

# 基于人工智能技术的引力波数据分析前沿

Enhancing Gravitational-wave Astronomy with AI Techniques

王赫

[hewang@ucas.ac.cn](mailto:hewang@ucas.ac.cn)

中国科学院大学·国际理论物理中心（亚太地区）  
on behalf of the LIGO-Virgo-KAGRA Collaboration

天文信息学与虚拟天文台 2022 年学术年会

2023年4月20日 15:45-16:00



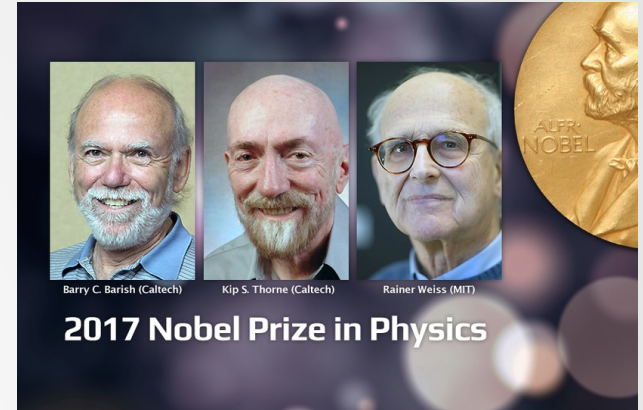
ICTP-AP  
International Centre  
for Theoretical Physics Asia-Pacific  
国际理论物理中心-亚太地区

# 目录

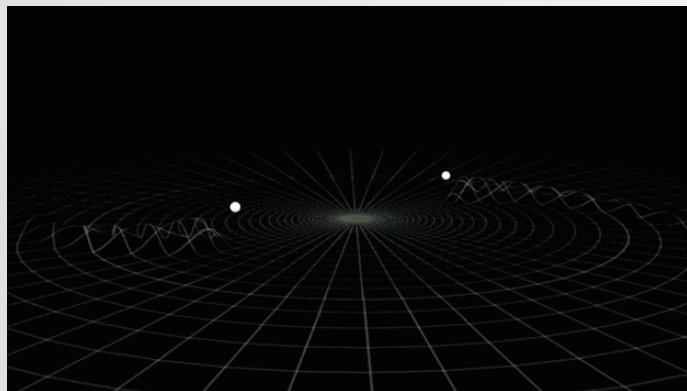
- 引力波天文学
- 人工智能与引力波数据处理
  - 引力波信号探测
  - 引力波数据降噪
- 未来展望与规划

# 引力波天文学

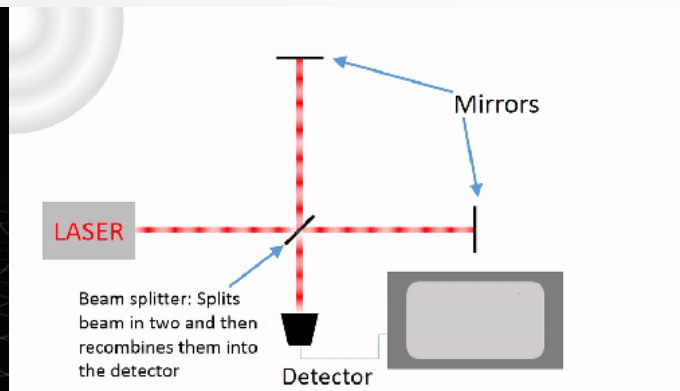
- 引力波是广义相对论中的一种强场效应
  - 2015年：首次实验探测到双黑洞并合引力波
  - 2017年：首次双中子星多信使探测，开启多信使天文学时代
  - 2017年：引力波探测成果被授予诺贝尔物理学奖
  - 至今：发现了超过 90 个引力波事件
- 未来：
  - 2023-2024年：有希望探测更多不同类型的引力波事件
  - 空间引力波探测计划 (LISA/Taiji/Tianqin) + XG (CE/ET)



2017 年诺贝尔物理学奖



双星并合系统产生的引力波波源



引力波振幅的测量



地面引力波探测器网络

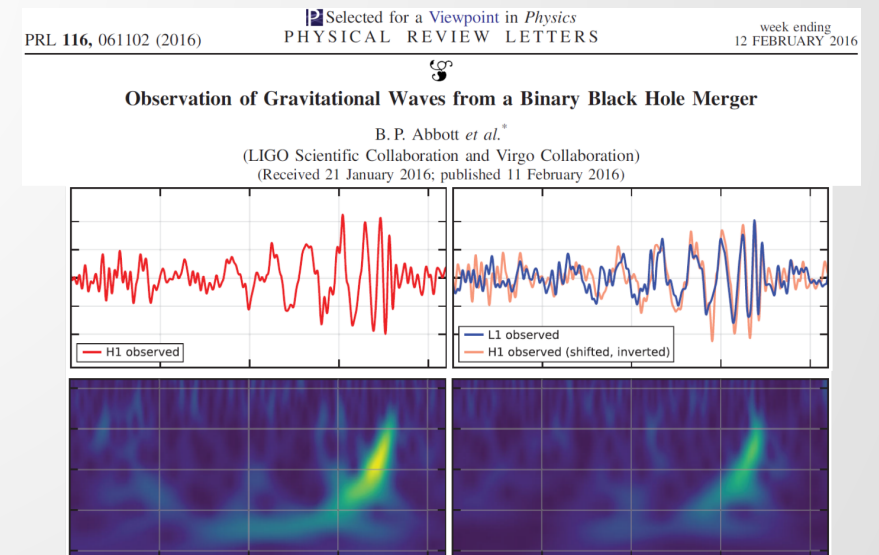
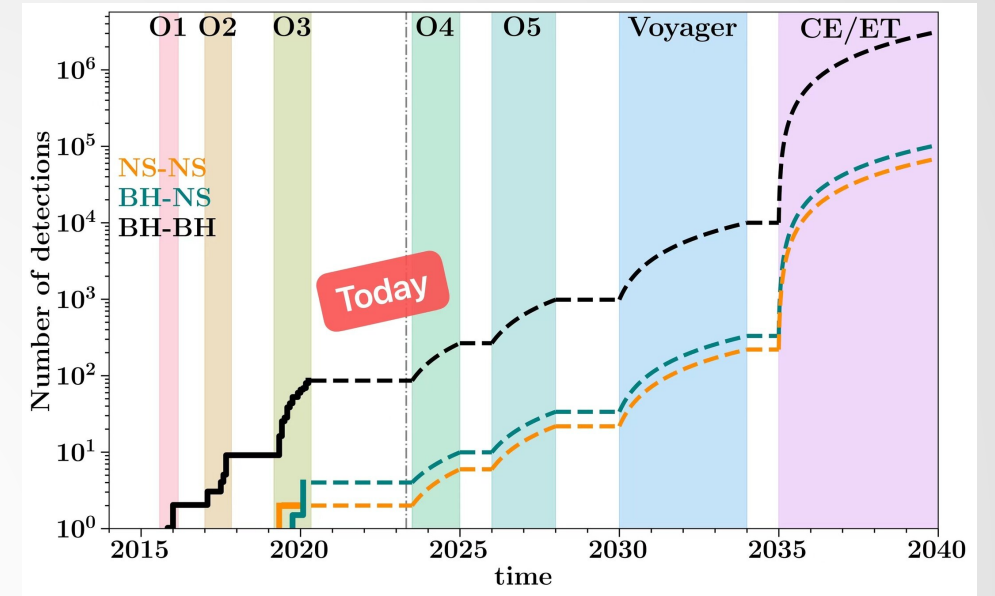
# 引力波天文学

- 基础理论的检验与修正
  - 基础物理学
    - 引力子是否有质量, 引力波的传播速度 ...
  - 天体物理学
    - 大质量恒星演化模型, 恒星级双黑洞的形成机制 ...
  - 宇宙学
    - 哈勃常数的测量, 暗能量 ...

- 伯纳德·舒尔茨曾列出成功观测引力波的五条关键要素:

1. 良好的探测器技术
2. 良好的波形模板
3. 良好的**数据分析**方法和技术
4. 多个独立探测器间的一致性观测
5. 引力波天文学和电磁波天文学的一致性观测

DOI:10.1063/1.1629411

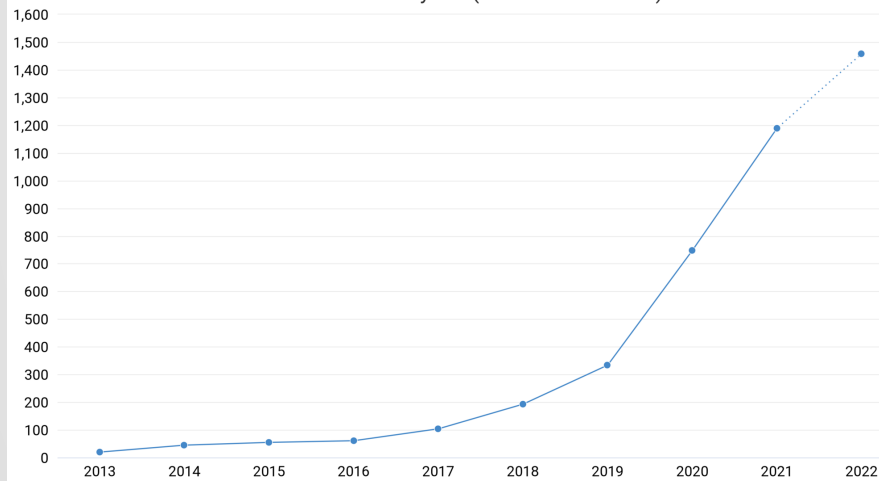


首次探测双黑洞并合引力波事件 GW150914

# 引力波天文学与人工智能

- 基于数据驱动机器学习技术的引力波数据分析方向成为新的**研究热点**
- 通过机器学习算法对大规模引力波数据进行分析和处理，可以更加准确地识别引力波信号以及实现引力波波源参数反演
- 机器学习技术还可以帮助引力波研究人员深入研究宇宙中的各种现象和物理规律，推动引力波天文学的发展

Citations in each year. (Criteria: see below)



Source: <https://app.dimensions.ai>  
Exported: December 20, 2022  
Criteria: 'gravitational waves learning' in title and abstract.  
© 2022 Digital Science and Research Solutions Inc. All rights reserved. Non-commercial redistribution / external re-use of this work is permitted subject to appropriate acknowledgement. This work is sourced from Dimensions® at [www.dimensions.ai](https://www.dimensions.ai)

Mach. Learn.: Sci. Technol. 2 (2021) 011002

<https://doi.org/10.1088/2632-2153/abb93a>

MACHINE  
LEARNING  
Science and Technology

TOPICAL REVIEW

## Enhancing **gravitational-wave** science with machine learning

Elena Cuoco<sup>1,2,3</sup>, Jade Powell<sup>4</sup>, Marco Cavaglia<sup>5,6</sup>, Kendall Ackley<sup>6,7</sup>, Michal Bejger<sup>8</sup>, Chayan Chatterjee<sup>9</sup>, Michael Coughlin<sup>10,11</sup>, Scott Coughlin<sup>12</sup>, Paul Easter<sup>13</sup>, Reed Essick<sup>13</sup>, Hunter Gabbard<sup>14</sup>, Timothy Gebhard<sup>15,16</sup>, Shaon Ghosh<sup>17</sup>, Leila Haegel<sup>18</sup>, Alberto Iess<sup>19,20</sup>, David Keitel<sup>21</sup>, Zsuzsa Márka<sup>22</sup>, Szabolcs Márka<sup>23</sup>, Filip Morawski<sup>24</sup>, Tri Nguyen<sup>25</sup>, Rich Ormiston<sup>25</sup>, Michael Pürrer<sup>26</sup>, Massimiliano Razzano<sup>3,27</sup>, Kai Staats<sup>12</sup>, Gabriele Vajente<sup>10</sup> and Daniel Williams<sup>14</sup>

nature reviews physics

Explore content ▾ About the journal ▾ Publish with us ▾

nature > nature reviews physics > expert recommendation > article

Expert Recommendation | Published: 03 October 2019

## Enabling real-time **multi-messenger astrophysics** discoveries with deep learning

nature  
computational  
science

PERSPECTIVE

<https://doi.org/10.1038/s43588-022-00288-z>

Check for updates

## Computational challenges for **multimodal** astrophysics

Elena Cuoco<sup>1,2,3</sup>, Barbara Patricelli<sup>1,3,4</sup>, Alberto Iess<sup>2,3</sup> and Filip Morawski<sup>5</sup>

ARTICLES

<https://doi.org/10.1038/s41567-021-01425-7>

nature  
physics

Check for updates

## Bayesian parameter estimation using conditional variational autoencoders for **gravitational-wave** astronomy

Hunter Gabbard<sup>1</sup>, Chris Messenger<sup>1</sup>, Ik Siang Heng<sup>1</sup>, Francesco Tonolini<sup>2</sup> and Roderick Murray-Smith<sup>2</sup>

nature  
astronomy

ARTICLES

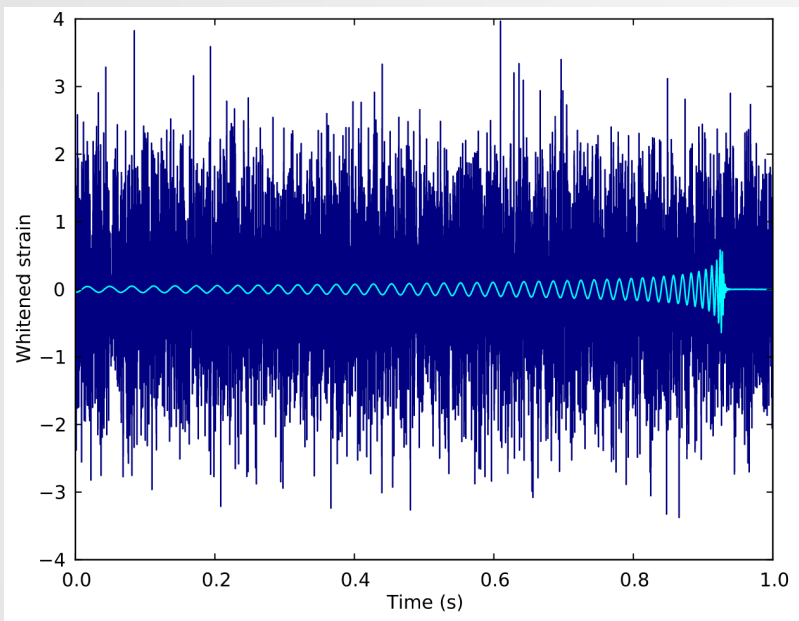
<https://doi.org/10.1038/s41550-021-01405-0>

Check for updates

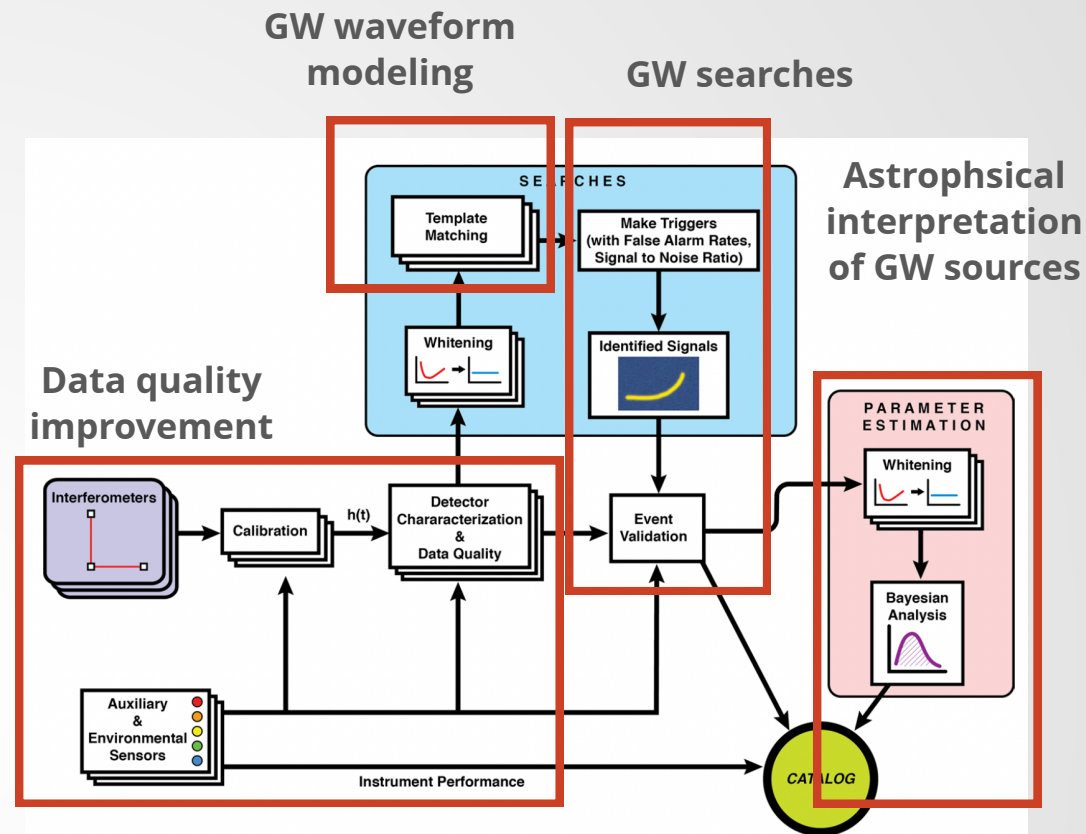
## Accelerated, scalable and reproducible AI-driven **gravitational wave** detection

E. A. Huerta<sup>1,2</sup>, Asad Khan<sup>3</sup>, Xiaobo Huang<sup>3</sup>, Minyang Tian<sup>3</sup>, Maksim Levental<sup>2</sup>, Ryan Chard<sup>1</sup>, Wei Wei<sup>3</sup>, Maeve Heflin<sup>3</sup>, Daniel S. Katz<sup>3</sup>, Volodymyr Kindratenko<sup>3</sup>, Dawei Mu<sup>3</sup>, Ben Blaiszik<sup>1,2</sup> and Ian Foster<sup>1,2</sup>

# 引力波天文学与人工智能



- 地面引力波探测科学数据的特点
  - 噪声特点：非高斯 + 非稳态
  - 信号特点：信噪比低  
(约噪声幅度的1/100, -60 dB)

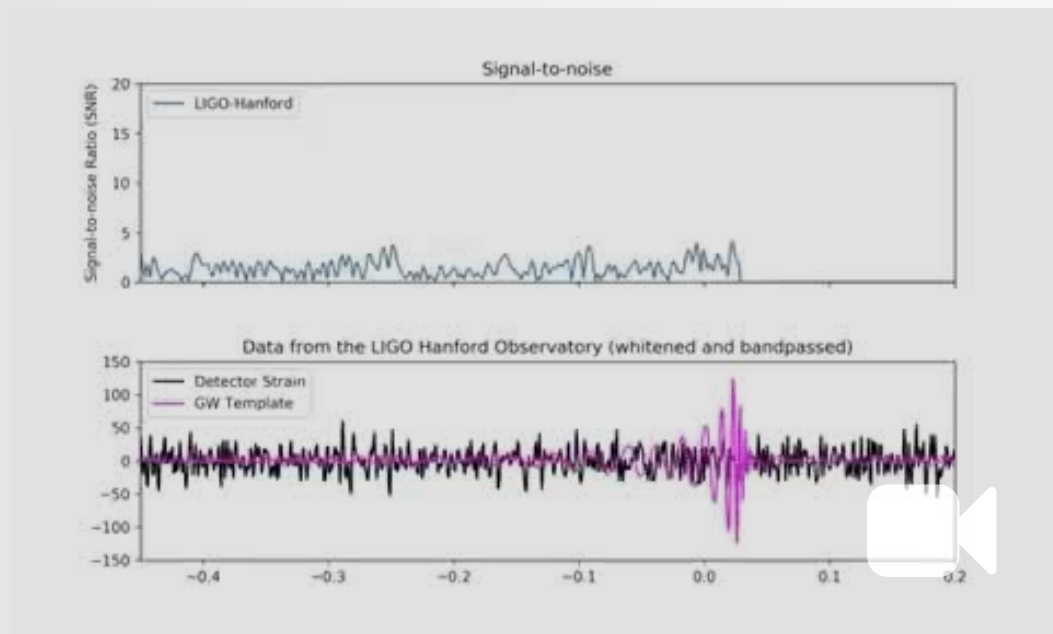


LIGO-Virgo 数据处理流程

Credit: Marco Cavaglia  
CQG. 37 (2020) 055002

- 深度学习技术应用在引力波科学数据处理流程各个环节：
  - **数据质量提升**：对非高斯短时脉冲噪声的抑制 (deglitch)
  - **波形建模**：快速生成理论精确模版
  - **信号搜寻**：快速准确识别引力波信号
  - **波源参数反演**：快速准确的后验概率分布估计

# 引力波天文学与人工智能：信号探测



## • 匹配滤波方法 (*Matched filtering*)

- 在**高斯且稳态**噪声环境下，提取**弱信号**的**最优线性**算法
- 从信号处理角度理解：
  - 输入：某一段时域数据
  - 输出：另一段时序数据 (探测统计量，即匹配滤波信噪比)
  - 求解：怎样的**线性滤波器**可以使得输出结果最大？

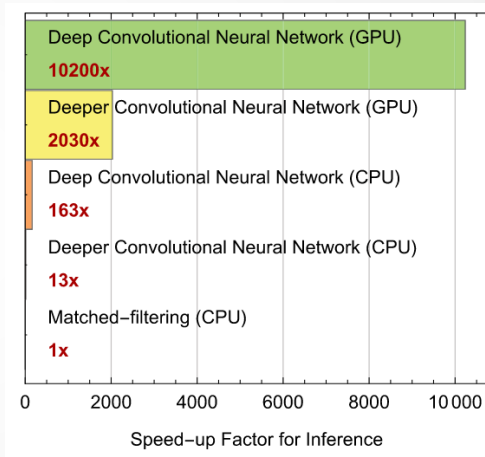
PHYSICAL REVIEW LETTERS **120**, 141103 (2018)

Editors' Suggestion

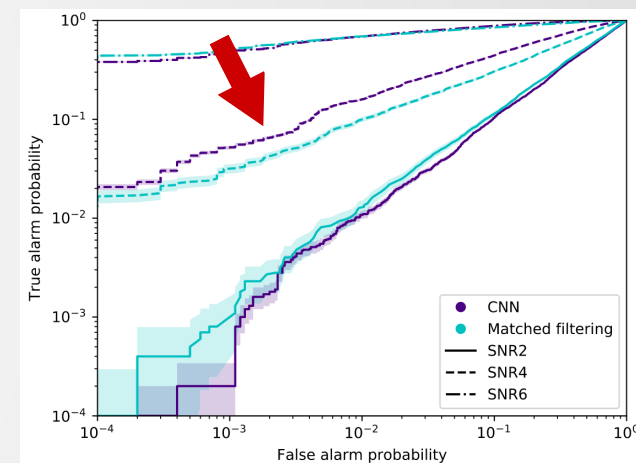
## Matching Matched Filtering with Deep Networks for Gravitational-Wave Astronomy

Hunter Gabbard,<sup>\*</sup> Michael Williams, Fergus Hayes, and Chris Messenger  
SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

(Received 16 December 2017; revised manuscript received 12 February 2018; published 6 April 2018)



PRD, 2018, 97(4): 044039.

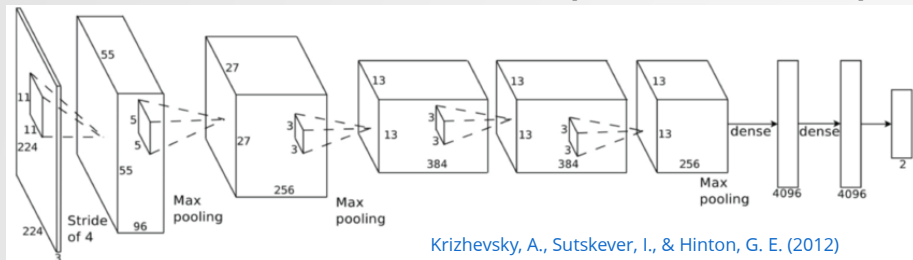


PRL, 2018, 120(14): 141103.

- 在原理验证阶段，**卷积神经网络**可以实现与传统数据处理方法相比拟的性能，在执行速度方面**完胜**(GPU加持)
- 许多研究者从不同的角度和问题出发，验证在**高斯稳态**噪声和**模拟**的理论波形下，卷积神经网络的优异性能。

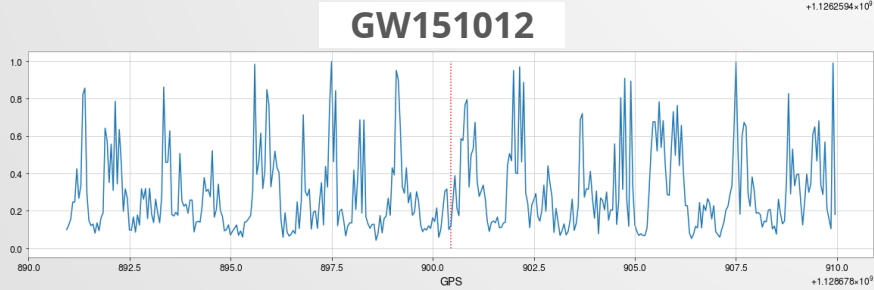
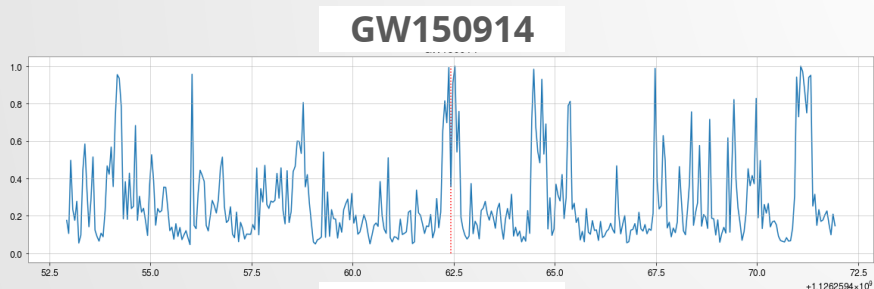
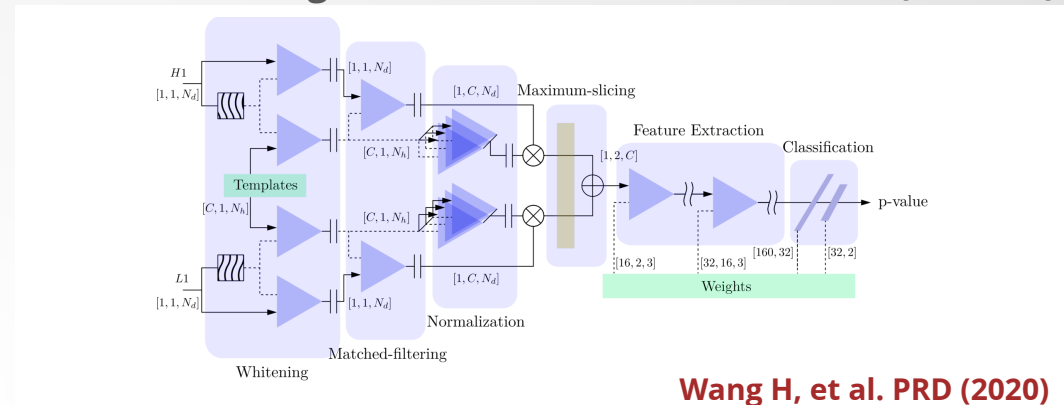
# 引力波天文学与人工智能：信号探测

## Convolutional Neural Network (ConvNet or CNN)

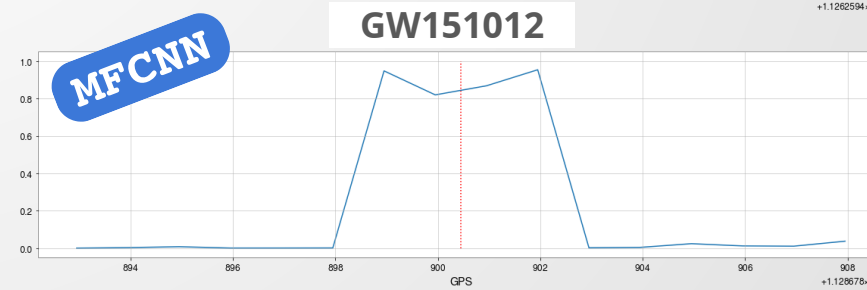
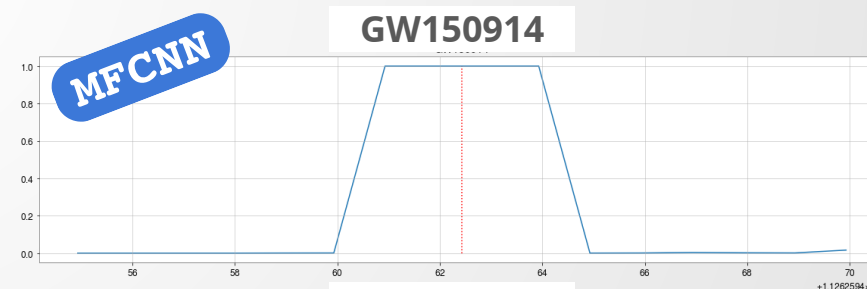


- 将 CNN 模型在真实噪声和真实引力波事件上测试，效果很糟糕 🤔

## Matched-filtering Convolutional Neural Network (MFCNN)



GPS 时间轴



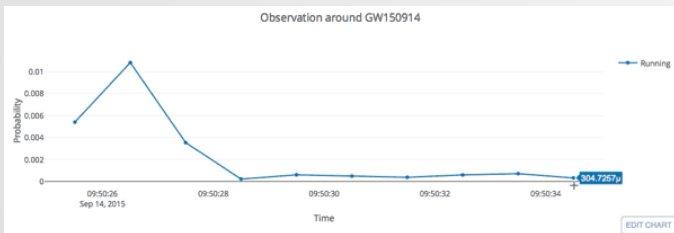
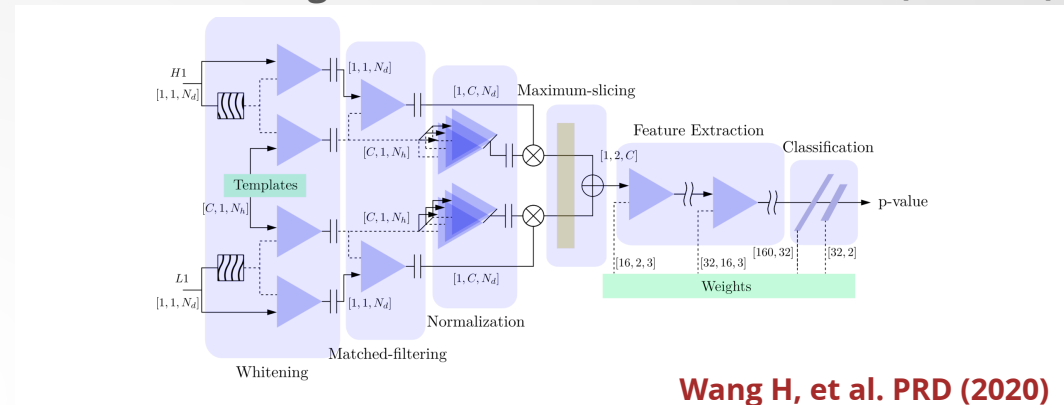
GPS 时间轴



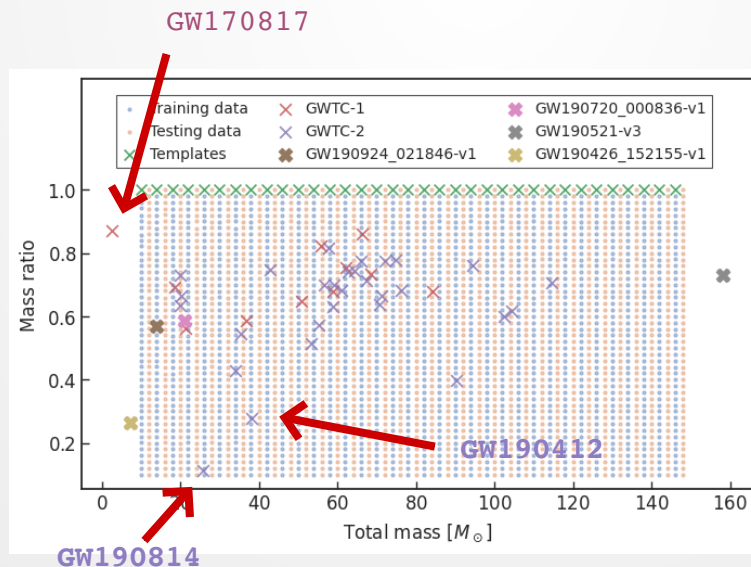
# 引力波天文学与人工智能：信号探测

- 改进并开发神经网络模型，以适应真实的引力波观测数据的任务
- 匹配滤波算法当中的波形模板 → 卷积层中的卷积核权重参数
- **匹配滤波感知层 (matched-filtering layer)**
- 可以准确探测到 GWTC-1 中的 11 个真实引力波事件，甚至包括 GW170817
- 引力波信号处理 → 智能引力波信号处理

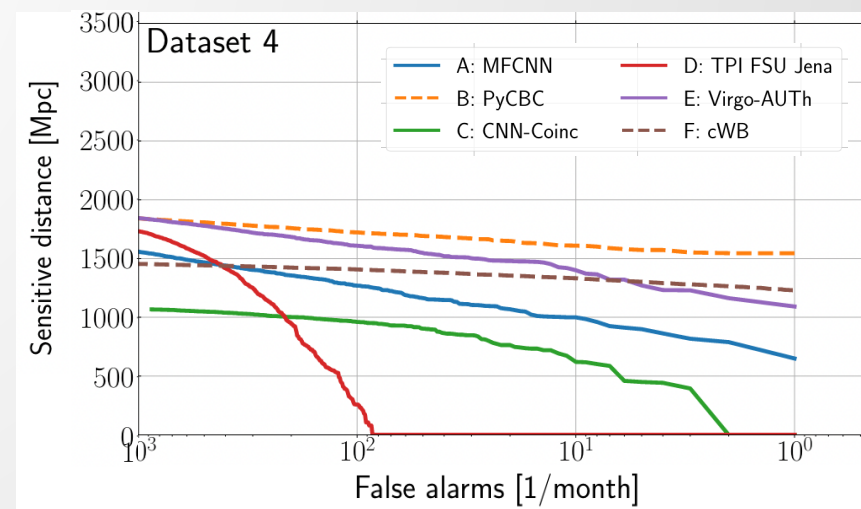
## Matched-filtering Convolutional Neural Network (MFCNN)



GW150914 的实时信号搜寻



GWTC-1/2 的参数空间分布



(MFCNN group) Wang H, et al. PRD (2023)

# 引力波天文学与人工智能：数据降噪



- 引力波智能降噪与数据分析：
  - 引力波观测数据降噪模型：WaveFormer
  - 引力波暂现源事件信号搜寻

He Wang, et al. *Intelligent noise suppression for gravitational wave observational data*, 26 February 2023, PREPRINT (Version 1) available at Research Square.

DOI: [10.21203/rs.3.rs-2452860/v1](https://doi.org/10.21203/rs.3.rs-2452860/v1)

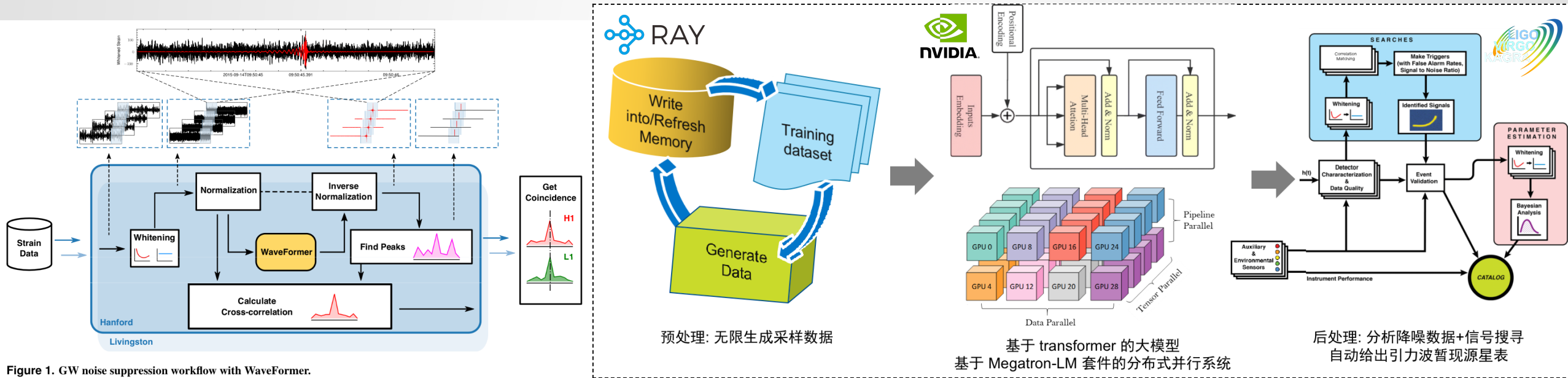
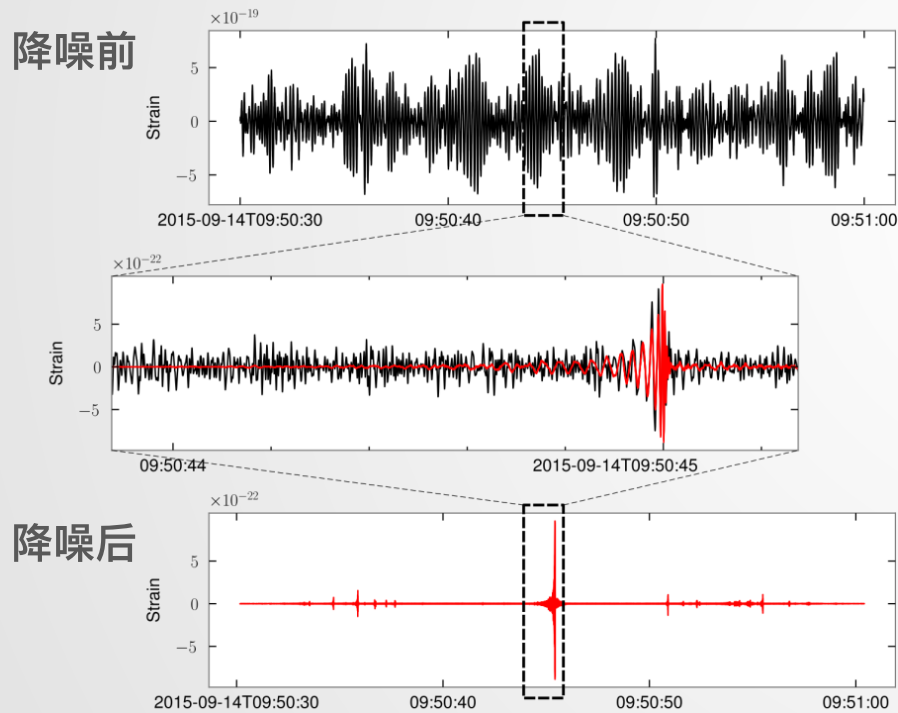


Figure 1. GW noise suppression workflow with WaveFormer.

# 引力波天文学与人工智能：数据降噪



- 引力波智能降噪与数据分析：
  - 引力波观测数据降噪模型：WaveFormer
  - 引力波暂现源事件信号搜寻



He Wang, et al. *Intelligent noise suppression for gravitational wave observational data*, 26 February 2023, PREPRINT (Version 1) available at Research Square.  
DOI: [10.21203/rs.3.rs-2452860/v1](https://doi.org/10.21203/rs.3.rs-2452860/v1)

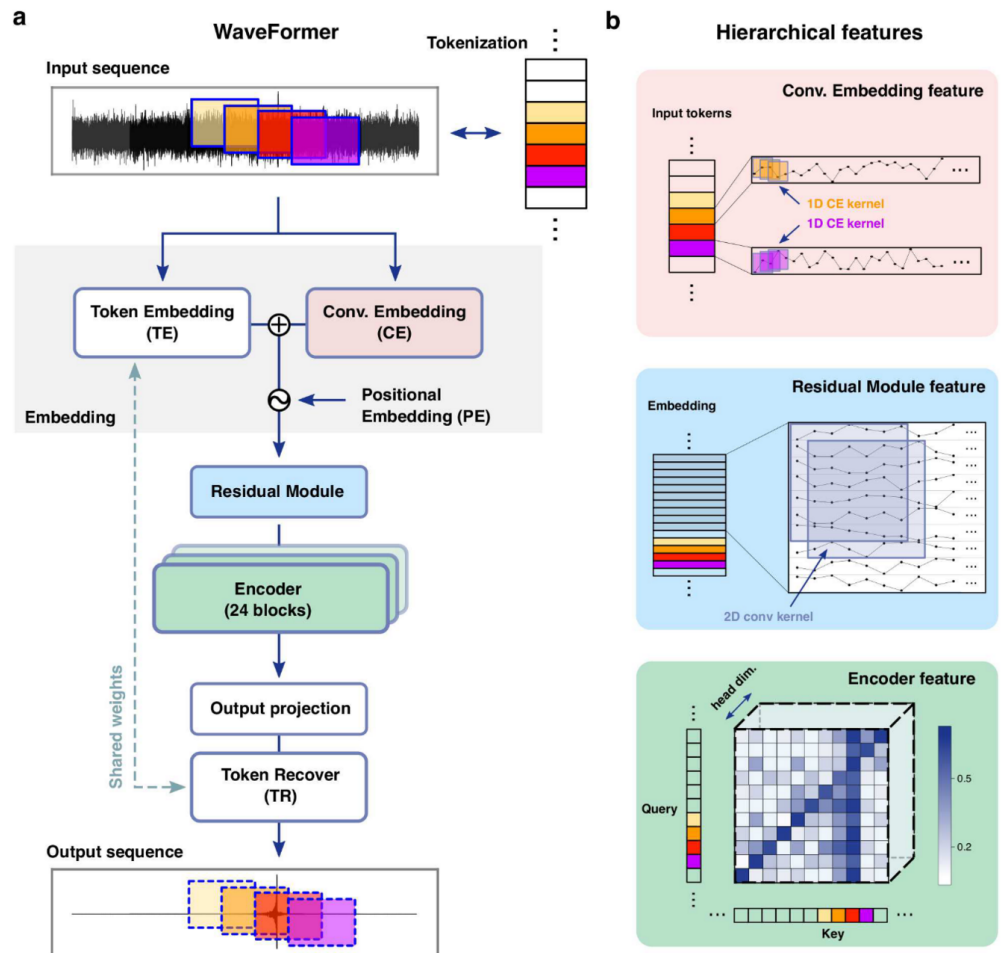


Figure 2. Science-driven WaveFormer architecture design and hierarchical feature extractions.

# 引力波天文学与人工智能：数据降噪



- 引力波智能降噪与数据分析：
  - 引力波观测数据降噪模型：WaveFormer
  - 引力波暂现源事件信号搜寻
- Billion-scale 大模型对全频段的引力波探测数据实现噪声去除和引力波信号的波形重构
- Highlights：
  - 抑制观测噪声幅度 (包括Glitch) 1-2 个数量级以上
  - 对引力波信号幅度和相位信息精确重构

He Wang, et al. *Intelligent noise suppression for gravitational wave observational data*, 26 February 2023, PREPRINT (Version 1) available at Research Square.

DOI: [10.21203/rs.3.rs-2452860/v1](https://doi.org/10.21203/rs.3.rs-2452860/v1)

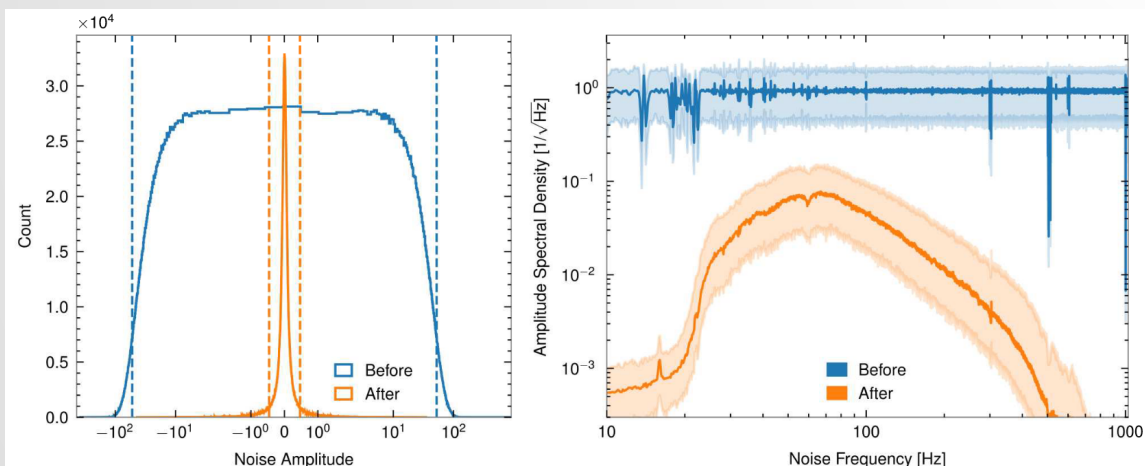


Fig. 3 Noise suppression performance of pure noise

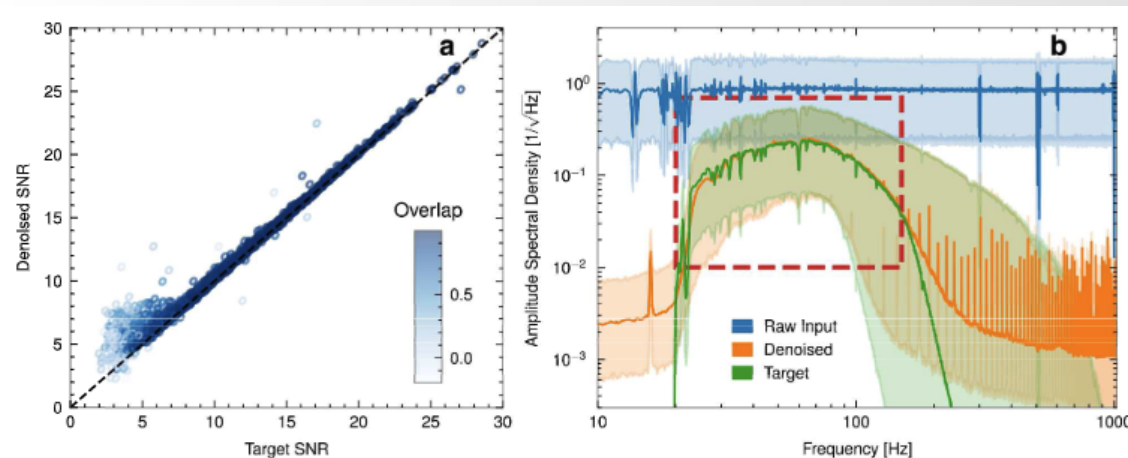
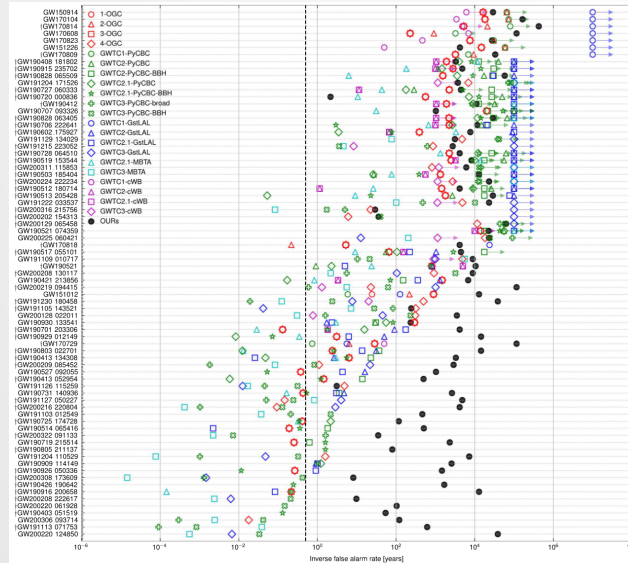


Fig. 4 Signal recovery results of simulated signals (Hanford O3b). **a.** Amplitude recovery performance. The closer scatter points are to the black dashed line, the better amplitudes are recovered. For most waveforms, its corresponding scatter point is quite close to the line. **b.** ASD distribution after noise suppression. Among the frequency region (red dashed rectangle box) that contains richest signal information, ASD of our denoised output is significantly consistent with that of target, which further indicates the great amplitude recovery performance.

# 引力波天文学与人工智能：数据降噪



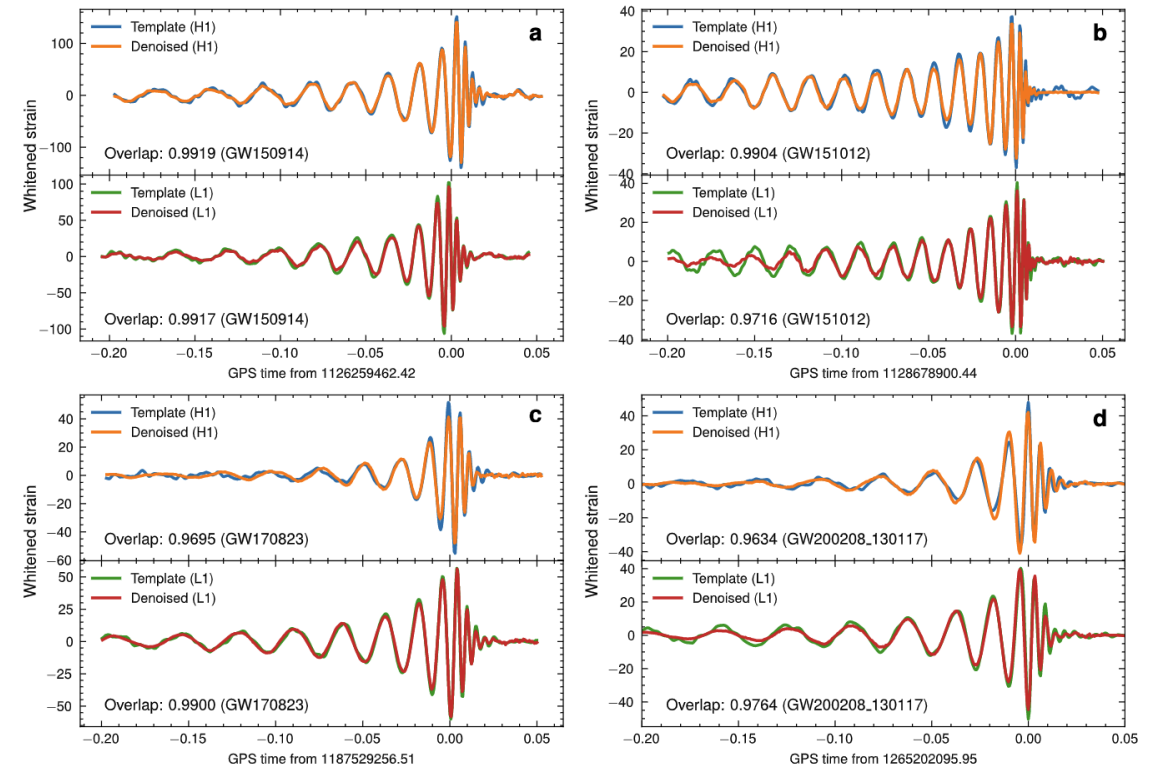
- 引力波智能降噪与数据分析：
  - 引力波观测数据降噪模型：WaveFormer
  - 引力波暂现源事件信号搜寻
- Billion-scale 大模型对全频段的引力波探测数据实现噪声去除和引力波信号的波形重构
- Highlights:
  - 抑制观测噪声幅度 (包括Glitch) 1-2 个数量级以上
  - 对引力波信号幅度和相位信息精确重构



**Figure 6.** IFAR for LIGO BBH events. The BBH events are collected from the first, second, and third Gravitational-Wave Transient Catalogs (GWTC-1/2/2.1/3) [6–9] and sorted by FAR (from low to high). The events marked with a dagger (†) contain Virgo data. The point with an arrow attached represents the minimum IFAR that can be achieved by the pipelines from the reported catalogs. The black dotted line represents a  $2.0\text{yr}^{-1}$  FAR threshold used in GWTC-2/2.1/3 catalogs. For BBH events from O3b (high-FAR events on bottom of the figure), we achieve 1386 years IFAR improvement on average.

He Wang, et al. *Intelligent noise suppression for gravitational wave observational data*, 26 February 2023, PREPRINT (Version 1) available at Research Square.

DOI: [10.21203/rs.3.rs-2452860/v1](https://doi.org/10.21203/rs.3.rs-2452860/v1)



**Fig. 5** Comparison of denoised signals from LIGO observation data for events: a. GW150914, b. GW151012, c. GW170823 and d. GW200208.130117 with their optimal templates. Hanford and Livingston are represented by H1 and L1. For most events, we achieved perfect phase information recovery; the ringdown part of GW150914, in particular, can be recovered very well. And no matter under what circumstances, like low network SNR (GW151012, GW200208.130117), or a high chirp mass system (GW170823), the amplitudes of our denoised signal both match those of the templates.

# 未来展望与规划

- 能够快速准确地获取高质量、高精度且可靠的引力波科学数据都是至关重要的
- 我国引力波数据分析基础薄弱，基础实施还不完善
- 引力波数据的现状、需求与挑战：

- 数据量庞大：>17.2TB
- 技术门槛较高：CernVM-FS (1-5MB/s, 42-209 days)
- 存储数据的平台非常多且散乱 (GitHub / GitLab / Zenodo / LIGO DCC / ...)
- 数据格式缺乏统一的标准和规范，普遍缺乏说明性引导和可视化预览

- NADC 青年数据科学家项目：

- 以LIGO科学合作组织公开发布的引力波探测数据集和引力波科学分析结果为基础，搭建引力波开放数据库数据源，对数据的一致性进行统一规范化整理。
- 以国家天文科学数据中心在线服务平台，开发适用于引力波天文研究的引力波探测开源数据门户。

## Large Scale Computing Resources

### CernVM File System

CernVM File System (CVMFS) provides access to LIGO, Virgo, and KAGRA data and software on Linux or MacOS computers, as if they are on a local disk.

IGWN Resources on CVMFS: + IGWN Public Data: [gwosc.osgstorage.org](http://gwosc.osgstorage.org) + IGWN Software: [oasis.opensciencegrid.org](http://oasis.opensciencegrid.org)

Explore CVMFS

### Open Science Grid (OSG)

Some researchers can use the Open Science Grid (OSG) to apply High Throughput Computing to their research.

1. Create an account at [osgconnect.net](http://osgconnect.net)
2. Use the [OSG Help Desk](#) to find documentation, support, and office hours
3. Read the [IGWN Guide to HTCondor](#) to learn how to manage jobs

OSG Connect may only be used for work relevant to research and/or education efforts associated with an academic, government, or non-profit institution in the United States.

	S5	S6	O1	O2	O3a	O3b	Total
占用空间(TB)	4.6	1.6	1.4	3.2	3.5	2.9	17.2
数据块总数	12,726	37,173	3,526	7,990	9,618	7,933	78,966

表 2 已公开的引力波观测数据存储汇总。

	GWTC	1-OGC	2-OGC	3-OGC	4-OGC
引力波信号总数	29,689	146,214	733,231	4,003,997	1,424,955

表 3 不同的引力波暂现源星表中疑似引力波信号总数。其中 GWTC<sup>[8,9,10,11]</sup>收集了前三次正式观测运行 O1-O3 中所有信号搜寻流水线所给出的疑似引力波信号；OGC (Open Gravitational-wave Catalog) <sup>[20,21,22,23]</sup>是马普所独立搜寻的公开暂现源星表，其中 1/2/3/4-OGC 分别对应于 O1/O2/O3a/O3b。

# 未来展望与规划

- 提升中国高能物理水平，实现重大科学目标
- 提升中国高能物理水平，实现重大科学目标
- 提升中国高能物理水平，实现重大科学目标

- 投资规模大: >17.2TB
- 运行门限值高: CERN 75 (1.5Mbit/s, 40-200 days)
- 中国高能物理水平提升
- 提升中国高能物理水平

```
for _ in range(num_of_audiences):  
    print('Thank you for your attention! 🙏')
```

This slide: <https://slides.com/iphysresearch/cvo2022>

## Large Scale Computing Resources

Resources used:  
CERN-IT-2019-001  
CERN-IT-2019-002

Resources used:  
CERN-IT-2019-003  
CERN-IT-2019-004

	100	200	300	400	500	600	Total
总运行时间	44	14	14	12	11	14	112
数据总量	1.7TB	3.7TB	5.7TB	7.7TB	9.7TB	11.7TB	43.2TB

	1000	2000	3000	4000	5000
总运行时间	2400	4800	7200	9600	12000

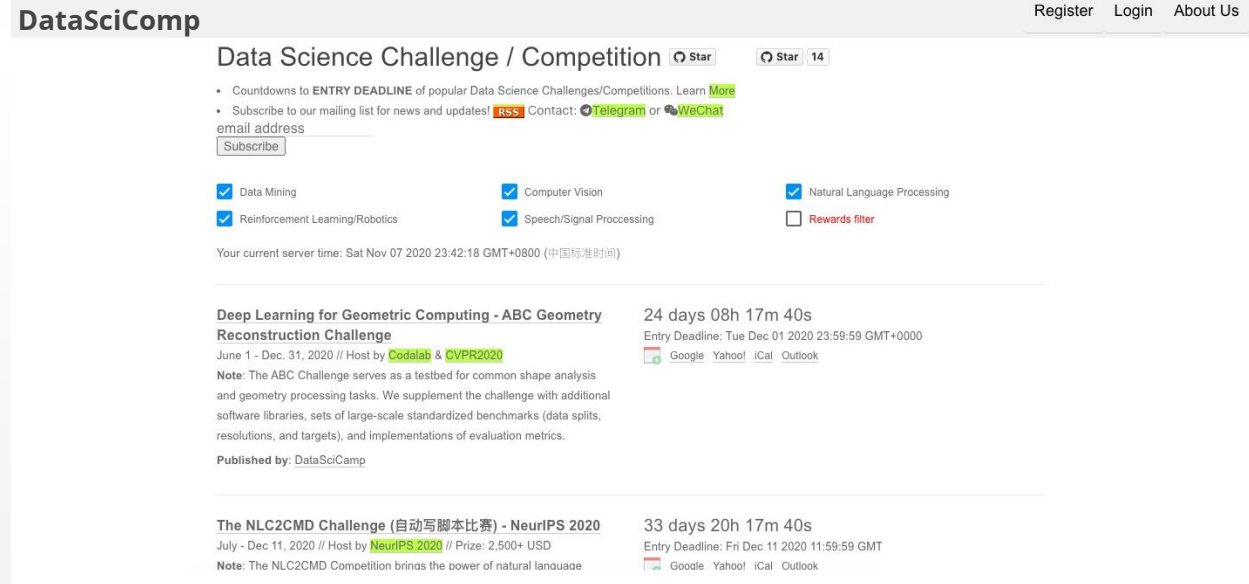
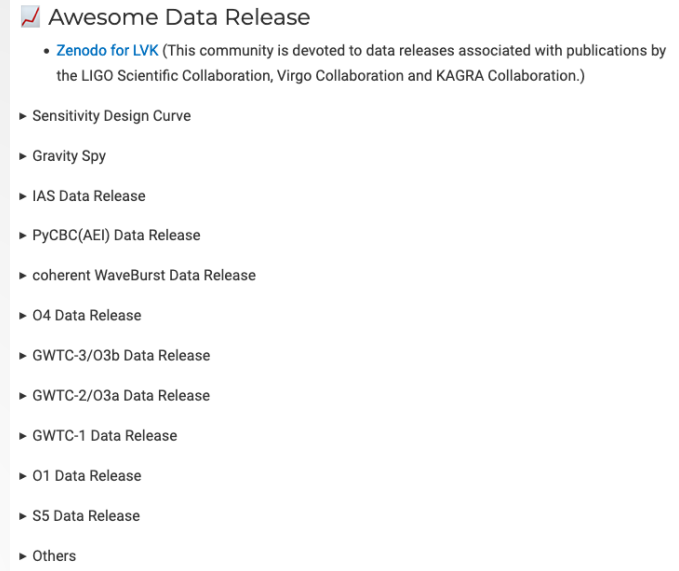
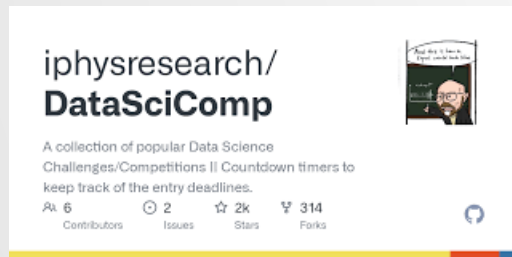
# Bonus Slides

## • 数据准备:

- 已经采集完近 17.2TB 引力波探测数据, 并收集了较为完整的各学术权威机构所发布的引力波科学数据结果

## • 技术储备:

- 曾设计并开发了DataSciComp门户网站
- NoSQL非关系型的数据库
- 前端Flask架构和Docker容器化技术
- 后端架构是基于微服务 (Microservice Architecture) 技术



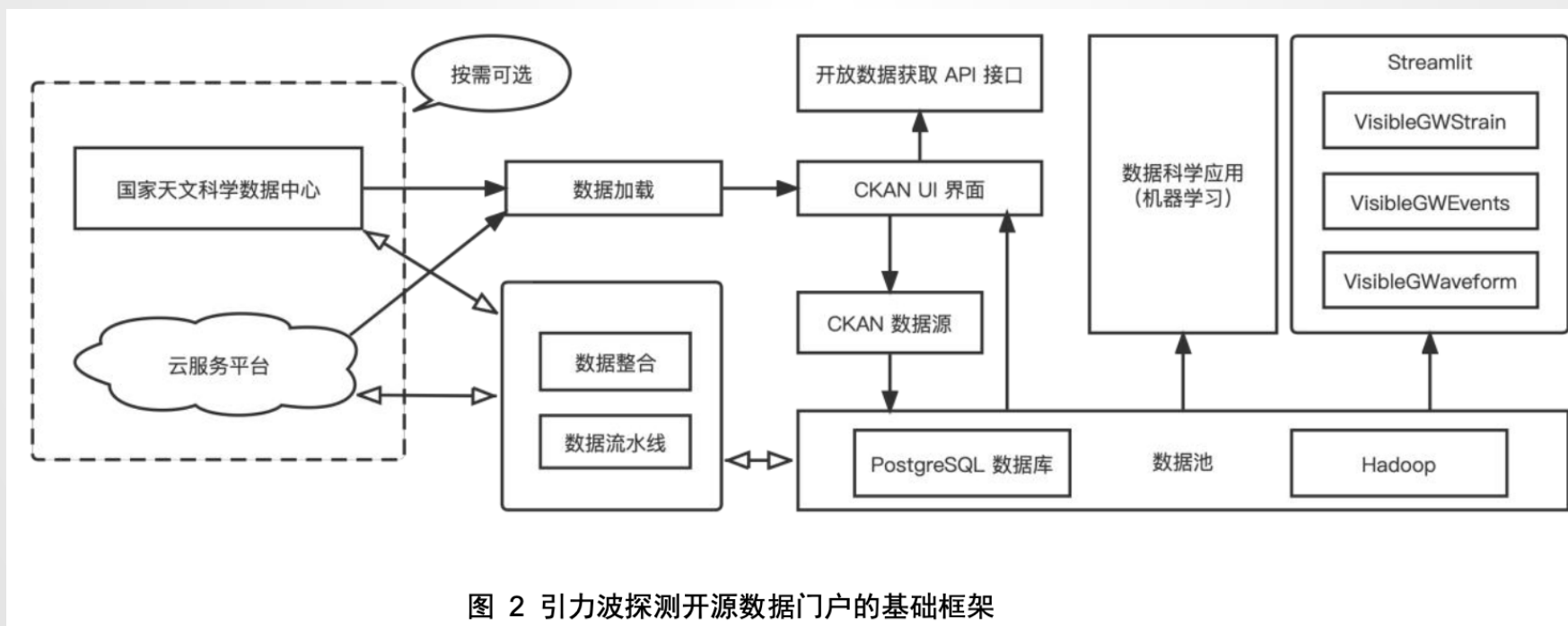
## • Demo 验证

- 可交互式数据可视化 VisibleGWevents
- <https://iphysresearch-visiblegwevents-test-firebase-4v0f1o.streamlit.app/>



# Bonus Slides

- **CKAN** (开源数据管理系统): an open-source DMS (data management system) for powering data hubs and data portals.
- **Streamlit**: 一个基于 Python 的 Web 应用程序框架，致力于以更高效、更灵活的方式可视化数据，并分析结果。
- **VisibleGWStrain**: 可调取并可视化O1-O3任意一段时域数据的进行观察和提取，展示数据质量等相关信息。
- **VisibleGWWaveform**: 可生成并可视化不同理论波形模板描述下的引力波波形信息（时域、频域和时频域）。
- **VisibleGWevents**: 可查询并可视化各类引力波暂现源星表（GWTC, OGC等）和对应波源参数反演的分析结果。



# Bonus Slides

- 预期研究成果

1. 引力波探测数据治理手册
2. 引力波探测开源数据门户
  - 开放数据获取 API 接口
  - 可交互式数据可视化应用
    - 引力波观测数据 (VisibleGWStrain)
    - 引力波模板波形 (VisibleGWWaveform)
    - 引力波科学分析结果 (VisibleGWevents)
3. 部分核心代码将会在开源平台Zenodo上以数字出版物的形式发布，并撰写软著。

- 潜在价值

- 为我国引力波天文学数据处理领域的科研人员带来便利
- 作为中国科学院“空间太极计划”科学数据平台的技术储备
- 教学辅助和学术参考价值
- 科普教育意义

<https://data.cardiffgravity.org/waveform-fitter/>

# Bonus Slides

- 预期研究成果

1. 引力波探测数据治理手册
2. 引力波探测开源数据门户
  - 开放数据获取 API 接口
  - 可交互式数据可视化应用
    - 引力波观测数据 (VisibleGWStrain)
    - 引力波模板波形 (VisibleGWWaveform)
    - 引力波科学分析结果 (VisibleGWevents)
3. 部分核心代码将会在开源平台Zenodo上以数字出版物的形式发布，并撰写软著。

- 潜在价值

- 为我国引力波天文学数据处理领域的科研人员带来便利
- 作为中国科学院“空间太极计划”科学数据平台的技术储备
- 教学辅助和学术参考价值
- 科普教育意义

<https://catalog.cardiffgravity.org/>