mHz–

**引用格式:** 王赫, 杜明辉, 徐鹏, 等. 空间引力波探测科学数据处理的挑战与人工智能技术的应用. 中国科学: 物理学 力学 天文学, 2024, 54: 270403  
Wang H, Du M H, Xu P, et al. Challenges in space-based gravitational wave data analysis and applications of artificial intelligence (in Chinese). Sci Sin-Phys Mech Astron, 2024, 54: 270403, doi: 10.1360/SSPMA-2024-0087

1 Hz频段的引力波信号. 2024年1月25日, 欧洲航天局科学计划委员会已正式批准了LISA任务<sup>1)</sup>, 授权启动了实验仪器和航天器的建造工作, 并计划在2030–2035年间发射, 这一进展标志着空间引力波探测任务的概念与技术已达到成熟阶段.

空间引力波探测的科学目标广泛而深远, 旨在通过对宇宙中的各种极端天体和物理过程进行观测, 推动天体物理学、基础物理学和宇宙学的研究, 包括利用银河系内UCB系统(英文缩写释义见表1)的引力波信号研究其形成与演化规律, 探索银河系的结构等; 利用MBHB信号追溯宇宙尺度下大质量黑洞的起源、形成和并合的历史, 探测宇宙膨胀速率等; 利用EMRI和IMRI信号探索宇宙中黑洞的性质和环境, 理解恒星级黑洞的天体物理学, 探索引力和黑洞的基本性质等; 借助SGWB加深对早期宇宙和粒子物理学的理解; 以及寻找引力波爆发源和其他未知波源等.

为实现上述科学目标, 空间引力波探测的目标波源涵盖多种类型的天体和事件, 包括总质量在 $10^5 M_{\odot}$ – $10^8 M_{\odot}$ 范围内的MBHB系统<sup>[23]</sup>, 其并合率估计为每年 $O(2)$ – $O(100)$ 次<sup>[24–27]</sup>; IMBHB系统, 总质量范围为 $10^2 M_{\odot}$ – $10^5 M_{\odot}$ <sup>[28]</sup>; EMRI和IMRI系统, 总质量在 $10^3 M_{\odot}$ – $10^7 M_{\odot}$ 范围内, 并合时质量比分别为 $10^{-6}$ – $10^{-3}$ 和 $10^{-3}$ – $10^{-1}$ <sup>[29,30]</sup>; SBHB系统的旋近阶段<sup>[31]</sup>; 银河系内UCB系统<sup>[32–37]</sup>, 其中大多数为双白矮星系统, 还包括少量由中子星-白矮星或双中子星构成的双星系统<sup>[38,39]</sup>, 只有一小部分信号(约 $O(10^4)$ )能够被单独解析, 而其余(约 $O(10^7)$ )将在0.5–3 mHz频段内构成一个各向异性且非平稳的“混淆”信号, 被称为银河系前景噪声<sup>[40]</sup>. 此外, 我们还期待探测到SGWB、引力波爆发源, 及各种未知波源等. 图1展示了空间引力波探测器LISA、太极、天琴的设计灵敏度曲线及主要目标波源. 图中的目标波源包括: (1) MBHB: 用颜色呈阶梯状变化的曲线表示, 曲线的颜色反映距离并合的时间, 曲线的粗细反映信号强度与灵敏度的比值(以太极的灵敏度为例); (2) UCB: 用蓝色点标记, 颜色的深浅表示信噪比(以太极的灵敏度为例), 其中已通过电磁观测发现的VB以空心圆区分; (3) EMRI: 以浅绿色曲线表示; (4) SBHB: 以灰色线段表示, 图中波源的参数范围为: 每个黑洞的质量  $\in (5, 80)M_{\odot}$ , 质量比

$\in (0.1, 1)$ , 红移  $\in (0.1, 1)$ . 波源的多样性显示出空间引力波探测的广阔前景, 也对科学数据处理提出了巨大挑战, 同时为人工智能技术的应用带来了新的契机.

科学目标的实现不仅需要高灵敏度的探测设备, 也需要科学的数据处理方法. 在本文中, “科学数据处理”是指从输入经过预处理的数据(时间延迟干涉通道), 到输出波源目录和参数后验分布之间的过程, 包括引力波信号的探测和波源参数的估计等, 它搭建了探测数据和科学目标之间的桥梁. 基于贝叶斯理论的数据处理方法在地面引力波探测中取得了巨大的成功, 对未来的空间引力波探测具有重要的借鉴意义. 贝叶斯统计推断的成功实施依赖于对噪声和信号的精确建模, 而空间引力波探测在信号和噪声方面独有的复杂性, 为数据处理提出了新的科学问题和技术挑战.

目前尚无成功开展空间引力波探测的先例可供参考, 而LISA、太极和天琴的技术验证卫星LISA探路者<sup>[41]</sup>、GRACE Follow-On<sup>[42]</sup>、太极一号<sup>[15,18]</sup>和天琴一号<sup>[22]</sup>等可以为空间引力波探测的数据状况和处理方法提供启发. 例如, 从LISA探路者数据中观察到的瞬态噪声事件(Glitch)、数据间断、非稳态噪声等均有可能在未来的引力波探测数据中出现(详见第3节). 此外, 还有其他可预期的特点. 首先, 地面引力波探测数据的总特征是强噪声、弱信号, 而空间引力波探测数据的突出特征是, 数量巨大( $O(10^4)$ 以上)的可探测波源在时域和频域交叠, 因此如何实现波源的识别和分离成为了新的问题; 其次, 与地面引力波探测器不同, LISA、太极和天琴的观测设备是在日心或地心轨道上运行的, 而UCB、EMRI等信号的可观测时间与探测器轨道运动的时间尺度相近, 因而观测信号会受到轨道运动的调制作用; 此外, 在太空中无法利用与地面引力波探测类似的多种辅助手段对设备运行状态和噪声统计特性进行测量, 需要利用有限的观测数据完成设备和噪声的标定.

全局拟合(Global Fit)<sup>[34,43,44]</sup>, 即同时对所有天体物理和仪器特征进行综合建模和参数估计, 是解决上述问题的思路之一. 在过去的20年中, 全局拟合算法已成为空间引力波探测科学数据处理研究的核心课题之一. LISA、太极和天琴团队所开展的数据模拟、数据挑战等活动均为数据处理流水线的开发奠定了基

1) Capturing the ripples of spacetime: LISA gets go-ahead.

表1 英文缩写释义

Table 1 Explanations for the English shortcuts

英文缩写	英文全称	中文全称
MBHB	Massive Black Hole Binary	大质量双黑洞
IMBHB	Intermediate-Mass Black Hole Binary	中等质量双黑洞
EMRI	Extreme Mass-Ratio Inspiral	极端质量比旋近
IMRI	Intermediate Mass-Ratio Inspiral	中等质量比旋近
SBHB	Stellar-mass Black Hole Binary	恒星级质量双黑洞
UCB	Ultra-Compact Binary	超致密双星
VB	Verification Binary	验证双星
SGWB	Stochastic Gravitational Wave Background	随机引力波背景
LDC	LISA Data Challenge	LISA数据挑战
AI	Artificial Intelligence	人工智能
TDI	Time Delay Interferometry	时间延迟干涉
MCMC	Markov Chain Monte Carlo	马尔可夫链蒙特卡罗
PSD	Power Spectral Density	功率谱密度

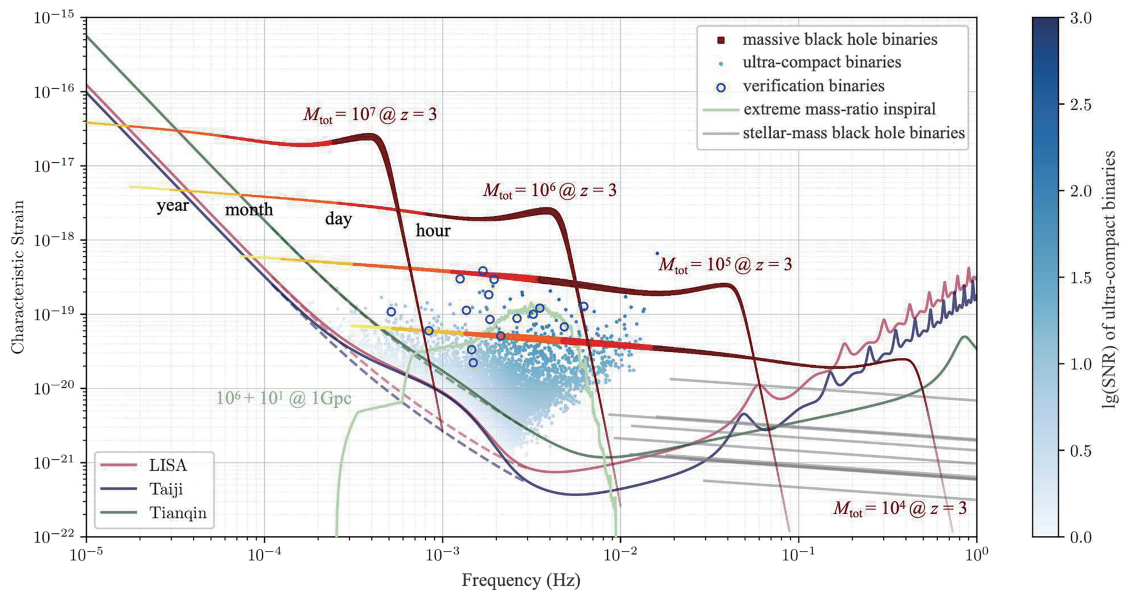


图1 (网络版彩图)空间引力波探测器LISA、太极、天琴的设计灵敏度曲线及主要目标波源. 各探测器的灵敏度均为等臂长近似下第一代Michelson-AET通道的总灵敏度, 单位为特征应变, 其中虚线表示仅计入探测器设备噪声(参考LISA [11]、太极 [15]、天琴 [19]的设计噪声指标), 实线为加入银河系前景噪声后的结果

Figure 1 (Color online) The design sensitivity curves and primary target sources of the space-based gravitational wave detectors LISA, Taiji, and TianQin. The sensitivity of each detector is the total sensitivity of the first-generation Michelson-AET channel under the equal-arm-length approximation, expressed in terms of characteristic strain. The dashed curves represent the sensitivities resulting from instrumental noise (according to the noise requirements of LISA [11], Taiji [15], and TianQin [19]), while the solid curves represent the results after incorporating the galactic foreground noise.

础, 如LISA团队发布的数据处理挑战Mock LISA Data Challenge<sup>2)</sup>、LDC<sup>3)</sup>, 太极团队提出的空间引力波探测

仿真数据原型Taiji Data Challenge<sup>[45]4)</sup>, 天琴团队开发的适用于多个空间引力波探测任务的科学数据模拟程

2) <https://asd.gsfc.nasa.gov/archive/astrogravs/docs/mldc/>.

3) <https://lisa-ldc.lal.in2p3.fr/>.

4) <http://taiji-tdc.ictp-ap.org/>.

序包GWSpace<sup>[46]5)</sup>等. 这些仿真数据和程序旨在尽可能涵盖真实数据中包含的复杂性, 包括随时间变化的探测器响应形式、多波源混叠、数据异常等, 以便在真实探测数据可用之前, 开发出相应的软件工具, 并验证数据处理流水线的可行性.

贝叶斯统计推断为科学数据处理提供了强大的理论框架, 借助这一框架, 可以实现基于数据的参数后验概率估计和模型选择等. 贝叶斯框架的基础是贝叶斯定理:

$$p(\boldsymbol{\theta} | \mathbf{d}, \mathcal{M}) = \frac{p(\mathbf{d} | \boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\theta} | \mathcal{M})}{p(\mathbf{d} | \mathcal{M})}, \quad (1)$$

其中,  $\mathbf{d}$ 是观测数据,  $\mathcal{M}$ 是我们选择的理论模型.  $p(\boldsymbol{\theta} | \mathbf{d}, \mathcal{M})$ 是参数集 $\boldsymbol{\theta}$ 的后验分布, 它与数据的似然函数 $p(\mathbf{d} | \boldsymbol{\theta}, \mathcal{M})$ 以及参数的先验分布密度 $p(\boldsymbol{\theta} | \mathcal{M})$ 有关. 分母 $p(\mathbf{d} | \mathcal{M})$ 被称为贝叶斯证据, 是后验分布对参数空间边缘化的结果:

$$\begin{aligned} \mathcal{Z} \equiv p(\mathbf{d} | \mathcal{M}) &= \int_{\Theta} p(\boldsymbol{\theta}, \mathbf{d} | \mathcal{M}) d\boldsymbol{\theta} \\ &= \int_{\Theta} p(\mathbf{d} | \boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta} | \mathcal{M}) d\boldsymbol{\theta}, \end{aligned} \quad (2)$$

可作为贝叶斯模型选择的标准. 在引力波数据分析中, 似然函数 $p(\mathbf{d} | \boldsymbol{\theta}, \mathcal{M})$ 量化了在给定模型 $\mathcal{M}$ 和参数 $\boldsymbol{\theta}$ 的条件下得到数据 $\mathbf{d}$ 的概率. 在预期的全局拟合算法中, 似然函数可表示为包含噪声模型和不同波源模型的联合概率分布:

$$p(\mathbf{d} | \boldsymbol{\theta}, \mathcal{M}) \propto e^{-\frac{1}{2}(\mathbf{d} - \sum_{\mathcal{M}} \mathbf{h}(\boldsymbol{\theta}_{\text{GW}}))^T \mathbf{C}(\boldsymbol{\theta}_{\text{noise}})^{-1} (\mathbf{d} - \sum_{\mathcal{M}} \mathbf{h}(\boldsymbol{\theta}_{\text{GW}}))}, \quad (3)$$

其中,  $\mathbf{h}$ 描述的是不同波源类型的波形模板 $\mathcal{M}$ 的集合. 值得注意的是, 由于波源的数量和种类未知, 所以参数 $\boldsymbol{\theta}_{\text{GW}}$ 的维度本身也是待定的.

本文将贝叶斯框架中的关键要素为线索, 综述空间引力波探测数据分析中的核心挑战及其可能的应对策略, 特别是在引力波信号建模与计算、噪声与数据异常处理、似然函数构造与计算等方面的最新进展.

首先, 我们将介绍引力波信号的建模与计算(第2节), 包括波源模板的构建和探测器响应的计算, 这是理解和分析引力波数据的基础. 接着, 文章将探讨噪声及数据异常的影响和处理方法(第3节), 包括对Glitch、

数据间断和非稳态噪声的识别与处理, 这些因素都可能影响信号的搜索和参数估计的准确性. 然后, 将讨论全局拟合与参数反演的技术(第4节), 特别是似然函数的构造与计算, 以及如何利用搜索技术和全局拟合方法来提高参数估计的效率和准确度. 最后, 文章将重点介绍AI技术在引力波信号建模、噪声与数据异常处理、信号识别与参数估计等方面的前沿应用(第5节), 展示了AI技术如何为解决空间引力波探测数据分析中的复杂问题提供新的思路和工具. 通过这些讨论, 希望能为读者提供一个关于空间引力波探测数据分析挑战及其解决方案的概览, 特别是在利用最新科技进展方面的洞见.

## 2 信号建模与计算

在基于贝叶斯框架的科学数据处理中, 理论模型(即式(1)–(3)中的 $\mathcal{M}$ 项)是必要的输入条件. 对于数据驱动的AI方法, 为保证训练集与实测数据的一致性, 也需要精确理论模型的支撑. 引力波信号的理论模型通常称为波形模板. 本文将波形模板的构造分为波源模板建模和探测器响应建模两部分.

在波源模板方面, 空间引力波探测的目标波源与地面不同, 因而不能完全沿用相同的模板. 完备的空间引力波探测波源模板需要涵盖更广泛的波源参数范围和更复杂的波源运动特性. 并且, 以MBHB为代表的波源信噪比通常较高(可达 $O(10^3)$ 以上), 对模板的精度要求也相应提高.

对于探测器响应而言, 由于信号的可观测时间较短, 第二代地面引力波探测器的响应可简单表示为引力波的极化模式在固定探测器上的投影, 而空间引力波探测目标信号的可观测时间可达数月甚至数年, 与探测器轨道运动的时间尺度相近, 因此有必要考虑由探测器运动导致的调制效应. 此外, 空间长基线激光干涉测量的一大技术难点在于, 载波激光的频率不稳定性与臂长差耦合, 导致干涉数据中的激光频率噪声比预期的引力波信号强约6–8个量级<sup>[12]</sup>, 因此需要在数据处理中引入TDI环节, 即通过适当延迟和组合干涉测量数据, 构造虚拟的等臂长干涉, 从而有效地抑制激光频率噪声<sup>[47–53]</sup>. 轨道的调制效应和TDI的加入均提

5) <https://github.com/TianQinSYSU/GWSpace>.

高了探测响应建模和计算的复杂性. 科学数据处理要求波形模板兼具精确度、完备性和计算效率. 上述复杂性与需求的矛盾, 为空间引力波探测波形模板的构建提出了挑战. 本节将分别从波源模板和探测器响应两个角度, 对面临的挑战和可能的应对思路进行介绍.

## 2.1 波源模板

空间引力波探测的目标波源多为致密天体组成的双星系统, 其中MBHB, SBHB和UCB中的双星质量相近(质量比通常在100以内), 目前可用的模板包括后牛顿模板<sup>[54,55]</sup>、经数值相对论校正的有效单体模板SEOBNR系列(时域)<sup>[56-61]</sup>、由数值相对论波形拟合的唯象模板IMRPhenom系列(频域)<sup>[62-65]</sup>、数值相对论替代(Surrogate)模板<sup>[66-68]</sup>等. 加入了离心率、高阶模、进动等复杂特性的模板(如SEOBNRE<sup>[69]</sup>, SEOBNRPHM<sup>[70]</sup>, IMRPhenomXPHM<sup>[65]</sup>等)不仅能够更精细地刻画波源和引力波信号的演变, 也有助于打破参数之间的简并关系, 提高参数的估计精度<sup>[71-74]</sup>. 基于这些模板的研究对数据处理算法开发和科学目标论证均有指导意义, 有待进一步开展. 上述模板的构建多是以数值相对论波形为基准的, 因此数值相对论的计算精度对于模板精度具有根本性的影响. 文献<sup>[75]</sup>计算了不同解析度的数值相对论模板之间的匹配因子, 研究表明SXS波形库<sup>[76]</sup>在LISA、太极、天琴的灵敏度下均可达到与地面引力波探测相当的精度水平. 文献<sup>[77]</sup>指出, 为实现波源参数的无偏估计, 波形模板mismatch的上限应反比于信号信噪比的平方. 根据这一标准, 结合全局拟合和科学目标研究的具体需求, 可进一步细化各类波源模板的精度指标.

在致密双星波源中, EMRI是一类特殊的天体系统. EMRI中双星的质量比约为 $10^3-10^6$ , 波形复杂度极高, 预期会观测到 $10^4-10^5$ 个周期<sup>[78,79]</sup>. 目前已有的建模方法包括精度较高、计算相对复杂的求解Teukolsky方程法<sup>[80-84]</sup>, 以及计算效率较高的AK<sup>[85]</sup>, AAK<sup>[86]</sup>, NK<sup>[87]</sup>近似模板等. 由于EMRI信号的波形复杂, 可观测时间长, 空间引力波探测对其模板的精度和效率要求较高, 兼顾精度和效率的方法仍在探索之中<sup>[88,89]</sup>. 此外, 部分可探测的EMRI系统可能具备更加复杂的特征, 例如双星极端质量比旋近系统(Binary EMRI, b-EMRI), 即系统中的小天体是由两个恒星级

黑洞组成的双星<sup>[90]</sup>. 这类特殊波源的观测和分析对引力理论和强引力场中的非线性动力学过程具有重要的研究价值, 同时也为波源模板的构建提出了新的挑战.

到达探测器的引力波信号不仅与波源有关, 也与引力波途径的空间环境有关. 文献<sup>[91]</sup>预测, 部分引力波会受到传播路径中大质量天体透镜效应的影响, LISA将在5年的任务周期内探测到 $\leq 4$ 个具有多像的引力波信号. 多像对信号的总信噪比具有增益作用, 在建模中不考虑透镜效应可能导致遗漏潜在的可探测波源<sup>[92-94]</sup>. 同一波源的多个像会在不同时间、不同轨道位置到达探测器, 这一特点有助于实现波源的精确定位. 文献<sup>[95]</sup>研究了透镜化EMRI信号的可观测效应, 表明该类信号可用于同时限制EMRI波源参数和宇宙学参数. 因此, 考虑透镜效应的模板构建和科学数据分析工作应当受到重视.

## 2.2 探测器响应

空间引力波探测器与地面引力波探测器响应形式的区别主要体现在两方面: 轨道运动的调制和TDI通道的组合. 具体而言, LISA (或太极)的三艘航天器构成近似等边三角形编队, 编队的引导中心(即质心)沿地球轨道绕太阳运行, 并落后(或领先)地球约 $20^\circ$ , 轨道周期为一年. 同时, 航天器编队还将围绕引导中心旋转, 旋转的周期也为一年. 而天琴则以地心为引导中心, 绕引导中心旋转的周期约为 $3.65 \text{ d}$ <sup>[96]</sup>. 探测器方向的转动将导致天线响应函数随时间变化, 而平动则会在引力波观测数据中引入多普勒效应. 对于持续时间长、精度要求高的信号, 上述效应均不可忽略. 调制效应为响应的建模和计算增加了复杂性, 但同时也有助于在参数估计中缓解外禀参数之间的简并, 提升对波源的定位精度.

TDI是抑制激光频率噪声和时钟噪声的必要环节. 根据抑制激光频率噪声的能力, 常见的TDI方案可分为第一代和第二代, 而各代TDI又可分为不同的构型, 如Michelson, Sagnac, Relay, Beacon, Monitor等<sup>[97,98]</sup>. 各种TDI构型对信号和噪声的响应形式各异, 这既增加了科学数据处理的复杂性和多样性, 同时又为区分不同的数据成分提供了便利. 在科学数据处理的文献中最常见的是第一代Michelson  $\{X, Y, Z\}$ 通道, 以及由 $\{X, Y, Z\}$ 组成的噪声近似正交的 $\{A, E, T\}$ 通

道<sup>[99]</sup>. 在等臂长近似下, 采用 $\{A, E, T\}$ 通道有如下优势: 首先,  $\{A, E, T\}$ 是由基本TDI通道线性变换所得的具有最优总信噪比的组合<sup>[99,100]</sup>; 其次, 三个通道的噪声是相互独立的, 在计算似然函数和总信噪比时, 无需考虑通道之间的关联; 此外,  $A, E$ 与 $T$ 通道对信号和噪声的响应有明显的区别: 在低频,  $T$ 通道对信号不敏感, 因而可作为“空”(null)通道之一, 用于标定设备噪声、区分Glitch与引力波爆发信号等<sup>[101-104]</sup>, 而在高频,  $T$ 通道对总灵敏度的贡献与 $A, E$ 通道相当, 因此也可在科学数据处理中作为信号通道<sup>[52]</sup>(图2). 当前大多数科学数据处理的文献均以等臂长近似下的第一代Michelson构型为例, 然而, LISA和太极探测器的臂长变化速度可达数米每秒, 需要采用第二代TDI方案实现激光频率噪声的有效抑制. 因此, 有必要进一步研究在臂长不等、臂长可变以及第二代TDI等实际场景下, 不同TDI通道的表现. 例如, 文献<sup>[105]</sup>表明, 在不等臂的情况下, Michelson, Monitor, Beacon构型的 $T$ 通道在10 mHz以下的灵敏度显著高于等臂的情况; 文献<sup>[106]</sup>指出, 第二代Michelson构型在高频的特征频率较多, 在特征频率处的引力波响应被大幅抑制, 为缓解其造成的信号损失, 可选用特征频率较少的hybrid Relay通道.

对探测器响应的计算可在时域或频域进行. 部分早期的文献延续地面引力波探测的习惯, 将响应表示为臂长的相对变化, 即应变<sup>[107]</sup>; 而近年来的文献通常用激光在收发航天器间的相对频率变化来表示响应<sup>[108-110]</sup>:

$$y_{ij}(t) \equiv \frac{v_{\text{receive}} - v_{\text{send}}}{v_{\text{send}}} \approx \frac{1}{2(1 - \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}_{ij}(t_i))} \left[ H_{ij} \left( t - \frac{L_{ij}(t)}{c} - \frac{\hat{\mathbf{k}} \cdot \mathbf{x}_j(t)}{c} \right) - H_{ij} \left( t - \frac{\hat{\mathbf{k}} \cdot \mathbf{x}_i(t)}{c} \right) \right], \quad (4)$$

式中,  $i, j$ 分别为收发激光的航天器编号,  $L_{ij}(t)$ ,  $\hat{\mathbf{n}}_{ij}(t)$ 分别为激光在航天器 $i, j$ 间传播的光程和方向矢量,  $\mathbf{x}_{i,j}(t)$ 表示航天器 $i, j$ 的位置,  $\hat{\mathbf{k}} = \hat{\mathbf{k}}(\theta, \phi)$ 表示入射引力波的单位波矢量,  $H_{ij}(t)$ 是引力波在探测臂 $ij$ 上的投影, 下标 $ij \in \{12, 23, 31, 21, 32, 13\}$ (详见文献<sup>[110]</sup>)等.

计算出单臂响应后, 还需要将其组合成TDI通道<sup>[98,109]</sup>. 在时域中计算的挑战主要是计算量巨大, 例如, 对于采样频率为0.1 Hz, 总时长一年的信号, 采样点数为 $O(10^6)$ 量级. 在每个采样点处都需要计算波形和响应, 如果考虑不同的TDI组合方式, 则计算的时间复杂度将进一步增大. 文献<sup>[110]</sup>介绍了支持GPU加速的时域TDI响应计算程序FASTLISARESPONSE<sup>6)</sup>. 与基于CPU的计算相比, 该程序可利用GPU提速数百倍, 且支持用户自定义的波源模板及数值轨道, 对空间引力波探测时域数据分析具有广泛的参考和应用价值.

即便在GPU并行计算的加持下, 时域计算的效率仍难以达到大规模数据分析任务(如全局拟合)的要求. 而在频域, MBHB信号的振幅、相位和响应均可表示为平缓变化的光滑函数, 因而可以通过计算相对稀疏的数据点来表示出信号的特征, 从而加快计算效率. 文献<sup>[111]</sup>表明, 刻画MBHB的频域响应实质上仅需计算 $O(10^3)$ 个采样点. 文献<sup>[71, 112]</sup>给出了频域中的TDI响应形式, 可简单概括为

$$\tilde{h}^{A,E,T}(f) = \sum_{\alpha} \mathcal{T}_{\alpha}^{A,E,T}[f, t_{\alpha}(f)] \tilde{h}_{\alpha}(f), \quad (5)$$

其中,  $\alpha \in \{+, \times\}$ . 如果要考虑高阶模的贡献, 则 $\alpha = \ell m$ .  $t_{\alpha}(f)$ 描述了时间与引力波频率的关系, 可通过

$$t_{\alpha}(f) = -\frac{1}{2\pi} \frac{d\Psi_{\alpha}(f)}{df} \quad (6)$$

计算, 其中 $\Psi_{\alpha}(f)$ 表示 $\alpha$ 模式频域波形的相位.  $\mathcal{T}$ 对时间的依赖反映了探测器轨道运动对信号的调制效应, 如图2所示. 基于上述形式, 且支持GPU加速的频域TDI响应计算程序BBHX<sup>7)</sup>可实现MBHB波形和似然函数的快速批量计算, 每个信号的平均计算时间仅为数微秒<sup>[111]</sup>.

根据UCB等近单频波源的特征, 文献<sup>[113]</sup>提出了可快速计算频域响应的快-慢分解方法, 将与探测器轨道运动和引力波振幅相关的“慢”部分和与引力波相位相关的“快”部分分开考虑, 前者可通过较少的采样点计算, 而后者在频域近似为单一频点. 基于这一算法的FastGB程序被用于生成LDC Radler数据集的UCB信号. 此后, 又有研究者开发了加入三体运

6) <https://github.com/mikekatz04/lisa-on-gpu>.

7) <https://github.com/mikekatz04/BBHx>.

8) <https://github.com/mikekatz04/GBGPU>.

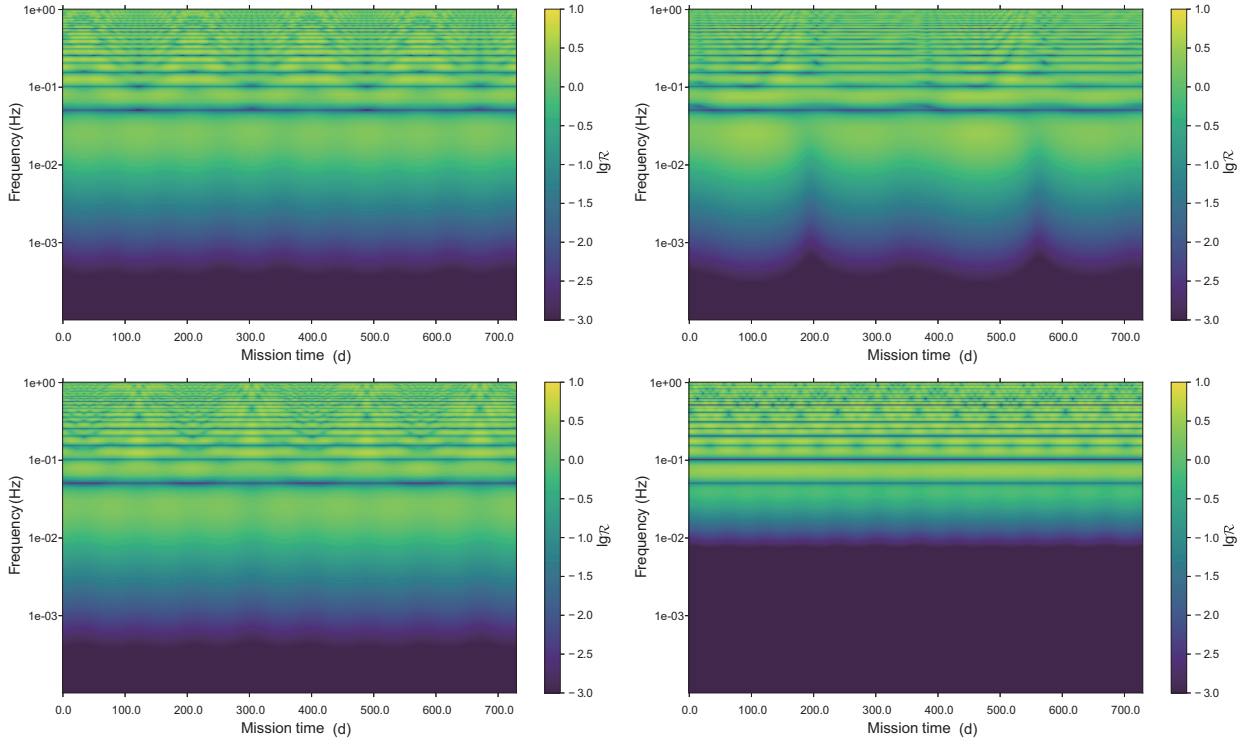


图2 (网络版彩图)等臂长近似下第一代Michelson-AET通道的响应函数随任务时间和频率的变化趋势. 上方两图分别展示了A通道对两个位于随机方位波源的响应, 左下图和右下图分别对应E通道和T通道的响应. 响应函数 $\mathcal{R}$ 定义为 $\mathcal{R}(t, f) \equiv \sqrt{[(\mathcal{T}_+)^2 + (\mathcal{T}_\times)^2]}/2$ , 对 $\mathcal{R}$ 的计算基于臂长为三百万公里的日心三星探测器

**Figure 2** (Color online) The variations of the response functions of the first-generation Michelson-AET channels under the equal-arm-length approximation, with respect to mission time and frequency. The top two panels show the responses of A channel to two randomly located sources. The left bottom panel and the right bottom panel correspond to the responses of E channel and T channel, respectively. The response function  $\mathcal{R}$  is defined as  $\mathcal{R}(t, f) \equiv \sqrt{[(\mathcal{T}_+)^2 + (\mathcal{T}_\times)^2]}/2$ , and the calculation is based on a heliocentric three-satellite detector with nominal arm-lengths of 3 million kilometers.

动[114]、离心率[115]等复杂波源运动特性的GPU加速版本GBGPU<sup>8</sup>).

上述频域算法在计算效率方面提升显著, 但现有的研究工作在推导频域响应时往往会采用理想的等臂长解析轨道, 不符合真实探测场景下的复杂轨道特征, 因而计算结果存在系统误差. 文献[110]研究了这一近似对参数估计的影响, 结果表明, 高信噪比、长时间信号受到的影响较大. 例如, 利用等臂解析轨道模型对信噪比为500的UCB信号进行参数估计, 其结果相对于真值的偏离可达 $2\sigma$ 水平以上(偏离程度与波源具体参数有关). 因此, 兼顾效率和精度的计算方法仍然是科学数据处理中的重要研究课题. 其他加速计算的方法[116–121], 如降阶积分方法(Reduced Quadrature Model)[122–124]、自适应频率分辨率方法(Adaptive Frequency Resolution)[124, 125]等, 均提供了可能的思路.

### 3 噪声与数据异常

目前空间引力波探测器LISA、太极、天琴均处于开发和地面试验阶段, 而地面与在轨的工作状态不同, 这意味着我们无法在发射前精确掌握探测器噪声的统计性质(对应式(3)中的C项), 需要借助在轨测量数据进一步标定[102–104, 126, 127]. 此外, 基于式(3)的引力波数据分析通常假设噪声是高斯、稳态的, 数据是连续的, 而在实际探测中, 噪声的非平稳性、非高斯性及各种可能的数据异常, 如环境或设备因素导致的Glitch[41, 128]、数据间断[129, 130]等, 都可能导致引力波事件的误警、漏警或参数估计偏差. 在本节中, 我们将介绍上述噪声及异常对空间引力波探测科学数据分析所造成的挑战及可能的应对思路.

#### 3.1 瞬态噪声事件

Glitch在时域数据中表现为脉冲或“毛刺”, 通常是

由设备或环境因素导致的。Glitch在地面引力波探测的数据中普遍存在,而在空间引力波探测器的技术验证卫星中,也发现了类似的现象。例如,在LISA探路者卫星引力参考传感器数据中发现了两类典型的Glitch,即快速上升-缓慢指数下降(Fast Rise and Slow Exponential Decay, FRED)型和正弦-高斯(Sine-Gaussian)型,每次Glitch事件的持续时间为数秒到数小时,发生的平均频率约为1天/次,导致Glitch的原因目前尚未确定,根据与其他数据(磁场、温度、压强等)的关联分析,测试质量周围环境气体的排出是最有可能的物理来源<sup>[41]</sup>。在GRACE的加速度数据中普遍存在振幅极高(可达 $10^{-5}$  m/s<sup>2</sup>量级)、持续时间很短(秒量级)的“twangs”信号,出现的频率约为600次/天<sup>[131]</sup>。GRACE的加速度计采用静电悬浮控制,同类技术将被用于空间引力波探测器引力参考传感器的非敏感轴向,twangs有可能通过轴间耦合进入到敏感轴向的数据中。此外,在GRACE Follow-On的星间激光干涉数据中还存在跳点等异常情况。基于设备和技术的继承性,类似的Glitch均有可能出现于空间引力波探测的科学测量数据。

Glitch对科学数据分析的危害在于,它可能会与引力波爆发信号混淆<sup>[132]</sup>;在建模中不考虑Glitch可能导致波源参数估计偏差<sup>[133]</sup>;在高斯稳态噪声的前提下,匹配滤波算法能够给出引力波探测的最优解,而Glitch导致的非高斯和非稳态特性使得这一前提不再成立,进而导致整体的探测阈值提高<sup>[134]</sup>等。

参考LISA探路者的经验<sup>[41]</sup>,科学测量数据与各类辅助数据的联合监测,可以为Glitch的判别和溯源提供线索。在数据处理层面,现有的研究工作对Glitch的处理思路主要是识别和扣除。文献<sup>[128]</sup>以LISA探路者的数据为例,采用含参数的shapelet函数作为模板,利用匹配滤波算法实现了自动化的Glitch搜寻和扣除。文献<sup>[132]</sup>推导了源于测试质量和干涉测量系统的Glitch在TDI通道中的传递形式,表明不同TDI通道的响应可以为区分Glitch和引力波爆发信号提供参考(例如, $T$ 通道中的低频信号被抑制,而Glitch则相对显著),并测试了贝叶斯模型选择算法在区分Glitch和引力波爆发信号方面的表现。文献<sup>[133]</sup>展示了同时估计Glitch模型参数与MBHB波源参数的效果,在讨论的模型范围内实现了MBHB和Glitch的精确重构。在文献<sup>[44]</sup>提出的全局拟合原型中,也展望了基于跨维度参数估计

方法<sup>[34,135,136]</sup>的Glitch处理策略,即在数据中搜索引力波信号的同时,可以同时估计维度未知的Glitch参数。针对Glitch导致的非高斯性,可以考虑用student- $t$ 分布<sup>[137,138]</sup>、广义双曲分布<sup>[139]</sup>、高阶Edge worth展开<sup>[140]</sup>等方式为似然函数建模。

### 3.2 数据间断

数据间断包括计划性间断和不可预期间断。其中,前者由天线调整<sup>[11]</sup>、频率计划切换、设备维护等原因导致;而后者则源于未预期的在轨异常事件。此外,在数据处理中,有时会为规避Glitch等异常数据的影响而删除其所在的数据段,因此也会人为地引入间断<sup>[130]</sup>。

数据间断同样会对空间引力波探测的科学数据处理造成困难。首先,间断会导致信息丢失,信号的信噪比积累时间不足;其次,当时域数据含有间断时,在频域会发生谱泄露,致使参数估计的偏差或误差增大<sup>[130,141]</sup>;文献<sup>[129]</sup>通过计算Fisher信息矩阵表明,不可预期间断的持续时间可达数天,对参数估计的影响较大,而可预期的间断的持续时间较短,仅在MBHB并合发生在间断内的情况下会造成可观影响。此外,间断本身也会导致噪声呈现出非稳态特性,详见第3.3节。

针对数据间断的处理,文献<sup>[141,142]</sup>的方法是在间断边缘设置过渡平滑的窗函数,从而减少频谱泄露。文献<sup>[130]</sup>提出了基于贝叶斯框架的数据扩充(Bayesian Data Augmentation)技术,即以窗函数的形式对间断建模,并通过迭代执行波源和间断的参数估计,最终获得稳定的波源参数后验分布。为实现含间断的UCB信号分析,文献<sup>[143]</sup>根据UCB信号在频域的稀疏性,提出了非参数化的间断修复算法。

### 3.3 非稳态噪声

非稳态噪声是指噪声的统计特性随时间变化。噪声的非平稳性有以下来源:首先,银河系内的物质主要分布于星系盘中,而太阳系位于银河系的猎户臂上,因此在太阳系内的观者看来,UCB波源的分布具有各向异性。同时LISA、太极等空间引力波探测器的轨道运动以年为周期,因此不可解析的UCB前景也会以年为周期变化,形成非稳态的噪声源。具体而言,当探测器平面的法线指向银河系中心时,前景噪声的强度较

高<sup>[101]</sup>. 其次, 在长达数年的探测任务期间, 探测器的载荷和卫星平台的工作状态也会因自身或环境因素发生漂移, 导致设备噪声的非平稳性. 以LISA探路者卫星的情况为例, 由于测试质量所在电极笼内的真空条件不断变化, 加速度噪声的水平持续降低<sup>[144, 145]</sup>; 同时, 姿态-光程耦合噪声(即航天器或光学平台姿态抖动导致的测距噪声)与光学器件的排布密切相关, 也会随着工作环境发生变化<sup>[144, 146, 147]</sup>. 最后, Glitch和数据间断也会为数据引入非平稳特征.

噪声的性质随时间变化, 就意味着需要定期开展噪声的测量和标定. 使用更新不及时的噪声PSD会影响低信噪比波源的探测, 或导致参数估计的结果出现偏差. 同时, 非稳态噪声对科学数据分析的影响是基础性的. 引力波的科学数据分析通常在频域进行, 噪声平稳的前提下, 不同频率处的噪声在统计上相互独立, 因而匹配滤波的计算大为简化, 噪声仅以PSD的形式出现在分母, 即<sup>[148]</sup>

$$(d|h) \equiv 4\text{Re} \int_{f_{\min}}^{f_{\max}} \frac{\tilde{d}(f)\tilde{h}^*(f)}{S(f)} df, \quad (7)$$

式中,  $(d|h)$ 表示观测数据 $d$ 与波形模板 $h$ 构成的匹配滤波器,  $S(f)$ 为噪声的单边PSD. 非稳态会导致不同频率的噪声出现统计关联, 上述优势将不复存在, 计算的复杂度大幅增加<sup>[130, 149]</sup>.

对稳定性的判别是处理非稳态噪声的前提. 文献<sup>[150]</sup>介绍了VOCAL Test, SOMED Test等判别时间序列统计稳定性的方法, 并在LISA探路者的数据上进行了测试. 时频域(如小波域)分析为刻画和处理非稳态噪声提供了便利<sup>[149]</sup>. 如果噪声的统计特性随时间缓慢变化, 则其在短时间内近似保持稳态, 相邻的时频点仍可视作统计独立, 从而简化匹配滤波的计算. 同时, 小波函数也为表征Glitch及引力波爆发信号提供了天然的基底. 为发展时频域数据分析, 在同一表示空间构造波形模板也是必要的. 文献<sup>[151]</sup>针对非稳态前景噪声对随机引力波背景的干扰, 提出了将数据按周分段加权, 并通过对权重寻优抑制前景噪声的方法, 有效提高了对随机引力波背景的探测能力. 已成功应用于地面引力波探测的非稳态噪声处理方法也可供借鉴, 如文献<sup>[152–154]</sup>等.

## 4 全局拟合与参数反演

### 4.1 似然函数

在空间引力波探测的科学数据处理中, 优化似然函数是一种重要的解决思路. 似然函数用于评估给定模型参数下观测数据的概率(对应式(3)), 是贝叶斯统计推断的关键组成部分. 不同的似然函数在计算复杂度和精确度方面各有优势和局限, 最常见的似然函数假设有两个:

- Whittle似然函数<sup>[155–157]</sup>: 这种似然函数假定噪声是高斯且稳态的, 通过将每个探测器独立构建的似然函数相乘, 构建成一个全面的似然函数. 它已经被广泛应用于引力波数据分析中<sup>[158, 159]</sup>. 这种方法的优点是更精确, 但计算代价较高(详细可查阅文献<sup>[160, 161]</sup>), 而且最重要的是对于高自相关性的时间序列, 该方法可能会诱导偏差进入到似然函数表达式中, 关于Whittle似然函数的缺点已在诸多文献中被广泛的讨论<sup>[138, 157, 162]</sup>.

- Heterodyned似然函数<sup>[163–166]</sup>: 这种似然函数通过假设在两个高似然点评估的模型相似, 因此比率是一个平滑变化的函数, 能够在很短的时间内评估周围点的似然. 它在计算效率上较Whittle似然函数有显著提高, 可以大幅减少贝叶斯推断的计算成本<sup>[164]</sup>, 且所需内存比完整的Whittle似然分析少<sup>[167]</sup>. 但Heterodyned似然函数的一个缺点是必须事先知道一个高似然的基准点, 而Whittle似然是完全通用的. 为了解决这个问题, 要在进行贝叶斯推断之前, 对数据进行基于模板的匹配滤波搜索, 并使用最佳拟合模板作为高似然点<sup>[168]</sup>. 另外, Heterodyned似然函数也可能不适用于包含高阶模的波形<sup>[169]</sup>. 因此, 当无法找到高似然基准点, 或者需要包括高阶模时, 完整的Whittle似然可能更为合适.

为了拓展噪声模型的理想假设并提高数据异常值的鲁棒性, 研究者们会考虑Student- $t$ 或Hyperbolic分布作为似然函数的概率模型(即对式(3)的形式建模)以实现噪声的建模. 这种方法特别考虑到数据中可能存在的非高斯噪声(详见第3.1节), 从而能够更有效地处理异常值和噪声, 尤其适用于引力波数据分析. 在最近, 有研究工作<sup>[170]</sup>通过数据下采样和近似似然函数的构建, 实现了快速的参数估计, 有效降低了计算成

本,提高了数据处理效率,为处理大规模数据集和复杂信号分析提供了可行性. 每种似然函数都有其特定的应用场景和优缺点,选择合适的似然函数对于有效和准确地分析引力波数据至关重要. 在实际应用中,研究者需要根据具体的数据特性和分析目标,选择最适合的似然函数.

## 4.2 搜索技术

将似然函数与对参数的先验知识结合,即可利用式(1)得到给定观测数据和理论模型下的参数后验分布. 较早的研究工作多采用Fisher信息矩阵预测波源参数的估计精度<sup>[171-176]</sup>. 应用Fisher信息矩阵的前提是参数的后验分布可近似为多元高斯分布,原则上仅适用于高信噪比的情况. 这种方法通常仅限于理论研究,无法单独用于实际的引力波数据处理.

本文着重探讨能够从实测数据中提取出物理信息的数据分析方法. 当前,空间引力波探测信号搜索流水线开发的重点是天体物理波源,特别是最明亮的MBHB和数量最多的UCB,并且已经取得了一系列进展.

针对MBHB信号, PYCBC-INFERENCE<sup>[168]</sup>和Bilby in Space<sup>[167]</sup>分别以地面引力波探测中广泛使用的PyCBC和Bilby框架为基础,针对空间引力波探测的特点进行了改造,利用马尔科夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)或嵌套采样方法实现了MBHB的参数估计. Cornish等人<sup>[166]</sup>以及Katz等人<sup>[177]</sup>的工作则基于PTMCMC方法,其中特别强调了GPU加速在提高参数估计效率方面的巨大优势<sup>[177]</sup>.

对于UCB信号,可解析的波源可达数万个,波源参数的个数为 $O(10^4) - O(10^5)$ 量级. 为缓解参数空间维度过高的问题,文献<sup>[36, 178, 179]</sup>的策略是不考虑参数的后验分布,仅专注于通过优化算法寻找与数据最匹配的模拟信号,从而实现参数的最大似然估计. Strub等人<sup>[37, 180]</sup>使用高斯过程回归方法构建似然函数,并利用GPU并行加速计算,他们的方法不仅能够得出波源参数的最大似然估计,还能得到估计结果的不确定性描述. 由于MCMC方法可以给出参数后验分布的详细信息,因此应用更为广泛,如使用BAM (Blocked Annealed Metropolis-Hastings)技术的研究工作<sup>[32, 181, 182]</sup>,以及基于RJMCMC技术的研究工作<sup>[34]</sup>等.

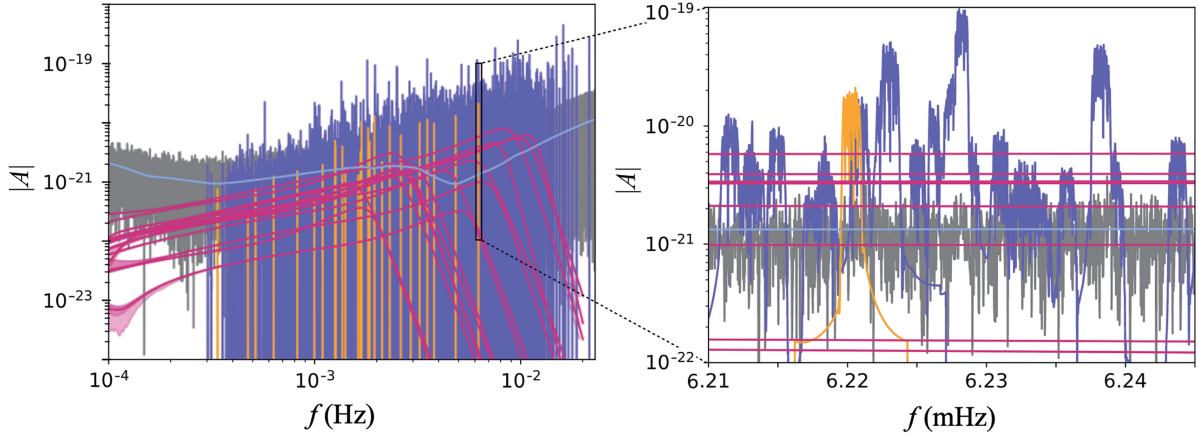
此外, BALROG<sup>[183, 184]</sup>是可用于多种目标波源(如UCB, SBHB, MBHB等)参数估计的程序包,目前仍处于开发阶段. 还有研究者通过融合多种MCMC搜索技术(如PTMCMC和RJMCMC等)开发了新的随机采样程序包Eryn<sup>[136]</sup>,展现出了极大的潜力.

尽管上述工作在波源搜索方面取得了显著进展,但这些研究普遍采用了将噪声视为已知先验信息的做法,即通过简单地使用PSD设计灵敏度来模拟噪声,而没有对数据中的噪声进行经验性的建模拟合. 在研究MBHB时,这些工作通常将UCB前景信号视为噪声的一部分,并在PSD的生成中包含这些信号. 由于这些信号数量众多,振幅和相位各不相同,根据中心极限定理,尽管并非平稳,但它们的累积效可被认为是高斯分布的. 这种方法虽然在处理MBHB信号时不会造成显著的差异,但对于可观测时间较长的信号,这种假设就显得过于简化. 实际上,当考虑UCB引起的混淆噪声时,研究发现,MBHB的内秉参数估计可能会偏离真实值<sup>[168]</sup>. 尽管注入值通常落在90%置信区间内,但与没有混淆噪声存在的情况相比,后验概率分布的峰值已经发生了偏移. 这些局限性激发了对一种全局拟合流水线的需求,该流水线能够识别并推断大部分UCB的参数属性,并在对MBHB进行推断之前从数据中移除这些信号. 此外,这种方法还应该能够处理PSD的时间依赖性,这将在下一节中详细讨论.

## 4.3 全局拟合

全局拟合方法<sup>[34, 43, 44]</sup>的核心思想在于对空间引力波数据中存在的所有天体物理和仪器特征同时进行综合建模. 这种方法不仅仅关注单一波源的信号,而是尝试捕捉数据中所有波源的综合影响,对整个数据集进行全面分析,以识别和建模所有潜在的信号和噪声源,如图3所示. 通过这种方式,可以更准确地理解数据中的每个组成部分,从而提高整体分析的准确性和可靠性.

在实际操作中,全局拟合的过程首先要重构并扣除数据中最为显著的波源信号,随后利用剩余的残差数据进行深入分析<sup>[34, 44]</sup>. 这种策略有效地减少了错误累积和偏差,尤其是在处理众多混叠信号时尤为重要. 全局拟合使得我们能够更加高效地处理和分析不同的信号源,通过确保信号与噪声参数的联合更新,避免了



**图 3** (网络版彩图) 基于LDC Sangria数据集的全局拟合分析结果<sup>[11]</sup>, 以TDI-A通道的幅度谱密度表示, 单位为 $\text{Hz}^{-1/2}$ . 图中灰色为LDC Sangria数据集的原始数据, 浅蓝色为分析中采用的噪声模型, 其余颜色表示根据波源参数估计中心值重构的引力波信号, 包括: UCB (紫色)、VB (橙色)、MBHB (品红色)

**Figure 3** (Color online) The results of global fitting analysis based on the LDC Sangria dataset [11], expressed in terms of the amplitude spectral density of TDI-A channel in unit of  $\text{Hz}^{-1/2}$ . The original LDC Sangria dataset is shown in gray, while the noise model employed in the analysis is depicted in light blue. The other curves represent the reconstructed gravitational wave signals based on the estimated central values of the source parameters, including UCB (purple), VB (orange), and MBHB (magenta).

偏差和源混淆对分析结果的潜在负面影响. 该方法旨在同时探测和描述数据中存在的所有波源, 并定义一种计算效率高的策略, 以便在接收到新数据时迅速更新全局拟合结果.

现有的全局拟合流水线原型是基于MCMC随机采样算法实现的<sup>[34,43,44]</sup>. MCMC算法通过构建一个马尔可夫链序列, 其元素 $\theta(t_i)$ ,  $i = 0, 1, \dots$ 是(渐进地)从目标后验分布 $f(\theta)$ 中独立抽样的. 只要算法满足细致平衡条件(Detailed Balance Sheet Condition):

$$f(\theta)p(\theta \rightarrow \theta') = f(\theta')p(\theta' \rightarrow \theta), \quad (8)$$

则链中的样本分布将会收敛到目标后验分布. 式中,  $p(\theta \rightarrow \theta')$ 是马尔可夫链从点 $\theta$ 移动到点 $\theta'$ 的概率. 最广泛使用的MCMC算法是Metropolis-Hastings<sup>[185,186]</sup>. 算法的第一步是定义一个初始状态 $\theta(t_0)$ . 然后, 在每个后续步骤 $i$ , 通过从给定的提议分布(Proposal Distribution)  $q(\theta' | \theta(t_i))$ 中随机抽取来提出一个新状态. 新提出的状态然后以一定的概率被接受. 如果移动被接受, 设置 $\theta(t_{i+1}) = \theta'$ , 否则设置 $\theta(t_{i+1}) = \theta(t_i)$ . 任何合理的提议分布的选择都将生成一个正确且稳定分布的马尔可夫链. 然而,  $q$ 的选择好坏对于其效率至关重要, 即在合理的计算时间内实现MCMC链的收敛. 此外, 另一方面的技术难题需要克服的是提高接受率, 这对

于收敛更至关重要, 有时候仅仅改进提议分布还不足以有效采样参数空间. 为了解决这些问题, 已经提出了各种MCMC增强方法, 如哈密顿蒙特卡罗<sup>[187,188]</sup>的No-U-Turn抽样<sup>[189]</sup>、吉布斯抽样<sup>[190-192]</sup>等.

目前, 已经有一些研究工作开始涉及到对高维度(High-Dimensional)、多模式(Multi-Modal)、可变维度(Trans-Dimensional)的强烈需求(可对应于式(3)中的 $\theta_{\text{GW}}$ ), 尤其是UCB的搜索技术研究(可见第4.2节). 此外, 文献[44]是首次尝试对LDC数据里的仪器噪声、UCB、VB和MBHB全局拟合的研究工作. 在上述全局拟合的解决思路中, 运用的贝叶斯统计推断方法各具特点.

- **Blocked Metropolis Hastings**<sup>[193]</sup>: 这种方法适用于处理高维数据. 其优势在于能够实现模块化, 即将复杂的问题分解为多个子问题, 使得处理更为高效. 它的模块化设计也有助于简化实现过程, 特别是在涉及复杂模型和大数据集的情况下. 文献[44]利用该算法分层地从不同源类型到单独源进行模块化组合, 由此实现某模块的参数更新时, 确保其他模块的参数保持不动.

- **并行退火(Parallel Tempering)**<sup>[194-196]</sup>: 并行退火的概念是为了在具有多峰模式特征的似然函数上有效地采样而引入的. 这个想法基于将后验密度转换

为具有不同温度的密度,  $T$ , 定义为  $p_T(\theta | y) \propto p(y | \theta)^{1/T} p(\theta)$ . 当  $T = 1$  对应的目标后验分布, 当  $T = \infty$ , 它是先验概率分布. 中间温度通过减少高低可能性区域之间的对比度来“平滑”后验. 在并行退火计算过程中, 一组马尔可夫链可并行构建, 每一链都在为不同的温度  $T$  采样转换后的后验, 这些链周期性地交换信息. 该算法的核心思想是最热的链在参数空间中更广泛地探索, 并且把它们遇到的高似然区域的信息传播到较冷的链, 这个信息是通过提议在不同链之间实施状态交换来进行传递的. 在文献[44]中, 该算法的调温方案是不同模块单独开发和调优, 只有零温度链参数或状态在不同进程间共享, 每个并行调温采样器都是多线程的, 理想情况下每个链使用单个CPU. 这种动态改进蒙特卡罗方法的模拟算法是通过并行退火改善链的综合效果, 尤其对于跨维度采样器的模型转换至关重要.

- 可逆跳跃(Reversible Jump)<sup>[135]</sup>: 这是一种适用于可变维度统计模型的贝叶斯推断方法. 它允许在不同复杂度的模型之间进行转换, 动态的进行模型选择, 从而提高分析的灵活性和适应性. 该算法在技术实现上的挑战主要是两个方面: 一是在提议不同维度模型之间的移动时需要满足的维度匹配(Dimension Matching)要求; 另一个是从算法效率的角度而言, 选择允许在动态参数空间上平滑过渡的提议分布也至关重要, 这在许多情况下需要大量的努力. 由于这些技术原因, 所有这些软件工具均是为了解决它们各自特定的目标问题而专门开发的<sup>[34, 44, 136]</sup>.

在全球拟合技术的应用中, 面临着以下几个主要的局限性和挑战(总结自文献[44]):

(1) 资源优化挑战: 这涉及到如何在最小化等待时间和最大化计算节点效率的过程中, 进行资源分配和优化策略. 在处理大规模、高维度的数据时, 这是一个复杂的平衡问题.

(2) 迭代成本平衡难题: 在不同计算步骤(如对MBHB和GBMCMC<sup>[34]</sup>内部的跨维度模型更新)之间平衡成本是一个挑战. 需要考虑到整体收敛时间和计算资源, 以优化整个分析过程.

(3) 具体的技术限制:

- Heterodyne步骤限制: 在MBHB模型处理中, Heterodyne步骤成为了一个瓶颈, 其在每次迭代中的成本对整体计算成本有重要影响.

- 收敛时间过长: 对于频率最高的UCB, 其收敛时

间较长, 这增加了迭代次数和整体分析的时间和资源消耗.

这些局限性表明, 空间引力波探测科学数据处理的挑战并不仅仅局限于软件算法层面. 在资源分配、处理高成本步骤和管理长时间收敛过程等方面, 软硬件结合的当前方法, 包括Littenberg和Cornish<sup>[44]</sup>提出的GLASS工具, 仍有改进空间. 为了提高效率和准确性, 不仅可能需要进一步的测试验证或采用更高的计算成本, 也可能需要充分利用软硬件结合的优势, 如CPU/GPU异步计算架构等. 需要强调的是, 要实现空间引力波观测频段内所有波源参数和噪声模型的全局拟合和可靠分析, 还有很多工作要做. GLASS所提供的工具仅是全局拟合解决方案的初步探索, 要达到“生产就绪”的状态, 还需要进行多轮迭代和完善.

除了MCMC随机采样方法之外, 在地面引力波数据处理领域, 研究者们也提出了一些其他的替代方案<sup>[121, 197-199]</sup>, 这些方法同样值得我们关注. 特别是随着人工智能技术的兴起, 其在引力波数据分析中的应用展现出了巨大的潜力和优势<sup>[200-204]</sup>. 人工智能技术不仅能够加速数据处理流程, 提高分析效率, 还能在某些情况下提供更为精确的参数估计和信号识别. 这些替代方案和AI技术的出现, 为地面引力波数据处理提供了新的视角和工具, 同时也为未来空间引力波科学数据处理的研究开辟了新的道路.

## 5 人工智能技术的应用

在空间引力波探测领域, 人工智能技术, 尤其是深度学习和GPU加速计算, 有望成为数据处理的关键工具<sup>[205]</sup>. 目前, 这些技术的应用范围从引力波信号建模、噪声与数据异常处理, 到信号识别与参数估计, 展现了其在精度、效率和泛化性方面的独特优势.

### 5.1 引力波信号建模

在引力波信号建模方面, 人工智能技术的应用主要集中在EMRI的快速生成和精确建模上, 是最早实现人工智能技术应用的案例. EMRI是空间引力波探测的重要信号源, 其中复杂的相对论效应导致波形产生高达数万个周期的变化, 波形复杂度极高, 预期会观测到 $10^4-10^5$ 周期. Chua等人<sup>[206]</sup>提出了使用人工神经网络实现了EMRI信号的波源参数与降

阶(Reduced-Order)域之间的映射关系, 由此简化复杂波形的建模流程, 显著提高了引力波推断的效率和精度, 开辟了快速生成完全相对论性EMRI波形模板的新途径. 随后他们进一步开发的AAK模型<sup>[207]</sup>程序包FastEMRIWaveforms<sup>[208]</sup>考虑了完整相对论效应可以实现高精度的快速EMRI波形生成, 现已成为非常流行的EMRI波形生成程序, 为LISA等空间引力波探测任务的数据分析提供了强有力的支持.

人工智能技术在引力波信号建模中的应用展示了巨大的潜力. 其核心优势在于利用神经网络模型替代传统的、计算量大且耗时的波形模板生成过程中的部分计算环节. 这种数据驱动的方法不仅大幅提高了计算效率, 还能在不受模型不确定性过大影响的前提下, 保持波形生成的高准确度. 此外, 类似的研究方法也已经被应用于恒星级双黑洞并合的模板生成中<sup>[209]</sup>. 与此同时, 一些研究者尝试采用端到端的方法, 直接从致密双星系统的历史波形数据生成未来的波形<sup>[210-212]</sup>, 或者直接从波源参数生成引力波波形模板<sup>[213]</sup>. 这些创新方法在提升波形生成的速度和准确性方面表现出极大的潜力. 然而, 由于人工智能模型本身的可解释性较差, 这些方法在引力波信号建模的质量可控性方面仍面临挑战, 需要未来的研究继续探索和完善.

## 5.2 噪声与数据异常

在空间引力波探测中, 对噪声和数据异常的处理是一项至关重要的任务. 目前, 人工智能技术在区分数据异常Glitch和其他引力波波源的问题上已显示出显著的应用潜力<sup>[214,215]</sup>. 特别是在文献<sup>[215]</sup>中, 研究者针对LISA数据中可能出现的多种Glitch类型, 提出了一种基于混合神经网络的数据处理流水线. 这种方法不仅能检测并扣除多种Glitch, 还能显著提高TDI数据的质量. 该研究展示了人工智能技术在模式识别和建模Glitch异常特征方面的强大能力. 此外, 其他研究者从信号建模的角度出发, 进行了进一步的探索<sup>[216]</sup>. 他们提出的深度学习模型能够在包含各种非稳态噪声的环境中准确提取引力波信号, 展现出对Glitch、数据间断和非稳态噪声的高度鲁棒性和适应性. 这些研究表明, 基于人工智能技术的噪声和异常处理方法不仅依赖较少的模型假设, 而且具有强大的泛化性能, 特别适用于复杂的空间探测环境中. 这些进展为未来的空间

引力波探测任务提供了重要的技术支持, 有望极大地提高数据处理的效率和准确性.

## 5.3 信号识别与参数估计

在信号识别与参数估计方面, 大多数人工智能技术应用的研究工作都集中在特定波源信号的识别和参数反演上. 与传统方法相比, 基于人工智能的算法将计算的时间成本从“在线”的数据分析转移到了“离线”的模型训练, 尤其适用于要求低延迟和实时性的场景, 如多信使观测预警等.

利用神经网络实现波源参数估计的一般原理为: 设 $p$ 为目标后验分布,  $q$ 为用神经网络构造的待估计后验分布, 则训练的目标应设定为 $p$ 与 $q$ 的差距达到最小. 用Kullback-Leibler散度衡量这一差距, 可表示为

$$D_{\text{KL}}(p|q) = \int d\mathbf{d} \int d\mathbf{p} p(\boldsymbol{\theta}|\mathbf{d}) \ln \left[ \frac{p(\boldsymbol{\theta}|\mathbf{d})}{q(\boldsymbol{\theta}|\mathbf{d})} \right] \\ \sim -\mathbb{E}_{\boldsymbol{\theta} \sim p(\boldsymbol{\theta})} \mathbb{E}_{\mathbf{d} \sim p(\mathbf{d}|\boldsymbol{\theta})} [p(\boldsymbol{\theta}|\mathbf{d})]. \quad (9)$$

换言之, 虽然对神经网络的训练目标是逼近后验分布, 但在计算损失函数时并不需要计算似然函数, 只需根据参数的先验分布对参数随机采样, 然后根据波形模板和噪声模型随机生成数据样本, 最后代入待估计后验分布 $q(\boldsymbol{\theta}|\mathbf{d})$ 中并对所有样本求均值. 这种损失函数的构建方法展示了一种不依赖于似然函数(Likelihood-Free)的思路, 原则上, 即便面对似然函数形式复杂、计算困难的情况, 这一方法仍可适用. 式中 $q$ 可采用不同的形式, 如多维高斯分布、标准化流等.

在EMRIs的探测和参数估计中, 人工智能技术的应用展现了显著的进步. 近期的研究工作<sup>[217-219]</sup>都集中在利用卷积神经网络模型来识别和反演EMRI波形信号. 这些研究覆盖了不同的引力波项目(如天琴<sup>[217]</sup>、LISA<sup>[218]</sup>和太极<sup>[219]</sup>), 采用了不同代际的TDI数据(TDI-1.0<sup>[217]</sup>, TDI-1.5<sup>[218]</sup>, TDI-2.0<sup>[219]</sup>), 并测试了多种波形族(如AK<sup>[217]</sup>, AAK<sup>[218,219]</sup>和EOB<sup>[219]</sup>). 这些工作共同指出, 通过人工智能模型, 特别是卷积神经网络, 可以在极短的时间(约10 ms)内有效地探测EMRI信号, 显著减少了传统匹配滤波方法对大量高精度波形模板的需求. 此外, 恽倩芸等人<sup>[220]</sup>的研究通过引入Unet和VGG模型, 以及采用Q变换的频谱图作为输入特征, 进一步提高了信号识别和参数估计的准确性

和效率,显著降低了后续贝叶斯参数估计的参数空间和计算成本。

尽管人工智能技术在效率上远超传统的匹配滤波等方法,但目前的研究在训练网络时使用的训练集相对于EMRI的高维参数空间来说还不够充分,且某些参数的范围设定较窄,限制了模型的普适性和泛化能力。此外,现有研究主要处理信噪比较高的情况,对于信噪比较低的一般情况尚未进行深入探索(关于EMRI数据分析方法的详尽综述可参见文献[89])。尽管存在这些限制,这些研究为EMRI波形的快速准确识别提供了新的技术途径,并对未来空间引力波探测任务的数据处理和分析具有重要的指导意义。

对于MBHB并合信号,其作为空间引力波探测中最强的瞬态信号源,预计的质量范围在 $10^4$ – $10^7$ 太阳质量之间。这些事件不仅可以观测到双星系统的晚期旋近、并合,还能捕捉到铃宕阶段,预计每年的观测事件率在几个到几百个。在此背景下,深度学习技术的应用为MBHB信号的提取和参数反演提供了新的视角。阮文洪等人<sup>[221]</sup>的研究展示了深度学习方法在LISA数据上寻找MBHB信号的能力。通过构建匹配滤波卷积神经网络<sup>[222]</sup>,他们的模型能够在几秒钟内分析一年的LDC数据,准确识别所有MBHB系统并合的引力波信号,且没有任何误报。此方法的成功应用,为空间引力波探测提供了一种新的路径,特别是在快速搜索和计数MBHB系统方面,甚至包括对MBHB并合信号的提前预警<sup>[223]</sup>,为未来多信使观测的早期预警提供了可能。在MBHB波源参数反演的问题上,最早开创性的工作是Chua等人<sup>[224]</sup>通过神经网络学习贝叶斯后验,展示了深度学习在引力波推断中的应用,其算法在采样效率上比传统方法高出3个数量级,为混叠MBHB信号的全波源参数反演提供了新思路。近期的研究工作<sup>[225–227]</sup>则通过归一化流技术,探索了更多的参数维度推断,探测器投影对称性和额外多模式的泛化性质,以及对Glitch的鲁棒性,进一步推进了基于空间引力波探测科学数据的引力波信号快速搜索和参数估计,为未来引力波源的快速识别和参数精确测定开辟了新途径。

随着深度学习技术在空间引力波数据处理领域的日益成熟,已经出现了几种新颖的应用方法论。例如,有研究<sup>[228]</sup>提出通过单一模型同时处理来自多种波源的信号。这种统一框架能够识别和分析包括MBHB,

EMRI,UCB以及随机引力波背景等不同类型的引力波信号,无需为每种波源类型设计和训练独立的模型。这不仅显著提高了数据处理的效率,也增强了模型的泛化能力,使其能够更加灵活地适应新的或未知类型的引力波信号。

除了这种端到端的建模思想外,另一种研究方向是将深度学习技术融入到基于MCMC的参数反演中<sup>[229–231]</sup>。研究者们尝试使用归一化流模型作为一个自适应的提议分布函数,通过拟合全部或部分待推断参数的马尔科夫链的后验分布函数来指导蒙特卡罗采样的方向。这种方法不仅可以实现更快的收敛速度和提高接受率,还能在减轻全局拟合过程中的巨大计算负担的同时,确保在贝叶斯原理下实现空间引力波探测的全局拟合和参数反演。这些进展不仅展示了人工智能技术在空间引力波探测中的应用潜力,也为未来的引力波研究提供了新的技术路径,有望在未来的探测任务中实现更高效和精确的数据分析和信号处理。

空间引力波探测在科学数据处理方面面临的主要挑战之一是所谓的“维度灾难”。这一问题源于引力波数据本身及其源参数的高维度特性。这使得传统的数据分析方法在处理这些数据时效率低下,难以捕捉到所有潜在的信息。然而,以数据驱动为核心的人工智能方法,尤其是深度学习技术,提供了一种有效的解决方案。通过将高维复杂的数据分布映射到低维的简单特征空间——深度学习中的隐空间(Latent Space)——这些技术能够实现原始数据向更适合分析或建模格式的转换。这种转换不仅有助于数据挖掘和特征提取,还能显著提高数据处理的效率和准确性。此外,人工智能技术的另一个显著优势在于其对GPU等高性能硬件的充分利用。这种硬件加速能够有效应对大数据量和高维度数据处理的挑战,为引力波数据分析提供了强大的计算支持。尽管人工智能在空间引力波探测的科学数据处理领域还处于初步探索阶段,但其在全局拟合技术等关键难题上的应用前景已经显现。通过人工智能技术的辅助,我们有望缓解甚至解决这些难题,为未来的空间引力波探测项目提供强有力的技术支持,开启引力波天文学新篇章。

## 6 总结与展望

本综述对空间引力波探测数据分析所面临的独特

挑战进行了简要介绍, 并列出了近年来的研究工作中提出的方法和思路。

值得注意的是, 空间引力波探测数据的突出特点在于数量巨大的引力波信号、数据间断、Glitch、非稳态噪声等共存于数据中, 并且在时域和频域高度交叠。因此有必要综合考虑上述问题, 并将应对的方法融入到全局拟合的框架中。例如, 在评估波形模板精度时, 需要考虑该类波源的估计误差对进一步搜寻其他

波源的影响; 在同时估计Glitch与波源参数时, 可以采用维度可变的RJMCMC搜索技术, 以应对Glitch数量未知的情况; 集成了多种数据处理技术的流水线势必面临计算效率的问题, 因而有必要优化算法效率, 发挥GPU和CPU等高性能计算硬件的运算潜力, 引入新兴的人工智能方法, 实现软硬件的优化组合。综上所述, 为通过数据分析实现空间引力波探测的科学目标, 仍有大量研究工作亟待开展。

## 参考文献

- 1 Aasi J, Abbott B P, Abbott R, et al. Advanced LIGO. *Class Quantum Grav*, 2015, 32: 074001
- 2 Acernese F, Agathos M, Agatsuma K, et al. Advanced Virgo: A second-generation interferometric gravitational wave detector. *Class Quantum Grav*, 2015, 32: 024001
- 3 Akutsu T, Ando M, Arai K, et al. Overview of KAGRA: Detector design and construction history. *Prog Theor Exp Phys*, 2021, 2021: 05A101
- 4 Abbott R, Abbott T D, Acernese F, et al. GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run. *Phys Rev X*, 2023, 13: 041039
- 5 Nitz A H, Kumar S, Wang Y F, et al. 4-OGC: Catalog of gravitational waves from compact binary mergers. *Astrophys J*, 2023, 946: 59
- 6 Venumadhav T, Zackay B, Roulet J, et al. New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo. *Phys Rev D*, 2020, 101: 083030
- 7 Zackay B, Venumadhav T, Dai L, et al. Highly spinning and aligned binary black hole merger in the Advanced LIGO first observing run. *Phys Rev D*, 2019, 100: 023007
- 8 Zackay B, Dai L, Venumadhav T, et al. Detecting gravitational waves with disparate detector responses: Two new binary black hole mergers. *Phys Rev D*, 2021, 104: 063030
- 9 Mehta A K, Olsen S, Wadekar D, et al. New binary black hole mergers in the ligo-virgo o3b data. arXiv: 2311.06061
- 10 Danzmann K, team L. LISA: Laser interferometer space antenna for gravitational wave measurements. *Class Quantum Grav*, 1996, 13: A247–A250
- 11 Colpi M, Danzmann K, Hewitson M, et al. LISA definition study report. arXiv: 2402.07571
- 12 Amaro-Seoane P, Audley H, Babak S, et al. Laser interferometer space antenna. arXiv: 1702.00786
- 13 Baker J, Bellovary J, Bender P L, et al. The laser interferometer space antenna: Unveiling the millihertz gravitational wave sky. arXiv: 1907.06482
- 14 Gong X, Xu S, Bai S, et al. A scientific case study of an advanced LISA mission. *Class Quantum Grav*, 2011, 28: 094012
- 15 Luo Z, Wang Y, Wu Y, et al. The Taiji program: A concise overview. *Prog Theor Exp Phys*, 2020, 2021: 05A108
- 16 Hu W R, Wu Y L. The Taiji Program in Space for gravitational wave physics and the nature of gravity. *Natl Sci Rev*, 2017, 4: 685–686
- 17 Wu Y L, Luo Z R, Wang J Y, et al. Taiji program in space for gravitational universe with the first run key technologies test in Taiji-1. *Int J Mod Phys A*, 2021, 36: 2102002
- 18 Wu Y L, Luo Z R, Wang J Y, et al. China's first step towards probing the expanding universe and the nature of gravity using a space borne gravitational wave antenna. *Commun Phys*, 2021, 4: 34
- 19 Mei J, Bai Y Z, Bao J, et al. The TianQin project: Current progress on science and technology. *Prog Theor Exp Phys*, 2020, 2021: 05A107
- 20 Luo J, Chen L S, Duan H Z, et al. TianQin: A space-borne gravitational wave detector. *Class Quantum Grav*, 2016, 33: 035010
- 21 Hu Y M, Mei J, Luo J. Science prospects for space-borne gravitational-wave missions. *Natl Sci Rev*, 2017, 4: 683–684
- 22 Luo J, Bai Y Z, Cai L, et al. The first round result from the TianQin-1 satellite. *Class Quantum Grav*, 2020, 37: 185013
- 23 Wise J H. The formation of the first black holes. arXiv: 2304.09311
- 24 Li K, Bogdanović T, Ballantyne D R, et al. Massive black hole binaries from the TNG50-3 simulation. I. Coalescence and LISA detection rates. *Astrophys J*, 2022, 933: 104
- 25 Li K, Bogdanović T, Ballantyne D R, et al. Massive black hole binaries from the TNG50-3 simulation. II. Using dual AGNs to predict the rate of black hole mergers. *Astrophys J*, 2023, 959: 3

- 26 Steinle N, Middleton H, Moore C J, et al. Implications of pulsar timing array observations for LISA detections of massive black hole binaries. *Mon Not R Astron Soc*, 2023, 525: 2851–2863
- 27 Fang Y, Yang H. Probing the delay time of supermassive black hole binary mergers with gravitational waves. *Mon Not R Astron Soc*, 2023, 523: 5120–5133
- 28 Stokov V, Fragione G, Berti E. LISA constraints on an intermediate-mass black hole in the galactic centre. *Mon Not R Astron Soc*, 2023, 524: 2033–2041
- 29 Barack L, Pound A. Self-force and radiation reaction in general relativity. *Rep Prog Phys*, 2019, 82: 016904
- 30 Gair J R, Babak S, Sesana A, et al. Prospects for observing extreme-mass-ratio inspirals with LISA. *J Phys-Conf Ser*, 2017, 840: 012021
- 31 Sesana A. Prospects for multiband gravitational-wave astronomy after GW150914. *Phys Rev Lett*, 2016, 116: 231102
- 32 Crowder J, Cornish N J. Solution to the galactic foreground problem for LISA. *Phys Rev D*, 2007, 75: 043008
- 33 Amaro-Seoane P, Aoudia S, Babak S, et al. Low-frequency gravitational-wave science with eLISA/NGO. *Class Quantum Grav*, 2012, 29: 124016
- 34 Littenberg T B, Cornish N J, Lackeos K, et al. Global analysis of the gravitational wave signal from galactic binaries. *Phys Rev D*, 2020, 101: 123021
- 35 Karnesis N, Babak S, Pironi M, et al. Characterization of the stochastic signal originating from compact binary populations as measured by LISA. *Phys Rev D*, 2021, 104: 043019
- 36 Zhang X H, Mohanty S D, Zou X B, et al. Resolving galactic binaries in LISA data using particle swarm optimization and cross-validation. *Phys Rev D*, 2021, 104: 024023
- 37 Strub S H, Ferraioli L, Schmelzbach C, et al. Bayesian parameter estimation of galactic binaries in LISA data with Gaussian process regression. *Phys Rev D*, 2022, 106: 062003
- 38 Nelemans G, Yungelson L R, Portegies Zwart S F. The gravitational wave signal from the galactic disk population of binaries containing two compact objects. *Astron Astrophys*, 2001, 375: 890–898
- 39 Breivik K, Coughlin S, Zevin M, et al. COSMIC variance in binary population synthesis. *Astrophys J*, 2020, 898: 71
- 40 Liu C, Ruan W H, Guo Z K. Confusion noise from galactic binaries for Taiji. *Phys Rev D*, 2023, 107: 064021
- 41 Armano M, Audley H, Baird J, et al. Transient acceleration events in LISA Pathfinder data: Properties and possible physical origin. *Phys Rev D*, 2022, 106: 062001
- 42 Bachman B, de Vine G, Dickson J, et al. Flight phasemeter on the laser ranging interferometer on the GRACE follow-on mission. *J Phys-Conf Ser*, 2017, 840: 012011
- 43 Cornish N J, Crowder J. LISA data analysis using Markov chain Monte Carlo methods. *Phys Rev D*, 2005, 72: 043005
- 44 Littenberg T B, Cornish N J. Prototype global analysis of LISA data with multiple source types. *Phys Rev D*, 2023, 107: 063004
- 45 Ren Z, Zhao T, Cao Z, et al. Taiji data challenge for exploring gravitational wave universe. *Front Phys*, 2023, 18: 64302
- 46 Li E K, Wang H, Chen H-Y, et al. GWSpace: A multi-mission science data simulator for space-based gravitational wave detection. arXiv: 2309.15020
- 47 Ni W T, Shy J T, Tseng S M, et al. Progress in mission concept study and laboratory development for the ASTROD (Astrodynamical Space Test of Relativity using Optical Devices). Allahdadi F A, Casani E K, Maclay T D, ed. In: Proceedings of Small Spacecraft, Space Environments, and Instrumentation Technologies, 1997. 105–116
- 48 Tinto M, Armstrong J W. Cancellation of laser noise in an unequal-arm interferometer detector of gravitational radiation. *Phys Rev D*, 1999, 59: 102003
- 49 Armstrong J W, Estabrook F B, Tinto M. Time-Delay Interferometry for space-based gravitational wave searches. *Astrophys J*, 1999, 527: 814–826
- 50 Estabrook F B, Tinto M, Armstrong J W. Time-delay analysis of LISA gravitational wave data: Elimination of spacecraft motion effects. *Phys Rev D*, 2000, 62: 042002
- 51 Vallisneri M. Geometric time delay interferometry. *Phys Rev D*, 2005, 72: 042003
- 52 Tinto M, Dhurandhar S V. Time-delay interferometry. *Living Rev Relativ*, 2021, 24: 1
- 53 Tinto M, Dhurandhar S, Malakar D. Second-generation time-delay interferometry. *Phys Rev D*, 2023, 107: 082001
- 54 Cutler C, Flanagan É E. Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform? *Phys Rev D*, 1994, 49: 2658–2697
- 55 Buonanno A, Iyer B R, Ochsner E, et al. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys Rev D*, 2009, 80: 084043

- 56 Buonanno A, Damour T. Effective one-body approach to general relativistic two-body dynamics. *Phys Rev D*, 1999, 59: 084006
- 57 Bohé A, Shao L, Taracchini A, 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*, 2017, 95: 044028
- 58 Cao Z, Han W B. Waveform model for an eccentric binary black hole based on the effective-one-body-numerical-relativity formalism. *Phys Rev D*, 2017, 96: 044028
- 59 Ossokine S, Buonanno A, Marsat S, et al. Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation. *Phys Rev D*, 2020, 102: 044055
- 60 Chiamello D, Nagar A. Faithful analytical effective-one-body waveform model for spin-aligned, moderately eccentric, coalescing black hole binaries. *Phys Rev D*, 2020, 101: 101501
- 61 Gadre B, Pürrer M, Field S E, et al. A fully precessing higher-mode surrogate model of effective-one-body waveforms. arXiv: 2203.00381
- 62 Pürrer M. Frequency-domain reduced order models for gravitational waves from aligned-spin compact binaries. *Class Quantum Grav*, 2014, 31: 195010
- 63 Pratten G, Husa S, García-Quirós C, et al. Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes. *Phys Rev D*, 2020, 102: 064001
- 64 García-Quirós C, Colleoni M, Husa S, et al. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Phys Rev D*, 2020, 102: 064002
- 65 Pratten G, García-Quirós C, Colleoni M, et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys Rev D*, 2021, 103: 104056
- 66 Blackman J, Field S E, Scheel M A, et al. Numerical relativity waveform surrogate model for generically precessing binary black hole mergers. *Phys Rev D*, 2017, 96: 024058
- 67 Varma V, Field S E, Scheel M A, et al. Surrogate models for precessing binary black hole simulations with unequal masses. *Phys Rev Res*, 2019, 1: 033015
- 68 Islam T, Varma V, Lodman J, et al. Eccentric binary black hole surrogate models for the gravitational waveform and remnant properties: Comparable mass, nonspinning case. *Phys Rev D*, 2021, 103: 064022
- 69 Liu X, Cao Z, Shao L. Upgraded waveform model of eccentric binary black hole based on effective-one-body-numerical-relativity for spin-aligned binary black holes. *Int J Mod Phys D*, 2023, 32: 2350015
- 70 Ramos-Buades A, Buonanno A, Estellés H, et al. Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes. *Phys Rev D*, 2023, 108: 124037
- 71 Marsat S, Baker J G, Canton T D. Exploring the Bayesian parameter estimation of binary black holes with LISA. *Phys Rev D*, 2021, 103: 083011
- 72 Pratten G, Schmidt P, Middleton H, et al. Precision tracking of massive black hole spin evolution with LISA. *Phys Rev D*, 2023, 108: 124045
- 73 Wang H, Harry I, Nitz A, et al. The challenge of eccentricity when observing stellar-mass binary black holes with space-based gravitational wave detectors. 2023
- 74 Gasparotto S, Vicente R, Blas D, et al. Can gravitational-wave memory help constrain binary black-hole parameters? A LISA case study. *Phys Rev D*, 2023, 107: 124033
- 75 Wang Z, Zhao J, Cao Z. Accuracy of numerical relativity waveforms with respect to space-based gravitational wave detectors. *Commun Theor Phys*, 2024, 76: 015403
- 76 Boyle M, Hemberger D, Izzo D A B, et al. The SXS collaboration catalog of binary black hole simulations. *Class Quantum Grav*, 2019, 36: 195006
- 77 Chatziioannou K, Klein A, Yunes N, et al. Constructing gravitational waves from generic spin-precessing compact binary inspirals. *Phys Rev D*, 2017, 95: 104004
- 78 Hinderer T, Flanagan É É. Two-timescale analysis of extreme mass ratio inspirals in Kerr spacetime: Orbital motion. *Phys Rev D*, 2008, 78: 064028
- 79 Pound A, Wardell B. *Black Hole Perturbation Theory and Gravitational Self-Force*, Singapore: Springer Singapore, 2020. 1–119
- 80 Teukolsky S A. Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations. *Astrophys J*, 1973, 185: 635–648
- 81 Han W B. Gravitational radiation from a spinning compact object around a supermassive Kerr black hole in circular orbit. *Phys Rev D*, 2010, 82: 084013
- 82 Han W B, Cao Z. Constructing effective one-body dynamics with numerical energy flux for intermediate-mass-ratio inspirals. *Phys Rev D*, 2011, 84: 044014

- 83 Han W B. Gravitational waves from extreme-mass-ratio inspirals in equatorially eccentric orbits. *Int J Mod Phys D*, 2014, 23: 1450064
- 84 Han W B. Fast evolution and waveform generator for extreme-mass-ratio inspirals in equatorial-circular orbits. *Class Quantum Grav*, 2016, 33: 065009
- 85 Barack L, Cutler C. LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. *Phys Rev D*, 2004, 69: 082005
- 86 Chua A J K, Moore C J, Gair J R. Augmented kludge waveforms for detecting extreme-mass-ratio inspirals. *Phys Rev D*, 2017, 96: 044005
- 87 Babak S, Fang H, Gair J R, et al. “Kludge” gravitational waveforms for a test-body orbiting a Kerr black hole. *Phys Rev D*, 2007, 75: 024005 [Erratum: *Phys. Rev. D* **77**, 04990 (2008)]
- 88 Yang S C, Han W B. Introduction of templates for low-frequency gravitational waves (in Chinese). *Acta Sci Nat Univ Sunyatseni*, 2021, 60: 99 [杨舒程, 韩文标. 低频引力波波形成模板介绍. *中山大学学报(自然科学版)*, 2021, 60: 99]
- 89 Zou X, Mohanty S, Xie Q, et al. An overview of EMRI data analysis. *Acta Sci Nat Univ Pekinensis*, 2024
- 90 Chen X, Han W B. Extreme-mass-ratio inspirals produced by tidal capture of binary black holes. *Commun Phys*, 2018, 1: 53
- 91 Sereno M, Sesana A, Bleuler A, et al. Strong lensing of gravitational waves as seen by LISA. *Phys Rev Lett*, 2010, 105: 251101
- 92 Takahashi R, Nakamura T. Wave effects in the gravitational lensing of gravitational waves from chirping binaries. *Astrophys J*, 2003, 595: 1039–1051
- 93 Lin X, Zhang J, Dai L, et al. Detecting strong gravitational lensing of gravitational waves with TianQin. *Phys Rev D*, 2023, 108: 064020
- 94 Cao Z, Li L F, Wang Y. Gravitational lensing effects on parameter estimation in gravitational wave detection with advanced detectors. *Phys Rev D*, 2014, 90: 062003
- 95 Toscani M, Burke O, Liu C, et al. Strongly lensed extreme mass-ratio inspirals. *Phys Rev D*, 2024, 109: 063505
- 96 Hu X C, Li X H, Wang Y, et al. Fundamentals of the orbit and response for TianQin. *Class Quantum Grav*, 2018, 35: 095008
- 97 Vallisneri M, Crowder J, Tinto M. Sensitivity and parameter-estimation precision for alternate LISA configurations. *Class Quantum Grav*, 2008, 25: 065005
- 98 Wang G, Ni W T. Numerical simulation of time delay interferometry for TAIJI and new LISA. *Res Astron Astrophys*, 2019, 19: 058
- 99 Prince T A, Tinto M, Larson S L, et al. LISA optimal sensitivity. *Phys Rev D*, 2002, 66: 122002
- 100 Wang G, Ni W T, Han W B, et al. Algorithm for time-delay interferometry numerical simulation and sensitivity investigation. *Phys Rev D*, 2021, 103: 122006
- 101 Adams M R, Cornish N J. Detecting a stochastic gravitational wave background in the presence of a galactic foreground and instrument noise. *Phys Rev D*, 2014, 89: 022001
- 102 Muratore M, Vetrugno D, Vitale S, et al. Time delay interferometry combinations as instrument noise monitors for LISA. *Phys Rev D*, 2022, 105: 023009
- 103 Muratore M, Hartwig O, Vetrugno D, et al. Effectiveness of null time-delay interferometry channels as instrument noise monitors in LISA. *Phys Rev D*, 2023, 107: 082004
- 104 Wang G, Li B, Xu P, et al. Characterizing instrumental noise and stochastic gravitational wave signals from combined time-delay interferometry. *Phys Rev D*, 2022, 106: 044054
- 105 Wang G, Ni W T. Revisiting time delay interferometry for unequal-arm LISA and TAIJI. *Phys Scr*, 2023, 98: 075005
- 106 Wang G. Time delay interferometry with minimal null frequencies. arXiv: 2403.01490v1
- 107 Cornish N J, Rubbo L J. LISA response function. *Phys Rev D*, 2003, 67: 022001
- 108 Estabrook F B, Wahlquist H D. Response of Doppler spacecraft tracking to gravitational radiation. *Gen Relat Gravit*, 1975, 6: 439–447
- 109 Vallisneri M. Synthetic LISA: Simulating time delay interferometry in a model LISA. *Phys Rev D*, 2005, 71: 022001
- 110 Katz M L, Bayle J B, Chua A J K, et al. Assessing the data-analysis impact of LISA orbit approximations using a GPU-accelerated response model. *Phys Rev D*, 2022, 106: 103001
- 111 Katz M L, Marsat S, Chua A J K, et al. GPU-accelerated massive black hole binary parameter estimation with LISA. *Phys Rev D*, 2020, 102: 023033
- 112 Marsat S, Baker J G. Fourier-domain modulations and delays of gravitational-wave signals. arXiv: 1806.10734
- 113 Cornish N J, Littenberg T B. Tests of Bayesian model selection techniques for gravitational wave astronomy. *Phys Rev D*, 2007, 76: 083006
- 114 Robson T, Cornish N J, Tamanini N, et al. Detecting hierarchical stellar systems with LISA. *Phys Rev D*, 2018, 98: 064012
- 115 Katz M L, Danielski C, Karnesis N, et al. Bayesian characterization of circumbinary sub-stellar objects with LISA. *Mon Not R Astron Soc*, 2022, 517: 697–711
- 116 Williams M J, Veitch J, Messenger C. Nested sampling with normalizing flows for gravitational-wave inference. *Phys Rev D*, 2021, 103: 103006

- 117 Lee E, Morisaki S, Tagoshi H. Mass-spin reparametrization for a rapid parameter estimation of inspiral gravitational-wave signals. *Phys Rev D*, 2022, 105: 124057
- 118 Islam T, Roulet J, Venumadhav T. Factorized parameter estimation for real-time gravitational wave inference. arXiv: 2210.16278
- 119 Wong K W K, Isi M, Edwards T D P. Fast gravitational-wave parameter estimation without compromises. *Astrophys J*, 2023, 958: 129
- 120 Pathak L, Reza A, Sengupta A S. Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources. *Phys Rev D*, 2023, 108: 064055
- 121 Tiwari V, Hoy C, Fairhurst S, et al. Fast non-Markovian sampler for estimating gravitational-wave posteriors. *Phys Rev D*, 2023, 108: 023001
- 122 Canizares P, Field S E, Gair J, et al. Accelerated gravitational wave parameter estimation with reduced order modeling. *Phys Rev Lett*, 2015, 114: 071104
- 123 Qi H, Raymond V. Python-based reduced order quadrature building code for fast gravitational wave inference. *Phys Rev D*, 2021, 104: 063031
- 124 Morisaki S. Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions. *Phys Rev D*, 2021, 104: 044062
- 125 Vinciguerra S, Veitch J, Mandel I. Accelerating gravitational wave parameter estimation with multi-band template interpolation. *Class Quantum Grav*, 2017, 34: 115006
- 126 Hartwig O, Muratore M. Characterization of time delay interferometry combinations for the LISA instrument noise. *Phys Rev D*, 2022, 105: 062006
- 127 Savalle E, Gair J, Speri L, et al. Assessing the impact of instrumental calibration uncertainty on LISA science. *Phys Rev D*, 2022, 106: 022003
- 128 Baghi Q, Korsakova N, Slutsky J, et al. Detection and characterization of instrumental transients in LISA Pathfinder and their projection to LISA. *Phys Rev D*, 2022, 105: 042002
- 129 Dey K, Karnesis N, Toubiana A, et al. Effect of data gaps on the detectability and parameter estimation of massive black hole binaries with LISA. *Phys Rev D*, 2021, 104: 044035
- 130 Baghi Q, Thorpe J I, Slutsky J, et al. Gravitational-wave parameter estimation with gaps in LISA: A Bayesian data augmentation method. *Phys Rev D*, 2019, 100: 022003
- 131 Peterseim N. Twangs-high-frequency disturbing signals in the 10 Hz accelerometer data of the grace satellites. 2014
- 132 Robson T, Cornish N J. Detecting gravitational wave bursts with LISA in the presence of instrumental glitches. *Phys Rev D*, 2019, 99: 024019
- 133 Spadaro A, Busicchio R, Vetrugno D, et al. Glitch systematics on the observation of massive black-hole binaries with LISA. *Phys Rev D*, 2023, 108: 123029
- 134 Littenberg T B, Cornish N J. Separating gravitational wave signals from instrument artifacts. *Phys Rev D*, 2010, 82: 103007
- 135 Green P J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 1995, 82: 711–732
- 136 Karnesis N, Katz M L, Korsakova N, et al. Eryn: A multipurpose sampler for Bayesian inference. *Mon Not R Astron Soc*, 2023, 526: 4814–4830
- 137 Röver C. Student-t based filter for robust signal detection. *Phys Rev D*, 2011, 84: 122004
- 138 Röver C, Meyer R, Christensen N. Modelling coloured residual noise in gravitational-wave signal processing. *Class Quantum Grav*, 2010, 28: 015010
- 139 Sasli A, Karnesis N, Stergioulas N. Heavy-tailed likelihoods for robustness against data outliers: Applications to the analysis of gravitational wave data. *Phys Rev D*, 2023, 108: 103005
- 140 Martellini L, Regimbau T. Semiparametric approach to the detection of non-Gaussian gravitational wave stochastic backgrounds. *Phys Rev D*, 2014, 89: 124009
- 141 Carre J, Porter E K. The effect of data gaps on LISA galactic binary parameter estimation. arXiv: 1010.1641
- 142 Pollack S E. LISA science results in the presence of data disturbances. *Class Quantum Grav*, 2004, 21: 3419–3432
- 143 Blelly A, Bobin J, Moutarde H. Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data. *Mon Not R Astron Soc*, 2021, 509: 5902–5917
- 144 Armano M, Audley H, Baird J, et al. Calibrating the system dynamics of LISA pathfinder. *Phys Rev D*, 2018, 97: 122002
- 145 Armano M, Audley H, Baird J, et al. Beyond the required LISA free-fall performance: New LISA pathfinder results down to 20  $\mu\text{Hz}$ . *Phys Rev Lett*, 2018, 120: 061101
- 146 Armano M, Audley H, Auger G, et al. Sub-Femto-g free fall for space-based gravitational wave observatories: LISA pathfinder results. *Phys Rev Lett*, 2016, 116: 231101
- 147 Hartig M S, Schuster S, Wanner G. Geometric tilt-to-length coupling in precision interferometry: Mechanisms and analytical descriptions. *J Opt*, 2022, 24: 065601
- 148 Sathyaprakash B S, Schutz B F. Physics, astrophysics and cosmology with gravitational waves. *Living Rev Relativ*, 2009, 12: 2

- 149 Cornish N J. Time-frequency analysis of gravitational wave data. *Phys Rev D*, 2020, 102: 124038
- 150 Edwards M C, Maturana-Russel P, Meyer R, et al. Identifying and addressing nonstationary LISA noise. *Phys Rev D*, 2020, 102: 084062
- 151 Lin S, Hu B, Zhang X H, et al. White dwarf binary modulation can help stochastic gravitational wave background search. *Sci China-Phys Mech Astron*, 2023, 66: 299512
- 152 Cornish N J, Littenberg T B. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Class Quantum Grav*, 2015, 32: 135012
- 153 Zackay B, Venumadhav T, Roulet J, et al. Detecting gravitational waves in data with non-stationary and non-Gaussian noise. *Phys Rev D*, 2021, 104: 063034
- 154 Mozzon S, Nuttall L K, Lundgren A, et al. Dynamic normalization for compact binary coalescence searches in non-stationary noise. *Class Quantum Grav*, 2020, 37: 215014
- 155 Gurland J, Whittle P. Hypothesis testing in time series analysis. *J Am Statist Assoc*, 1954, 49: 197
- 156 Whittle P. Curve and periodogram smoothing. *J R Statist Soc Ser B*, 1957, 19: 38–47
- 157 Choudhuri N. Contiguity of the Whittle measure for a Gaussian time series. *Biometrika*, 2004, 91: 211–218
- 158 Romano J D, Cornish N J. Detection methods for stochastic gravitational-wave backgrounds: A unified treatment. *Living Rev Relativ*, 2017, 20: 2
- 159 Thrane E, Talbot C. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publ Astron Soc Aust*, 2019, 36: e010
- 160 Veitch J, Raymond V, Farr B, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys Rev D*, 2015, 91: 042003
- 161 Ashton G, Hübner M, Lasky P D, et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *Astrophys J Suppl Ser*, 2019, 241: 27
- 162 Contreras-Cristán A, Gutiérrez-Peña E, Walker S G. A note on Whittle's likelihood. *Commun Stat-Simul Comput*, 2006, 35: 857–875
- 163 Cornish N J. Fast fisher matrices and lazy likelihoods. arXiv: 1007.4820
- 164 Zackay B, Dai L, Venumadhav T. Relative binning and fast likelihood evaluation for gravitational wave parameter estimation. arXiv: 1806.08793
- 165 Cornish N J. Heterodyned likelihood for rapid gravitational wave parameter inference. *Phys Rev D*, 2021, 104: 104054
- 166 Cornish N J. Low latency detection of massive black hole binaries. *Phys Rev D*, 2022, 105: 044007
- 167 Hoy C, Nuttall L K. Bilby in space: Bayesian inference for transient gravitational-wave signals observed with LISA. arXiv: 2312.13039
- 168 Weaving C R, Nuttall L K, Harry I W, et al. Adapting the PyCBC pipeline to find and infer the properties of gravitational waves from massive black hole binaries in LISA. *Class Quantum Grav*, 2023, 41: 025006
- 169 Leslie N, Dai L, Pratten G. Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics. *Phys Rev D*, 2021, 104: 123030
- 170 Linley J. Fast LISA likelihood approximations by downsampling. arXiv: 2402.01819
- 171 Finn L S. Detection, measurement, and gravitational radiation. *Phys Rev D*, 1992, 46: 5236–5249
- 172 Vallisneri M. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys Rev D*, 2008, 77: 042001
- 173 Vecchio A. LISA observations of rapidly spinning massive black hole binary systems. *Phys Rev D*, 2004, 70: 042001
- 174 Berti E, Buonanno A, Will C M. Estimating spinning binary parameters and testing alternative theories of gravity with LISA. *Phys Rev D*, 2005, 71: 084025
- 175 Lang R N, Hughes S A. Measuring coalescing massive binary black holes with gravitational waves: The impact of spin-induced precession. *Phys Rev D*, 2006, 74: 122001
- 176 Arun K G, Babak S, Berti E, et al. Massive black-hole binary inspirals: Results from the LISA parameter estimation taskforce. *Class Quantum Grav*, 2009, 26: 094027
- 177 Katz M L. Fully automated end-to-end pipeline for massive black hole binary signal extraction from LISA data. *Phys Rev D*, 2022, 105: 044055
- 178 Bouffanais Y, Porter E K. Detecting compact galactic binaries using a hybrid swarm-based algorithm. *Phys Rev D*, 2016, 93: 064020
- 179 Gao P, Fan X L, Cao Z J, et al. Fast resolution of galactic binaries in LISA data. *Phys Rev D*, 2023, 107: 123029
- 180 Strub S H, Ferraioli L, Schmelzbach C, et al. Accelerating global parameter estimation of gravitational waves from galactic binaries using a genetic algorithm and GPUs. arXiv: 2307.03763
- 181 Crowder J, Cornish N J. Extracting galactic binary signals from the first round of Mock LISA Data Challenges. *Class Quantum Grav*, 2007, 24: S575–S585

- 182 Littenberg T B. Detection pipeline for galactic binaries in LISA data. *Phys Rev D*, 2011, 84: 063009
- 183 Klein A, Pratten G, Buscicchio R, et al. The last three years: Multiband gravitational-wave observations of stellar-mass binary black holes. arXiv: 2204.03423
- 184 Finch E, Bartolucci G, Chucherko D, et al. Identifying LISA verification binaries among the Galactic population of double white dwarfs. *Mon Not R Astron Soc*, 2023, 522: 5358–5373
- 185 Metropolis N, Rosenbluth A W, Rosenbluth M N, et al. Equation of state calculations by fast computing machines. *J Chem Phys*, 1953, 21: 1087–1092
- 186 Hastings W K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, 57: 97–109
- 187 Brooks S, Gelman A, Jones G, et al. Handbook of Markov Chain Monte Carlo. Boca Raton: CRC Press, 2011
- 188 Betancourt M J, Girolami M. Hamiltonian monte carlo for hierarchical models. arXiv: 1312.0906
- 189 Homan M D, Gelman A. The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian Monte Carlo. *J Mach Learn Res*, 2014, 15: 1593–1623
- 190 Muller P. A genetic approach to posterior integration and Gibbs sampling. Technical Report. Department of Statistics, Purdue University, 1991
- 191 Gilks W R, Wild P. Adaptive rejection sampling for Gibbs sampling. *Appl Stat*, 1992, 41: 337
- 192 Ritter C, Tanner M A. Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *J Am Statist Assoc*, 1992, 87: 861–868
- 193 Andrieu C, de Freitas N, Doucet A, et al. *Mach Learn*, 2003, 50: 5–43
- 194 Swendsen R H, Wang J S. Replica Monte Carlo simulation of spin-glasses. *Phys Rev Lett*, 1986, 57: 2607–2609
- 195 Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. *J Phys Soc Jpn*, 1996, 65: 1604–1608
- 196 Vousden W D, Farr W M, Mandel I. Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations. *Mon Not R Astron Soc*, 2015, 455: 1919–1937
- 197 Pankow C, Brady P, Ochsner E, et al. Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. *Phys Rev D*, 2015, 92: 023002
- 198 Lange J, O’Shaughnessy R, Rizzo M. Rapid and accurate parameter inference for coalescing, precessing compact binaries. arXiv: 1805.10457
- 199 Fairhurst S, Hoy C, Green R, et al. Simple parameter estimation using observable features of gravitational-wave signals. *Phys Rev D*, 2023, 108: 082006
- 200 Delaunoy A, Wehenkel A, Hinderer T, et al. Lightning-fast gravitational wave parameter inference through neural amortization. arXiv: 2010.12931
- 201 Green S R, Simpson C, Gair J. Gravitational-wave parameter estimation with autoregressive neural network flows. *Phys Rev D*, 2020, 102: 104057
- 202 Green S R, Gair J. Complete parameter inference for GW150914 using deep learning. *Mach Learn-Sci Technol*, 2021, 2: 03LT01
- 203 Dax M, Green S R, Gair J, et al. Real-time gravitational wave science with neural posterior estimation. *Phys Rev Lett*, 2021, 127: 241103
- 204 Gabbard H, Messenger C, Heng I S, et al. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *Nat Phys*, 2022, 18: 112–117
- 205 Cuoco E, Powell J, Cavagliá M, et al. Enhancing gravitational-wave science with machine learning. *Mach Learn-Sci Technol*, 2020, 2: 011002
- 206 Chua A J K, Galley C R, Vallisneri M. Reduced-order modeling with artificial neurons for gravitational-wave inference. *Phys Rev Lett*, 2019, 122: 211101
- 207 Chua A J K, Katz M L, Warburton N, et al. Rapid generation of fully relativistic extreme-mass-ratio-inspiral waveform templates for LISA data analysis. *Phys Rev Lett*, 2021, 126: 051102
- 208 Katz M L, Chua A J K, Speri L, et al. Fast extreme-mass-ratio-inspiral waveforms: New tools for millihertz gravitational-wave data analysis. *Phys Rev D*, 2021, 104: 064047
- 209 Schmidt S, Breschi M, Gamba R, et al. Machine learning gravitational waves from binary black hole mergers. *Phys Rev D*, 2021, 103: 043020
- 210 Lee J, Oh S H, Kim K, et al. Deep learning model on gravitational waveforms in merging and ringdown phases of binary black hole coalescences. *Phys Rev D*, 2021, 103: 123023
- 211 Khan A, Huerta E A, Zheng H. Interpretable AI forecasting for numerical relativity waveforms of quasicircular, spinning, nonprecessing binary black hole mergers. *Phys Rev D*, 2022, 105: 024024
- 212 Shi R, Zhou Y, Zhao T, et al. Compact binary systems waveform generation with generative pre-trained transformer. arXiv: 2310.20172
- 213 Liao C H, Lin F L. Deep generative models of gravitational waveforms via conditional autoencoder. *Phys Rev D*, 2021, 103: 124051
- 214 Bayle J B. Simulation and data analysis for LISA: Instrumental modeling, time-delay interferometry, noise-reduction performance study, and discrimination of transient gravitational signals. Paris U. VII, APC, 2019

- 215 Houba N, Ferraioli L, Giardini D. Detection and mitigation of glitches in LISA data: A machine learning approach. *Phys Rev D*, 2024, 109: 083027
- 216 Xu Y, Du M, Xu P, et al. Gravitational wave signal extraction against non-stationary instrumental noises with deep neural network. arXiv: 2402.13091
- 217 Zhang X T, Messenger C, Korsakova N, et al. Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks. *Phys Rev D*, 2022, 105: 123027
- 218 Zhao T, Zhou Y, Shi R, et al. Decode: Dilated convolutional neural network for detecting extreme-mass-ratio inspirals. arXiv: 2308.16422
- 219 Yun Q, Han W B, Guo Y Y, et al. Detecting extreme-mass-ratio inspirals for space-borne detectors with deep learning. arXiv: 2309.06694
- 220 Yun Q, Han W B, Guo Y Y, et al. The detection, extraction and parameter estimation of extreme-mass-ratio inspirals with deep learning. arXiv: 2311.18640
- 221 Ruan W H, Wang H, Liu C, et al. Rapid search for massive black hole binary coalescences using deep learning. *Phys Lett B*, 2023, 841: 137904
- 222 Wang H, Wu S, Cao Z, et al. Gravitational-wave signal recognition of LIGO data by deep learning. *Phys Rev D*, 2020, 101: 104003
- 223 Ruan W H, Guo Z K. Pre-merger detection of massive black hole binaries using deep learning. arXiv: 2402.16282
- 224 Chua A J K, Vallisneri M. Learning Bayesian posteriors with neural networks for gravitational-wave inference. *Phys Rev Lett*, 2020, 124: 041102
- 225 Ruan W, Wang H, Liu C, et al. Parameter inference for coalescing massive black hole binaries using deep learning. *Universe*, 2023, 9: 407
- 226 Du M, Liang B, Wang H, et al. Advancing space-based gravitational wave astronomy: Rapid parameter estimation via normalizing flows. *Sci China-Phys Mech Astron*, 2024, 67: 230412
- 227 Xiong C Y, Sun T Y, Zhang J F, et al. Robust inference of gravitational wave source parameters in the presence of noise transients using normalizing flows. arXiv: 2405.09475
- 228 Zhao T, Lyu R, Wang H, et al. Space-based gravitational wave signal detection and extraction with deep neural network. *Commun Phys*, 2023, 6: 212
- 229 Ashton G, Talbot C. Bilby-MCMC: An MCMC sampler for gravitational-wave inference. *Mon Not R Astron Soc*, 2021, 507: 2037–2051
- 230 Korsakova N, Babak S, Katz M L, et al. Neural density estimation for galactic binaries in lisa data analysis. arXiv: 2402.13701
- 231 Wouters T, Pang P T H, Dietrich T, et al. Robust parameter estimation within minutes on gravitational wave signals from binary neutron star inspirals. arXiv: 2404.11397

# Challenges in space-based gravitational wave data analysis and applications of artificial intelligence

WANG He<sup>1,2\*</sup>, DU MingHui<sup>3\*</sup>, XU Peng<sup>2,3,4,5</sup> & ZHOU Yu-Feng<sup>6</sup>

<sup>1</sup>*International Centre for Theoretical Physics Asia-Pacific, University of Chinese Academy of Sciences, Beijing 100190, China;*

<sup>2</sup>*Taiji Laboratory for Gravitational Wave Universe (Beijing/Hangzhou), University of Chinese Academy of Sciences (UCAS), Beijing 100049, China;*

<sup>3</sup>*Center for Gravitational Wave Experiment, National Microgravity Laboratory, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China;*

<sup>4</sup>*Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences (UCAS), Hangzhou 310024, China;*

<sup>5</sup>*Lanzhou Center of Theoretical Physics, Lanzhou University, Lanzhou 730000, China;*

<sup>6</sup>*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

*\*Corresponding authors (WANG He, email: [hewang@ucas.ac.cn](mailto:hewang@ucas.ac.cn); DU MingHui, email: [duminghui@imech.ac.cn](mailto:duminghui@imech.ac.cn))*

As space-based gravitational wave detection projects such as LISA, Taiji, and Tianqin continue to advance, we are on the cusp of gaining a new viewpoint on observing the universe. However, the scientific data processing for these projects faces unprecedented challenges, including the superposition of numerous gravitational wave sources, non-stationary noises, and data anomalies. This review aims to make a brief summary of these challenges and their possible solutions, using the Bayesian statistical inference framework as a thread, and provide researchers with a relatively comprehensive perspective. Topics such as the construction of waveform templates, the modeling of detector responses, and the processing of noise and data anomaly are discussed, with a focus on the strategies for parameter estimation and global fitting, especially the evaluation of likelihood, and the utilization of various stochastic sampling techniques to improve the efficiency and accuracy of analysis. Notably, this review highlights the applications of artificial intelligence technologies in waveform modeling, noise and data anomaly processing, signal recognition, and parameter estimation, showcasing how artificial intelligence can pave new paths for solving complex problems in the data analysis of space-based gravitational wave detection.

**gravitational-wave detection, data analysis, parameter estimation, artificial intelligence**

**PACS:** 04.30.-w, 04.80.Nn, 07.05.Kf, 07.05.Mh

**doi:** [10.1360/SSPMA-2024-0087](https://doi.org/10.1360/SSPMA-2024-0087)