

doi: 10.3969/j.issn.1000-8349.2016.01.01

# 引力波数据分析

王小鸽<sup>1</sup>, ERIC Lebigot<sup>1</sup>, 都志辉<sup>1</sup>, 曹军威<sup>1</sup>,  
王运永<sup>2</sup>, 张帆<sup>2</sup>, 蔡永志<sup>2</sup>, 李木子<sup>2</sup>, 朱宗宏<sup>2</sup>,  
钱进<sup>3</sup>, 殷聪<sup>3</sup>, 王建波<sup>3</sup>, 赵文<sup>4</sup>, 张扬<sup>4</sup>,  
DAVID Blair<sup>5</sup>, JU Li<sup>5</sup>, ZHAO Chun-Nong<sup>5</sup>, WEN Lin-Qing<sup>5</sup>

(1. 清华大学, 北京 100084; 2. 北京师范大学 天文系, 北京 100875; 3. 中国计量科学研究院, 北京 100013; 4. 中国科学技术大学, 合肥 230026; 5. University of Western Australia, WA 6009, Australia)

**摘要:** 以引力波探测为基础的引力波天文学是一门正在崛起的新兴交叉学科, 它是继以电磁辐射为探测手段的传统天文学之后, 人类观测宇宙的一个新窗口, 对研究宇宙的起源和演化, 拓展天文学的研究领域都有极其重要的意义。激光干涉引力波探测器的出现, 更开辟了引力波探测的新纪元。引力波数据处理与分析已在世界各地迅速发展起来, 为引力波天文学的研究提供了锐利的武器。文中系统地介绍了引力波数据分析中常用的工具软件, 详细讨论了时间-频率分析、复合分析法、脉冲星计时分析法、匹配过滤器、模板、 $\chi^2$  检验、蒙特卡罗模拟等引力波数据分析中使用的基本方法。

**关键词:** 引力波; 激光; 干涉仪; 数据分析

**中图分类号:** P159.3 **文献标识码:** A

## 1 引 言

引力波是广义相对论最重要的预言<sup>[1, 2]</sup>, 引力波探测是当代物理学重要的前沿领域之一。以引力波探测为基础的引力波天文学<sup>[3, 4]</sup> 是一门新兴的交叉学科, 由于引力辐射独特的物理机制和特性, 使得引力波天文学研究范围更广、更全面。它以全新的手段和理念探寻宇宙中未知的质量体系, 提供其他天文观测方法不可能获得的信息, 加深人们对宇宙中天体结构的认识。继以电磁辐射 (如可见光、红外线、紫外线、X 射线、伽玛射线和射电) 为探测手段的

收稿日期: 2015-08-25; 修回日期: 2015-10-08

资助项目: 北京师范大学科学研究基金; 中央高校基础研究基金; 国家自然科学基金 (11373014, 11073005); 中国科学院战略先导项目 (XDB09000000); 973 项目 (2012CB821804, 2014CB845806)

通讯作者: 王运永, wangyunyong@bnu.edu.cn

传统天文学之后,引力波探测成为人类观测宇宙的一个新窗口。它对研究宇宙的起源和演化,拓展天文学的研究领域都有极其重要的意义。

由于引力波信号非常弱,引力波天文学研究不但需要高灵敏度、宽频带的探测器,以及高超的探测技术,也需要科学的数据处理和分析方法。当前,引力波探测器的主流设备——激光干涉仪引力波探测器已在世界各地蓬勃发展起来,以 Advanced LIGO, Advanced Virgo, KAGRA 和 GEO-HF 为代表的第二代激光干涉仪引力波探测器正处于紧张的安装调试阶段,不久即可投入运转,获取数据。与此同时,引力波数据处理、分析方法及相应的通用软件的研发也与第二代干涉仪的建造同步进行,并取得了巨大进展。不同特色、不同研究方向、不同研究课题的专用和通用软硬件也在世界各大实验室迅速建立起来,为引力波探测和引力波天文学的研究打下了坚实的基础。

## 2 数据分析通用工具

为了更好地进行数据分析,物理学家需要建立大量通用的软件系统。如工具软件、模拟计算软件、存放刻度常数修正参数和探测器参数的数据库、管理和控制作业进度的软件系统、事例筛选和重建系统、针对专门研究领域的分析流水线等。在做数据分析之前,必须对它们的功能、格式、内容、操作方法等进行深入透彻的学习和理解,才能得心应手,达到事半功倍的效果。

一般来说,每个实验室都有自己的一整套完善的软件体系,其中每个程序包都含有百万行以上的源程序。这些程序需要专门的人员认真编写、维护、改进和升级。以 LIGO 为例,它有三个专门小组,负责所有软件工作,它们是:(1) LIGO 数据分析系统 LDAS (LIGO Data Analysis System) 组;(2) 模型化和模拟计算 (Modeling and Simulation) 组;(3) 通用计算 (General Computing) 组。各大实验室的软件系统都是数十年来几代人辛勤劳动的结晶,在国际上有很大的通用性,可以互相借鉴和移植。

### 2.1 数据获取硬件系统

激光干涉仪是引力波探测器的主要部分,是光学仪器。数据获取首先要把光信号抽取出来,转换成电信号并进行数字化。这个获取数字信号的过程包括的硬件设备有:光路、光电转换器件、读出电子学 (readout electronics) 和在线控制插件等部分。光路由平面镜和透镜组成,它将需要的光信号抽取出来,引导到需要的方向和地点并汇聚在光探测器上。光探测器将光信号转换成电信号,电信号经过读出电子学放大、甄别和成形之后,进行模数转换。然后通过在线计算机控制的部件记录在存储设备上。硬件系统还包括标定系统,定期对电子系统进行定标,以确保数据的精度。除了干涉仪本身的数据之外,在引力波探测中还使用了众多的各种独立地对探测器的环境和探测器状态进行监控的设备和记录仪,对诸如温度、气压、风力、大雨、冰雹、地表震动、声响、电场、磁场等环境条件进行监测,以及对引力波探测器内部的平面镜和透镜的位置等探测器自身状态进行监测。这些监控探测器的信号也都转化成数字信号,以数据的形式纪录和保存,供后续和在线的数据分析时使用。不论是激光

干涉仪上获取的数据, 还是物理环境的监测设备所纪录的环境数据, 都需要统一的时间标志以便使所有探测器同步, 表明数据所表示的信号的发生时间。有这个时间标志, 也可以对坐落在世界各地的多个探测器在同一时间段探测到的信号进行关联分析, 获得更多的关于引力波的准确信息。数据获取系统的这个全球同步的时钟 (时间标志) 和时间发布设备也是数据获取的重要设备之一。

## 2.2 数据结构

从引力波探测器获取的原始数据都是一些数字, 不能直接拿来作物理分析, 要进行基本的整理加工, 如: 去掉无用的垃圾, 用标准信号对探测器及电子学进行刻度, 对特殊信息进行标示等。也就是说在进行实质性的物理分析之前, 要先做大量的、繁杂的、重复性的工作, 而且每个做物理分析的人都要做几乎相同的事情, 费时费力, 浪费资源。在世界各个实验室都建立了专门的小组承担这项任务, 按照统一的要求和格式把数据组织起来, 加入一定的标识符, 形成数据文件。该文件除了探测器获取的信息外, 还包含后续分析所需要的足够丰富的资料。这种把原始数据转变为直接用来作物理分析的数据文件的过程是数据的离线分析所需要的。以 LIGO 于 2014 年 8 月公开发布的 S5 科学数据为例, 数据文件包含以下几方面的附加信息:

第一, 提供了数据的获取时间, 即时间标签。以 timeline 的形式, 标识出哪段时间里 LIGO 探测器的数据是可以被用户利用的。同时还给出在给定时间段里可以使用的数据元的百分比; 这个百分比越高, 表明这段时间里探测器的状态越稳定, 可靠数据的比例越大。

第二, 包含对探测器的标识信息, 以表明数据来自哪个探测器。LIGO 目前的 S5 数据含有来自 5 个不同探测器的观测数据, 分别标识为 H1, H2, L1, V1, 和 G1。

第三, 为了科学研究, LIGO 在数据获取时, 有针对性地加入一些在理论上具有已知特征的引力波, 用以考察和测试开发的分析软件和算法是否能捕捉到这类事例。为了把这类人工植入的数据事件与真实采集到的信号加以区别, LIGO 在数据文件中对这类事件做了标识。

第四, 激光干涉仪和其他环境监测设备在运行中会有不稳定的瞬时故障, 这些仪器的小故障会带来一些类似引力波产生的信号事件 (instrumental glitches), 这类事件不是来源于天体活动, 而是来源于探测器设备或周围环境的变化。它们虽然是虚假触发, 但可能与天体活动产生的事件非常相似。通过适当的分析手段 (称为 veto study), 这些事例能够被判断为虚假事例, 它们需要在数据中加以标识, 以减少对后续数据分析过程的干扰。

第五, 数据的取样率。取样率与数据获取设备的处理能力和速度相关的, 也是数据的重要特征信息。

最后, 还有数据质量分类信息。在数据采集过程中, 对设备和系统运行状态的监控能给出当前时刻设备的状态是否稳定, 以及环境干扰成分是否会影响到数据的质量这些信息。如果设备由于某种原因而出现不稳定的情况, 那么得到的数据就是不可靠的; 如果监控发现某个时刻的环境干扰过大, 那么得到的数据的可靠性也需要质疑。于是, 采用了打数据质量标签的方法。LIGO 的数据有 CAT1 和 CAT2 两类标签, 用于标定某个时间段上数据的质量。不达到 CAT1 分类的数据不能用于搜索天体活动事件; 没有同时达到 CAT1 和 CAT2 质量

标准的数据, 分析结果会被否决掉。

以上这些信息以不同的格式表示和存储, 在访问时通过数据文件中给出的元数据所提供的对数据的描述, 来获取相应的数据结构信息。LIGO 提供了获取这些信息的工具。以 LIGO 数据中标识为例, 在 LIGO 发布的 S5 数据中, 用一位 2 进制数作为一个标识。当该位的值为 0 时, 表示否定; 为 1 时, 表示肯定。用 18 位标识数据质量, 用 6 位标识植入数据信息。

### 2.3 蒙特卡罗模拟

蒙特卡罗方法广义而简单地讲就是根据随机采样的观测和一些已知的关系, 反演出某个无法直接获取到的自然现象的模型或参数。这个方法在许多科学领域有广泛的应用, 特别是在一些无法直接获取到确定参数值, 而观测数据却相对容易得到的问题上, 采用蒙特卡罗方法往往是唯一的选择。例如计算数值积分问题, 当判断积分空间上随机的点与被积函数的关系很容易判断, 而被积函数又没有显式表达式的时候, 采用蒙特卡罗方法, 根据观测的随机点的分布来推算出积分值, 这是一个很有效的方法。

除此之外, 蒙特卡罗方法与马尔可夫链方法结合, 形成的马尔可夫链-蒙特卡罗方法 (简称为 MCMC 方法), 是一个求解反向推演问题 (inverse problem) 非常有效的方法。可以极大地加快计算的收敛速度。根据观测或最大似然值在参数空间中寻找最佳参数的问题, 采用马尔可夫链-蒙特卡罗方法求解是运用得很普遍有效的方法。例如, 利用爱因斯坦广义相对论和天体运动物理模型, 以及对噪音的建模, 可以仿真出天体运动产生的引力波在噪音背景上叠加所形成的、可以被激光干涉仪捕捉到的波形。由于模型的参数较多, 不同的参数得到波形会不同。对于一个探测到的事件, 要判断它是否是由某一类天体活动如中子星或黑洞双旋产生的, 也可以看成是一个反向推演问题。即根据探测的引力波数据, 求出引力波的发生源问题, 这种方法在引力波数据分析领域里也称为模型参数估计。

一类比较突出的成果是采用贝叶斯推断的框架来分析的。这个方法的基本思想是根据模型和观测数据, 推算出后验概率密度函数。但是贝叶斯推断方法的计算量大是一个突出的问题, 以双星旋绕系统来说, 用广义相对论模型描述两点质量中心以圆形轨道旋绕的系统产生的引力波波形模型有 15 个参数, 加上中子星或黑洞的模型参数、噪音的模型参数, 仪器的校准参数等, 就显得更为繁多。除了参数多这个因素之外, 似然函数的复杂性和波形生成过程也导致了巨大的计算量。近年来, 随机取样及马尔可夫链-蒙特卡罗方法<sup>[5-13]</sup>等技术的应用, 在提高计算贝叶斯推断问题的速度和效率方面取得了一些成果。

引力波源的参数估计算法, 以 LAL Inference 软件包实现的方法为例, 需要计算出约  $10^7 \sim 10^8$  个波形<sup>[14]</sup>, 用于与干涉仪的数据进行比较。这些波形的生成需巨大的计算, 以至成为参数估计分析的瓶颈。一旦干涉仪可探测的有效频域进一步扩大, 那么波形的长度就会随之增长, 参数估计分析需要的时间也就随之增加, 达到不可接受的程度。通过构造一个平衡分布, 马尔可夫-蒙特卡罗方法可获得与要计算的后验概率分布成比例的马尔可夫链, 产生的样本的概率密度分布也与目标后验概率密度分布成比例, 可以有效地减少采样的数量, 从而加快后验概率 PE (Posterior Estimation) 分析的速度。如何进一步加快 MCMC 算法的速度, 参考文献[14] 中讨论了若干方法。

### 2.4 数据通道和数据单元

激光干涉仪引力波探测器可以看成是一个准线性装置, 输入到干涉仪的信号和干涉仪的输出信号可形象地用图 1 表示。

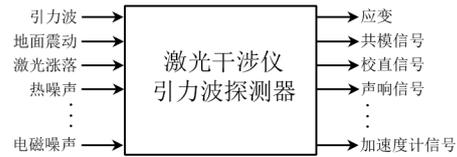
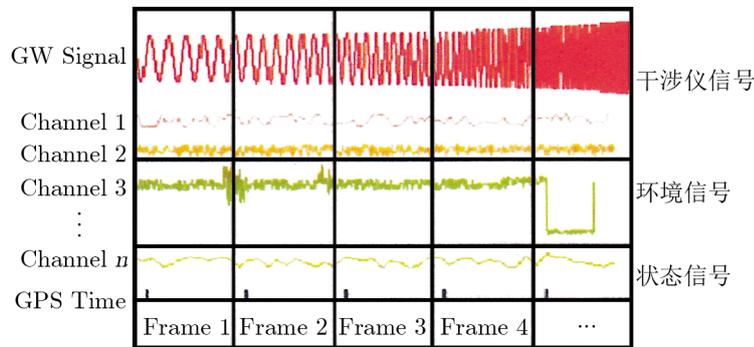


图 1 激光干涉仪引力波探测器输入输出信号关系

激光干涉仪引力波探测器输出各种不同类型的信号, 这种输出信号统称为数据。在干涉仪的不同部位获取的数据及不同的环境监视器上获取的不同类型的数据, 要分别给予不同的名称, 称之为数据通道。激光干涉仪引力波探测器及辅助设备的数据通道共有数百个, 分别具有不同的物理或技术含义。相同数据通道的数据存储在一起, 形成数据文件。激光干涉仪引力波探测器开动之后, 要连续运转相当长的时间, 如数日、数周甚至数月, 各个数据通道的数据源源不断地输出。为了便于记录和进行物理分析, 需要把数据流切割成小段, 以这个小数据段为单位进行记录, 这种小数据段被称为数据元。它是数据的基本存储单位。每个数据元都有自己的编号(即名称), 相邻数据元的编号是连续的, 它们在时间上形成一个连续序列。储存在数据元内的主要数据是在选定的时间间隔内的时间序列, 然而, 也可以在数据元内放其他类型的数据, 如: 谱形、表格、矢量或阵列等。为了便于世界各地的引力波探测器联网, 以进行符合测量, 国际上对数据元的格式和结构有统一的规定。数据通道与数据元的关系如图 2 所示。



注: 图中左侧给出的是数据通道及编号, 右侧给出的是数据名称。底部显示的是数据元。

图 2 数据道与数据元的关系<sup>[17]</sup>

激光干涉仪引力波探测器正常工作时, 满足一定选择条件的数据元如图 3 所示。图中展示的是 LIGO 数据通道 LI: LSC-AS-Q 中标号为 asq730885392 的一个数据元。它的横坐标是时间, 纵坐标是模数转换器 ADC 的计数。图中的尖锐突起是我们要选出来的。

## 3 数据分析方法

### 3.1 波形分析

激光干涉仪引力波探测器是一个宽频带幅度探测器, 它探测的是波形。实验物理学家在

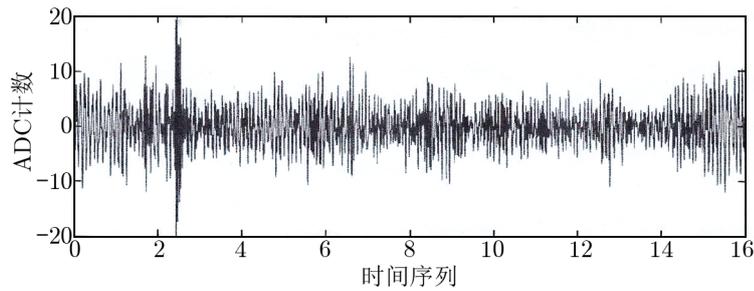


图3 满足选择条件的数据元的图形

引力波探测中不用“天体物理源”的概念来思考问题，而使用“波形结构形态学”，不同的天体物理源辐射出不同的波形。波形分析是引力波数据分析的基础。在引力波数据分析中，波形结构形态学研究的对象主要有以下几个方面。

(1) 具有有限时间间隔的连续性(或爆发性)引力波

我们自认为已经建立了合理的物理模型，其波形也认为是已经知道的。如双中子星(NS-NS)互绕、黑洞-黑洞(BH-BH)旋绕、中子星-黑洞(NS-BH)旋绕所辐射的Chirp形信号和余音信号等。

(2) 未知的脉冲形引力波

这是另一种类型的爆发性引力波，我们没有可靠、合理的物理模型来描述它们，其波形也是不知道的，如超新星爆发、黑洞坍缩等。这种类型的引力波波形是根据研究者个人的理解建立的。

(3) 窄频带、连续的、周期性引力波

我们有可靠、合理的物理模型，其波形被认为是相当确定，如具有一定椭度的脉冲星、不稳定旋转的中子星等。用波形分析法对这类事例进行分析，是快速进入引力波数据分析的捷径。

(4) 随机背景辐射

随机背景辐射是宽频带、连续性的引力波，它是随机的，不易与噪声区别。既包括宇宙大爆炸产生的引力波的残余本底，也包括各种不同的、连续性波形的叠加，这类引力波频带很宽，用波形分析法处理时要克服很多困难。

(5) 未知的引力波波型

还有很多使我们感到惊奇的波形，其波源及辐射过程都是我们始料未及的。这很可能是一种新的天文现象，需要引起高度关注。

### 3.2 时间-频率分析

时间-频率分析法是数据处理中最常见的方法之一<sup>[15, 16]</sup>，它用时间-频率分解法，把已经记录到数据文件中的各个数据元做傅立叶分解。并利用计算机把相应的频率、时间和功率画成三维图，将自己确定的频率间隔内存在的功率与已知的噪声功率统计分布进行比较，寻找与一般噪声信号不同的东西。还可以根据自己建立的条件进行判选，找出感兴趣的新奇事例。在引力波探测中，大量事件的波形是未知的，或者说，在足够精确度下是未知的，如黑

洞并合及质量巨大的星体核的坍缩等, 它们不能用匹配过滤器进行分析。在这种情况下, 时间—频率分析法给我们提供了一种方法, 它不必预先知道波形的相关知识, 而是直接去寻找感兴趣的引力波事例。时间—频率分析法中, 频率、时间、功率三维图在图 4 中给出。

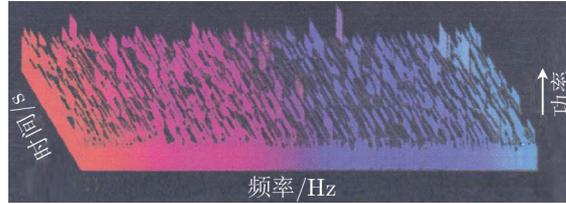


图 4 频率、时间、功率三维图频率—时间特点图<sup>[17]</sup>

时间—频率分析法的基本思路是: 在固定的时间间隔内把探测到的数据切割成片, 对这些数据切片进行傅立叶展开。把给定的频率间隔内存在的功率与已知的噪声功率的统计分布进行比较, 寻找引力波存在的证据。在这种分析中, 给定的频率间隔是时间的函数。

该方法的分析手段很多, 常用的有超功率法和 TF Cluster 分析法。

### 3.2.1 超功率法

将选定频带、选定时间段内数据中存在的功率, 与已知的噪声功率的统计分布进行比较, 如果探测器的噪声是稳定的且具有高斯分布, 那么噪声功率将遵循  $\chi^2$  分布, 其自由度数量等于时间—频率值的两倍。因此, 这种方法在探测与引力波强度有关的超阈功率时, 它的效率取决于信号的期待持续时间和频带宽度, 也与它的强度有关。

### 3.2.2 TF Cluster 方法

该方法是在时间—频率平面内寻找功率高于给定阈值的小区域组成的单元集团。因为大多数信号都分布在一些小区域内, 它们在时间—频率平面内有很高的空间关联性。这种集团化分析法对滤掉探测器噪声是很有用的。

举例来说, 利用蒙特卡罗算法模拟在白色高斯噪声本底下, 双星旋绕产生的 Chirp 信号, 得到一批数据, 利用这些模拟数据进行频率—时间分析, 可以对以上的讨论进行检验。我们选定的条件为:

- (1) 在时间—频率分解数据中对小单元的功率设置阈值;
- (2) 只保留超过功率阈值的小单元相互连接而构成的集团;
- (3) 对集团的总功率设置阈值后获取相应结果。

### 3.2.3 时间尺度选择

激光干涉仪信号频率变换分析的特定时间尺度可以根据对信号的预期为基础来选择。引力波信号预计会在不同的时间尺度上有所变化, 例如, 对双中子星或双黑洞来说, 两个星体在相对低的频率上彼此在轨道上旋转很长一段时间。而在双星合并之前, 他们的旋转频率在很短的时间迅速增加 (对于致密双星的并合, 这段时间通常为秒的量级)。因此, 在进行频率变换时, 不同的时间片要采用不同的时间尺度。例如, 一个用于爆发性引力波脉冲序列的数据分析流水线, 需要计算多个时间—频率变换, 每个变换在不同的时间尺度上进行 (每个时

间尺度都基本上等价于原来的信号),然后把在不同尺度上得到的大时频系数组合在一起,形成在时频域内的群集。如果在位于不同地域的多个独立的引力波探测器之间进行关联分析,则需要在关联性地组合所有的变换之前,为每个独立的引力波探测器做这样的分析。所谓关联性地组合所有变化,是指在组合时,要把信号的相位考虑在内,而不是仅组合不同探测器的时频系数的幅度。

时间尺度选择的另一个方法是采用小波变换进行单个多尺度变换。这样的变换基本上等同于以指数的尺度在不同时间尺度上做频率分析。通过在 LIGO 激光干涉仪<sup>[19]</sup>的环境探测器上发现超常的能量来探测不寻常的事件,如被 Kleine Welle 事件触发生成器或被数据分析流水线 Omega 探测出来的不寻常事件就是使用这类方法的例子<sup>[19]</sup>。在这些情况下,小波系数相对于常用的监测器噪声来说是非常之大的,它会引发一个包含所有环境辅助监测器读数信息的事件,其产生的测量向量可以用于差错 (glitch) 检验 (参阅 3.4 节)。

### 3.2.4 白化

在对激光干涉仪引力波探测器的输出信号进行时间-频率分析时,噪声问题一般是通过白化过程来处理。事实上,每个干涉仪信号噪声的时间-频率变换呈现出的曲线,是随着所分析的频率而变化的。探测器输出信号的表现异常(无论是由于引力波,或更可能的是由于毛刺引起的<sup>[20]</sup>)会由于信号幅度分布远高于噪声曲线而被观测到。由此,对异常事件探测分析的全过程包含了用噪声曲线来划分干涉仪信号的时频变换这样一个步骤。正是因为这个步骤在效果上把噪声曲线抹平了,所以这个步骤被称为白化。这个白化的信号与白噪声类似,在所有的频率上都有固定的能量,即其能量是个常数。信号白化以后,寻找包含超高能量的频率就可以通过简单的阈值来判断了。

### 3.3 模板的应用

在引力波的理论研究中,有些研究对象的结构和作用机制是清楚的,物理图像是直观的,如密近双星旋转问题;它们辐射的引力波可以用一个确定的时间函数,如  $S(t)$  来表示。有些研究对象在理论上研究得并不透彻,对其结构和物理机制了解得不太清楚,但是可以根据已有的知识建立起一个恰当的物理模型,推导出所辐射的引力波的时间函数  $S(t)$  和图像。

无论哪种情况,我们都认为这个引力波函数(如果存在的话)具有已知的形式  $S(t)$ 。它是我们在理论上预言的形式,一旦碰到它,我们能认出来。我们称这些函数是要探测的引力波的“模板”<sup>[21]</sup>。在实际应用中,通常把模板写成由成百上千个分立子模板构成的一个模板组合。在物理分析过程中,我们以模板为标准,逐一筛查获取的带有噪声的数据,寻找与模板类似的图样。含有这种图样的记录其强度和式样均与噪声单独存在时不同。

可以看出,以这种方法做引力波的数据分析,实际上是一个研究从探测器上获得的时间与我们建立的模板之间的匹配关系的过程。讨论模板在物理分析中的应用原理之前,我们首先了解函数的互相关运算。

设  $S_1(t)$  和  $S_2(t)$  是两个时间函数,这两个函数的互相关运算定义为:

$$S_1 * S_2(\tau) \equiv \int_{-\infty}^{\infty} S_1(t)S_2(t + \tau)dt \quad (1)$$

互相关运算的涵义是: 对于每一个给定的时间, 先让函数  $S_2(t)$  在时间上偏移  $\tau$  后与函数  $S_1(t)$  相乘, 乘积运算针对时间范围  $[-\infty, \infty]$  内所有被记录的数据, 运算进行完之后, 再把得到的所有乘积加在一起。对于不同的  $\tau$  值, 这些乘积之和不同, 它随  $\tau$  值的变化而变化, 即  $S_1 * S_2(\tau)$  是偏移时间  $\tau$  的函数, 我们称其为时间函数  $S_1(t)$  和  $S_2(t)$  的互相关函数。求解互相关函数  $S_1 * S_2(\tau)$  的过程称为函数  $S_1(t)$  和  $S_2(t)$  的互相关运算。互相关性是表示两个函数  $S_1(t)$  和  $S_2(t)$  相互关联程度的量度。

当  $S_1(t)$  和  $S_2(t)$  是同一个函数  $S(t)$  时, 我们定义它的自相关运算函数为:

$$S * S(\tau) \equiv \int_{-\infty}^{\infty} S(t)S(t + \tau)dt \quad . \quad (2)$$

$S * S(\tau)$  是时间函数  $S(t)$  的自相关函数, 自相关函数是测量一个时间函数的两个副本之间, 在各种不同的时间偏移时函数自身相关程度的方式。显然, 当函数自身在时间上对准时, 即当  $\tau = 0$ , 它的自相关函数  $S * S(\tau)$  有最大值。若该函数  $S(t)$  是一个周期性函数, 则在多个周期中自相关函数将有多个最大值。 $S * S(\tau)$  的宽度给出函数  $S(t)$  随时间变化的快慢。

现在回到原来的讨论。设我们建立的模板为  $S(t)$ , 它是我们要寻找的信号。设从探测器上获得的时间记录为  $V(t)$ , 则它们之间的互相关函数为:

$$V * S(t) = \int_{-\infty}^{\infty} V(\tau) \cdot S(t + \tau)d\tau \quad . \quad (3)$$

上式表明, 我们要对每一个可能的探测器输出信号的出现时间  $\tau$  求值。

现在我们考虑一个无噪声的时间序列, 即探测器的输出是由真实信号自己引起的 (等效于真实信号就是模板引起的信号)。设模板函数为  $S(t)$ , 则模板自己引起的探测器输出信号为:

$$V(t) = \alpha S(t - t_0) \quad . \quad (4)$$

很明显, 当模板  $S(t)$  与探测器输出的时间记录  $V(t) = \alpha S(t - t_0)$  出现符合时, 互相关函数的值为:

$$V * S(t) = \int_{-\infty}^{\infty} \alpha S^2(t + \tau)d\tau \quad . \quad (5)$$

此值大于偏离完全符合时的值。如果模板对准了无输出信号的时间段, 此时既无信号也无噪声输出, 即  $V(t) = 0$ , 则互相关函数的值也为零。由于我们的探测器总是有噪声的, 模板的互相关函数值基本上不为零, 且互相关函数  $V(t) * S(t)$  是另一个有噪声的时间序列函数。如果探测器输出的噪声具有高斯分布:

$$P(\nu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\nu^2/2\sigma^2) \quad , \quad (6)$$

那么只要没有信号存在,  $V(t) * S(t)$  的分布也是高斯型的。如果我们足够幸运, 碰到一个足够大的信号, 则探测器的输出就与模板匹配得很好, 这时  $V(t) * S(t)$  有一个特别大的值<sup>[7]</sup>。在  $V(t) * S(t)$  的直方图中, 它会明显地落在由噪声事例填充的各个通道之外, 任何一个这样

的事例 *Outlier* 都是一个合适的候选信号, 因为如果互相关值很大, 就不太可能是由噪声自身单独生成的。

对于给定的模板  $S(t)$ , 我们可以把噪声定义为:

$$N^2 \equiv \sqrt{\langle (V * S(\tau))^2 \rangle} . \quad (7)$$

它是噪声与模板之间互相关值的均方根值, 是  $V(t) * S(t)$  的直方图宽度的量度。

我们用  $S^2$  表示在任意时刻  $t$  出现的信号的强度, 它用期待的输出形式  $V(t)$  的互相关函数表示:

$$S^2 \equiv |V * S(t)| . \quad (8)$$

信噪比  $SNR$  是信号存在时的  $S^2$  值与噪声单独存在时的  $N^2$  之比的平方根:

$$SNR = \sqrt{S^2/N^2} . \quad (9)$$

信噪比  $SNR$  给出了在探测器的输出中, 出现只有噪声没有其他信号的事例的可能性, 也就是说,  $SNR$  给出了在探测器的输出中出现含有噪声以外其他信号的可能性有多大。  $SNR$  愈大, 输出中含有其他信号的可能性愈大。也就是说, 大的  $SNR$  值指明, 在该输出时间序列函数中, 存在着噪声以外的东西: 它可能是要探测的引力波信号, 也可能是其它干扰信号。

上述讨论表明, 从噪声中把信号挑拣出来的方法之一, 是在带有噪声的探测器输出  $V(t)$  和我们认为真实的信号模板之间建立互相关函数。因为, 我们要做的就是从多种噪声混杂体中抽取某些具有特殊形态的事例。可以证明<sup>[22]</sup>, 当处理的噪声是“白噪声”时, 模板方法是最佳的选择; 当处理的噪声不是“白噪声”时, 情况会复杂一些, 但基本方法类似。模板分析法是利用计算机直接计算探测器输出与模板之间的互相关函数。我们也称这种分析方法为数据处理中的离线分析, 因为它是先把探测器的输出数据记录在介质上, 然后再根据需要的物理问题, 建立可用的模板, 进行互相关运算, 研究感兴趣的物理问题。

匹配滤波器是引力波数据分析中功能最强大的工具之一, 它是在搜寻致密双星并合信号中逐步发展和健全起来的一种分析技术。比如, 中国、澳洲和美国合作的用来探测引力波的新方法——SPIIR<sup>[23]</sup>, 就是匹配滤波器的运用之一。在这个方法里, 将干涉仪信号与模板进行相关分析, 如果一个信号事件与一个模板的相关值较高, 就说明观测到的信号不完全是噪声的可能性比较高, 而且有可能是与该模板所关联的引力波信号。进而, 用多个模板与信号进行相关性计算, 以检查该信号是某个特定的引力波波形的可能性有多大。

在用激光干涉仪进行引力波探测时, 噪声是一个需要解决的主要问题。匹配滤波器有一个非常有用的性质: 由于干涉仪的输出信号可以表示为单纯的引力波信号和随机噪声的和, 在干涉仪输出信号与给定模板的相关性计算得到的相关值中, 可能找到一个最优的信噪比, 它就是我们选出的值得进一步分析的有用“事例”<sup>[24]</sup>。从这个意义上来说, 用匹配滤波器对数据进行过滤是一个不错的选择。

### 3.4 差错排除

到目前为止, 人们还未直接探测到引力波。根据理论计算和模型分析可推出, 引力波的强度是非常微弱的。它在臂长为千米量级的激光干涉仪引力波探测器中, 使测试产生的位移

只有质子直径的千分之一或更小。然而, 从目前激光干涉仪的发展水平和得到的技术参数来看, 非常有可能利用这种灵敏度极高的设备探测到引力波。但是这个仪器对环境的干扰又是非常敏感的, 用它观测到的具有一定能量的事件频率远远高于预期的能够让激光干涉仪捕捉到的真实引力波事件的发生频率。这就是说, 在捕捉到的事件中大多数并不是由于引力波产生的, 而是由于探测器仪器故障、缺陷或周围环境产生的噪声引起的, 习惯上大家把这种瞬间能量集中形成的“尖峰”称为“差错”。由于这种虚假事例太多, 在深入进行数据分析前, 必须把它们“否决”掉。通过否决权的使用可以排除绝大多数“差错”, 在进行物理分析时节省大量的时间和资源, 对后续的数据分析工作非常有利。否决权的研究就是我们常说的“veto study”, 其基本原理是: 通过对激光干涉仪关键部位及对周围环境监控信号(它们被记录在辅助数据道上)的分析, 对每个事件发生的时间段里是否有仪器故障或干扰做出判断, 进而对该段时间上引力波主数据道事件的可靠性做出裁决。如果这个时间段仪器不稳定, 或者环境中干扰噪声过强, 就认为这段时间里采集到的引力波数据信号不可信。值得注意的是, 由于引力波信号过于微弱, 且又有大量的噪声充斥在探测器的数据段内, 加上仪器可以捕捉到的产生引力波的天文学事件发生的频率较低, 如果过分地否决掉一些有噪音的事件, 也有可能把真正的引力波事件从搜索中遗漏。因此, 追求高效率和低误判率是研究事件否决权的主要目标。这项工作长期以来一直是一项重要的研究内容。

由于辅助数据道的种类多、数量广、内容丰富, 它们之间的关系复杂, 对它们的分析难度很大。详细内容请参阅参考文献[25-34]。

随着信息科学领域的迅速发展, 一些基于统计学的方法, 机器学习的方法也被应用到这个问题的研究中<sup>[35]</sup>。该论文的作者运用了随机森林算法<sup>[44]</sup>、人工神经网络<sup>[45, 46]</sup>和支持向量机三种不同的机器学习算法, 来分析引力波数据道中的噪音, 对引力波数据道上捕捉到的事件进行分类。过去几十年中这三种机器学习算法在计算机科学、生物、医学等领域被广泛应用。将这些算法延伸到引力波数据分析领域, 将在识别引力波数据中仪器异常产生的噪声事件研究中发挥重要作用。由于篇幅所限, 这里只能对它们做简单的介绍。

神经网络机器学习方法的思想来源于模拟人类大脑中数据处理识别问题, 是一种以模仿动物神经网络行为特征, 进行分布式并行信息处理的算法数学模型。随机森林方法是改进后的经典决策树分类方法, 它是一个包含了多个决策树的分类器, 通过建立多个决策树(森林一词的来源)并对其结果进行综合评估, 如取平均值等; 所以其结果并不是分类的标签, 而是分类的置信度的表示, 通常是  $[0, 1]$  之间的小数, 可以认为是分类结果的置信概率。支持向量机(support vector machine)采用的是与前面两种方法完全不同的又一种机器学习的二分类方法: 在样本空间里, 寻找一个超平面, 使得这个平面能最大可能地把两类训练样本分隔开; 这个超平面经过训练确定下来之后, 分类问题就转变成判断一个样本落在该超平面分隔的哪个子空间上的问题; 而超平面的寻找最终被转换成求解一个二次规划的计算问题。

虽然这三种机器学习的方法各有千秋, 但是经过调试优化之后, 它们在引力波数据噪音分析和事件分类上表现出来的能力基本相同。这个结论很有意思, 它告诉我们, 机器学习方法在应用中所能达到的效果, 很大程度上决定于数据质量, 与具体采用哪种方法关系不太大。

### 3.5 变化点分析法

还有另外一种方法,它同样不需要预先知道波形的有关知识就能辨认引力波信号,这就是变化点分析法。该方法的原理是在时间域内寻找探测器输出数据的统计学参数的变化。其做法是把输出数据分成很多小组,每个数据小组内统计学参数近似为常数,这个统计学参数用正态分布的平均值和方差来表示。变化点定义为噪声特性(以平均值和方差表示)变化的时间点。也就是说,如果在这个时间点的随便哪一边具有不同的统计分布值,且在属于这种不同统计分布值的那一边,其统计学参数值超过给定的阈<sup>[36]</sup>,我们就可以选择它。这种被辨认出的变化点划定了一个数据小组的起始和终止时间点,数据小组的统计学参数以起始和终止时间内的平均值和方差来表示。

一旦具有不同平均值和方差的数据小组确定之后,相邻的、有异常表现的数据小组就被聚集在一起,形成单个的有用事例。然后,这种事例就可以在世界各地的引力波探测器上进行符合分析,包括它的频带宽度、最高值数据小组出现的时间、刻度过的能量以及事例持续的时间等,从而确定事例的真实性。

### 3.6 联合与相干技术

联合与相干技术在引力波数据分析中主要有两大功能:判断事例真伪和确定波源位置。

#### 3.6.1 确定事例的真实性

将多个探测器联合起来进行关联测量与分析,对判断引力波信号的真伪是非常有利的。例如,假设将位于不同地域的三个探测器联合起来进行测量,若只有一台探测器的信号里发现了超高的能量,那么这个信号很有可能是来自某个短暂的噪声源(如某个开关的开启或关闭,异常的地面震动等)。如果真正的引力波到达地球,那么这三个探测器势必都会一致性地有所反应。它们不仅都会观测到超高能量的信号(能量的大小与探测器的位置和引力波极性的相对关系有关),而且它们的信号还会有相互关联的相位。也就是说,在三个探测器上观测到的引力波信号来源于同一个引力波,只不过仅仅有一点相移而已。从这一点可以推出:因为噪声信号通常是不相干的,所以,如果从不同的探测器上得到的信号能够被相干组合的话,即除了各自保留相位信息外能具有一致性,那么就可以认为该事例是引力波而不是噪声的可能性更高。

数据分析流水线 Wave Burst 及早先的基于最大似然<sup>[37]</sup>或基于 F 统计<sup>[38]</sup>的引力波数据分析手段都可用来做干涉仪信号的关联分析。

多个引力波探测器信号之间的关联性对于压低噪声和辨认真实的引力波信号来说是至关重要的。利用

$$t_1 = t_2 + L/c, \quad (10)$$

其中  $t_1$  和  $t_2$  是每个探测器的记录时间,  $L = |(\mathbf{x}_1 - \mathbf{x}_2) \cdot \mathbf{n}|$  是探测器之间的投影距离,  $\mathbf{n}$  是引力波源的方向,我们可以建立探测器信号之间的关联性  $Y$ ,

$$Y = \int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 S_1(t_1) Q(t_1 - t_2) S_2(t_2), \quad (11)$$

这里  $S_1(t)$  和  $S_2(t)$  分别是两个探测器的输出(应变),  $Q$  是过滤器。选择这个过滤器的目的

是让我们所期待的信号的关联性  $Y$  有最大的信号/噪声比。为了选择最佳过滤器  $Q(t)$ , 可以假设引力波探测器的噪声是稳定的, 具有高斯分布形式, 且不同的探测器之间互不相关。

### 3.6.2 确定事例源的空间位置

利用单个探测器很难辨认引力波是从太空中哪个地方来的。相互关联的信号把全球引力波探测网内各个探测器联合成单一的引力波天文台。原则上讲, 它应该有很好的空间分辨率。网内各个探测器信号到达的时间差给我们带来引力波传播方向的信息, 波源位置的确定仅仅是利用这些信息做些三角运算而已。

下面我们讨论怎样利用探测器信号的符合测量来辨认引力波波源的位置<sup>[40]</sup>。设两个探测器之间的输出信息是相互关联的, 我们可以很好地估算两个探测器输出信息之间的时间移动  $\Delta t_{\text{sig}}$ 。从这个意义上讲, 探测器之间的实时关联测量并不是必须的, 但是, 对于每个数据流来说, 好的时间标识是完全必要的。用两个探测器之间的时间差, 如果不考虑测量误差的话, 我们可以在太空中划定一个圆, 如图 5 所示。

由于测量精度是有限的, 存在测量误差, 我们在太空中划定的不是一个圆, 而是宽度为  $d\theta$  的一条带状环。很明显, 如果我们能精确地测定时间差 (如提高信噪比以减小随机误差, 精确设置波形和极化状态以减小系统误差) 就可以使  $\Delta\theta$  大大减小, 从而提高定位精度。

引力波信号是从这个带状环内发出的。这个带状环很像一个系统中的两条平行的纬度线或赤纬线。系统的极轴定义为两个探测器之间连接的延长, 如图 6 所示。图中  $D$  为两个探测器之间距离。定位圆相对于极化轴的倾斜角为  $\theta$ ,  $\theta = \cos^{-1} c\Delta t/D$ 。

利用三个探测器能够得到两个独立的时间差, 两个独立的时间差可以在太空中划定两个定位带状环, 这两个定位带状环的交叉在太空中形成两个“补丁”, 如图 7 所示。引力波信号是从这两个“补丁”中的一个内发出的。利用第四个探测器的观测将给出第三个时间差, 它足以在太空中把引力波源定位在一个“补丁”内。

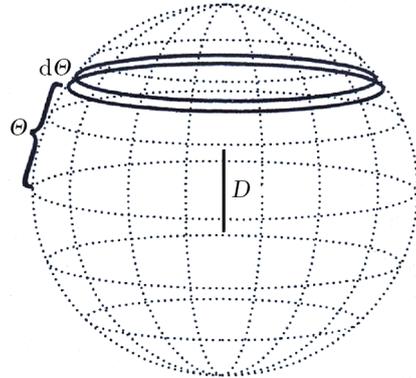


图 5 从外侧观测到的天体图<sup>[21]</sup>

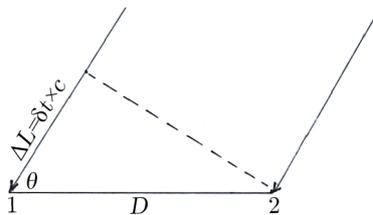


图 6 极化轴倾斜角  $\theta$

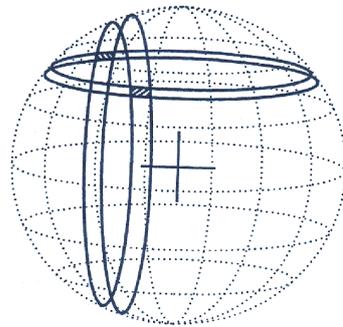


图 7 两个带有测量误差的圆在误差箱上重叠<sup>[21]</sup>

如果我们建立了一个国际引力波探测网,网内探测器的时间分辨率为 0.1 ms,探测器基线之间的距离为数量级为  $6 \times 10^3$  km (相当于地球半径),那么,波源的位置就可以被确定到大约  $(5 \text{ mrad})^2$ 。这是一个相当不错的角定位精度<sup>[21]</sup>。

### 3.7 数据分析流水线

在激光干涉仪引力波探测器运行时,产生的数据会源源不断地流入存储设备和分析“工具”中,数据分析“工具”通常是一套由一组软件构成的“流水线”,它的功能是读入数据,然后对数据做多个步骤的分析。

根据能否实时处理从干涉仪流入的数据流,数据分析流水线被分为在线的和离线的流水线两大类。高级 LIGO 探测仪的精度比初级 LIGO 高一个数量级,这意味着其可探测的引力波源的数量提升到原来的 1000 倍,也就是说,当 Advanced LIGO 运行时,比以往多得多的数据将进入到数据处理软件中进行处理。保证数据膨胀情况下软件仍然能正常、高效地运行,是 LIGO 数据分析软件系统需要面对并解决的重要问题。当前大家正在努力做的事情就是优化数据分析流水线的处理速度,使它能够用于实时的在线数据分析,而在进行离线分析时速度更快。计算速度是很重要的,因为激光干涉仪是连续灵敏的,数据通道又非常之多,数据源源不断地流出,若不及时快速处理,就会“堆积如山”,降低探测器的工作效率,使先进的硬件设备的威力发挥不出来,延长计算时间。

#### 3.7.1 数据分析流水线与在线数据分析

在引力波数据分析中,“流水线”是从将引力波数据分析与工业生产及多媒体流处理软件的类比中引申出来的。从计算的角度来看,数据从 LIGO 探测仪中流出后,需要经过缓存、处理和最后再输出的过程,这一过程与现实世界中的流水线非常类似。在计算机软件中,LIGO 的数据处理过程与多媒体流处理软件的运行机制亦非常吻合,因此,在 LIGO 数据处理软件的选择上,研究人员自然而然地想到了使用流式处理软件来对激光干涉仪引力波探测器输出的数据流进行深度处理,这就形成了引力波数据分析中大家所说的分析“流水线”。

GStreamer 是一款开源的多媒体处理基础设施软件,它创建于 1999 年,到目前为止,已被绝大多数 Linux 发行版本用来支撑底层多媒体软件系统的运行。在软件编写上,GStreamer 提供了一个管道系统,软件开发者可以动态或者静态地在管道中添加相应的处理单元(输入单元、过滤单元和输出单元),其中过滤单元对输入数据进行某种处理,处理完成后再立即输出。这样一来,数据就像流水一样,在 GStreamer 的流水线上流动起来了。

正因为 GStreamer 流式处理数据的特性与 LIGO 探测器数据处理模型的相似性,GStreamer 被 LIGO 合作研究组中的一部分研究人员选作了数据处理软件的基础构件,在 GStreamer 的基础上,经过 LIGO 合作研究组的共同努力,研究者们将数年间所开发的引力波数据分析子程序套件 LAL Suite 与 GStreamer 整合起来,开发并构建出了一套新的、专门用于引力波信号处理的管道流水处理系统,这一系统便是如今在 LIGO 合作研究组中得到广泛应用的 GstLAL 流水线。GstLAL 是专门为引力波数据处理量身定做的,有了它之后,研究人员可以省略大量重复性的基础工作,而更加专注于数据处理算法本身和特殊的物理问题,极大地提高了研究效率。

数据流水线的建立为高效地处理 LIGO 探测器数据提供了可能。在数据流水线中, 如果运行的数据处理算法可以实时地完成数据处理工作, 那么引力波的探测工作就可以实时化。它就是物理研究工作中所说的“在线分析”, 在线分析在实际研究中具有重大意义。例如, 实时数据处理的结果可能会给出某些正在发生的物理现象的线索, 这样就可以指导探测器对所探测的天空点的方位进行实时修正; 对于某些稍纵即逝或者持续时间非常短暂的物理观测过程, 实时数据处理将会带来巨大帮助。

在现有条件下, 要完成实时探测是一项相当具有挑战性的工作。首先, Advanced LIGO 所产生的超大数据量意味着要想实时处理, 计算机系统需要具有非常强的处理能力。其次, 由于现有的数据处理算法并非效率很高, 因此, 在算法层面上需要进行诸多的优化以减少实际处理中的计算量。再说, 实际的研究中, 数据处理通常不能达到实时, 而是有一定的延迟, 广义上讲, 延迟是数据从流出 LIGO 探测器到处理完成之间所经历的时间; 它是不可避免的, 我们的努力只能体现在构建低延迟的数据处理流水线上。

在进行低延迟引力波探测和数据分析时, 匹配过滤器 (matched filtering) 被成功地运用在 GstLAL 数据流水线中, 它是在 LIGO 科学合作研究组之致密双星并合小组的推动下完成的。这一方法的基础是 Wiener 最优滤波, 其基本思路是将预期的双星旋绕波形模板和引力波探测数据进行关联, 再按照探测器的反噪声谱密度进行加权<sup>[42]</sup>。为了减少计算消耗, 这一关联过程通常是在频率域内通过傅里叶变换的形式进行的。在初级 LIGO 探测中, 探测器数据被划分成一个个的“科学数据块”, 这些数据块再被划分为更小的“数据段”, 每个数据段的大小被选择为两倍于模板库的大小。这样一来, 为了达到实时计算的效果, 每个数据段必须在一半于数据段所占时间内完成匹配过滤。也就是说, 匹配滤波过程的最小延迟 (从信号到达探测器到信号被探测到的时间) 与最长模板大小成正比。

### 3.7.2 GPU 通用计算技术的应用

在第二代激光干涉仪引力波探测器如 Advanced LIGO 中, 可探测的最低频率从 40 Hz 降低到了 10 Hz, 大大提高了频带宽度, 而致密双星并合所产生的引力波事件绝大部分发生在低频区域; 因此, 在 Advanced LIGO 上用于匹配过滤的波形将会变得比初级 LIGO 上用的要长得多, 这也就意味着数据段长度会增长, 探测延迟随之增长。这表明, 使用这一处理方法来挑选出一个有用的引力波触发 (GW trigger) 所用的时间要长。经过这一延迟, 伽马射线暴 (GRB) 事件伴随的早期电磁信号几乎已经衰减至极其微弱的地步, 对于这种事例有可能造成部分信号的探测遗漏。

为了进一步减少时间延迟, Hooper 等人<sup>[23]</sup>提出了一个新的算法, 即并行合并无限冲击响应滤波算法 (SPIIR), 该算法对每一个信号使用一种迭代的方式来产生结果信号, 它的特点是不同的模板、同一模板中的不同信号之间完全没有任何相互干扰。由于通常情况下有成千上万的模板需要处理, 而每个模板都有几十到数百个信号, 因此, 在这一方法中使用计算机进行大规模并行处理确实是一种能提高计算效率的方法。在以往的软件流水线中, CPU 几乎是唯一的计算资源, 随着近年来 GPU 通用计算的兴起, Chung 等人将匹配过滤算法部分移植到 GPU 上<sup>[41]</sup>, 取得了数十倍的加速效果。在 GPU 加速方面, 清华大学与西澳大利亚大学开展了合作, 对 SPIIR 算法进行了面向 GPU 的移植与优化工作, 得益于 SPIIR 算法的高

并行度,以及深度优化数据结构的设计和算法细节的优化工作。到 2012 年底, SPIIR 在单机上使用 NVIDIA GTX480 显卡,与 Intel Core i7 920 单核性能对比,已获得了 58 倍的加速比<sup>[43]</sup>。到 2015 年初,经过进一步优化的 SPIIR GPU 算法在单机上使用 NVIDIA GTX980 显卡,与 Intel Core i7 3770 单核相比,已获得了最多 124 倍的加速比。GPU 通用计算技术的引入极大地提高了计算机集群单节点引力波数据处理的能力。

在数据流水线的软件架构中,数据以流的形式存在,每一个数据处理单元被称作一个元素,元素与元素之间通过接口进行连接。将所有这些元素在数据流水线中组装完成以后,数据便可以像水流一样在流水线中流动起来了。在 GPU 计算中,通常位于 CPU 端的内存不能被 GPU 直接读取,而是需要编程人员将需要计算的数据先从内存传输到 GPU 专用内存,然后再进行计算,最后将计算完成的数据从 GPU 端传输到 CPU 端。与计算过程相比, GPU 与 CPU 内存之间的数据传输由于需要使用 PCI-E 总线进行,因此要慢得多,这也给软件流水线与 GPU 的结合带来了诸多挑战:一是与 CPU 端内存相比, GPU 自身的存储空间要相对小得多,这就要求精心设计流水线算法数据流量,以避免使用超过 GPU 承载能力的存储需求;二是传统 GstLal 流水线只能传输 CPU 端内存指针,这就意味着无论 GPU 计算数据在元素后来的阶段是否能够被使用到,都必须将结果传输回 CPU 端,再通过接口使数据流动到下一个元素进行处理。可以想象,如果两个相邻的数据流水线元素都使用 GPU 进行计算,后一个元素正好要用到前一个元素的计算结果,那么这样从 GPU 到 CPU 的数据传输是完全没有意义的。为了解决这一问题, GstLAL 与 Gstreamer 社区进行了深入沟通,目前 Gstreamer 1.0 已经能够较好地支持 GPU 指针的传输,而 GstLAL 的 Gstreamer 部分向 1.0 的版本转换工作正在进行当中。可以预期,使用新版本 Gstreamer 的 GstLAL 软件流水线与 GPU 的结合将会进一步让引力波数据处理受益。

### 3.7.3 GPU 进行 SPIIR 滤波的过程

GPU 进行 SPIIR 滤波的全过程如图 8 所示, GPU 的执行模型主要分为两层: Grid-Block 层和 Block-Thread 层, Grid-Block 层对问题进行粗粒度划分, Block-Thread 层在粗粒度划分的基础上再对任务进行细粒度划分。图 8 右图展示了 Block-Thread 层的任务划分模型,对于一个 SPIIR 模板,分配一个 GPU Block 来进行它的 IIR 计算。通常来讲,一个 GPU 拥有数十个流处理器,在 GPU 中,一个 Block 是一个基本的调度单元,它可以被 GPU 上的流处理器调度执行。每个 Block 又可以细分为多个 Warp,一个 Warp 由 32 个线程组成,线程是 GPU 中的基本执行单元。在图 8 中,一个 Block 被细分为多个 Group,每一个 Group 中的线程完成计算后,通过一个同步操作,将计算结果同步至 GPU 中一个 Block 范围的共享内存中,再使用多个线程进行规约操作,最终得到计算结果 SNR (信噪比)。在 Grid-Block 层面, GPU 在 Int 数据类型可支持的范围内,支持任意大小的 Block 数目。这样 Block 数目乘以 Block 的大小 (Block 中的线程数目) 就是 GPU 启动的所有线程数,由此可以看到,在通常情况下 GPU 可以启动成千上万的线程进行大规模并行计算。当然, GPU 中这些线程并非都是同时开始执行的,而是通过调度器以 Warp 为单位对线程调度执行。每个 Warp 在任意时刻由硬件保证步调的一致性,这也是很多同步优化策略的硬件基础。需要注意的是, GPU 的线程与 CPU 线程虽然名称一致,但是它们的切换开销却有着本质的区

别: GPU 对硬件资源进行预分配, 因此每一个线程都保证拥有独立的资源, 这样 GPU 在对线程进行切换的时候不需要保存上下文, 完全没有切换开销; 而 CPU 因为资源共享的原因, 线程切换的开销相对来讲是非常大的。

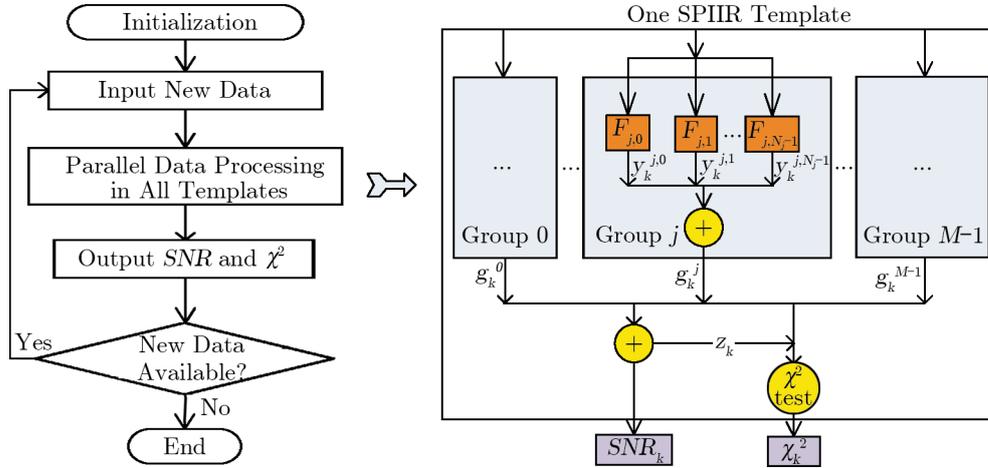


图 8 GPU 进行 SPIIR 的过程图

将 GPU 与 GstLAL 流水线结合起来, 极大提升了引力波数据处理的效率, 为实时处理第二代激光干涉仪引力波探测器 (如 Advanced LIGO) 的数据提供了可能性。目前, SPIIR 部分的延迟已被降至 10 s。

### 3.8 $\chi^2$ 检验

在引力波探测研究中, 虽然把匹配过滤器设计成在信号与模板相匹配时才给出大的响应, 但是, 探测器的“差错 (glitch)”噪声也会触发过滤器, 给出大的输出。“差错”的波形与模板一点也不一样, 它们与模板一点也不匹配。过滤器的这种输出并非有用信号引起的, 它们是假输出。

$\chi^2$  统计检验提供了一个去除这些虚假事例的有效手段, 因为这种统计检验能够确定过滤器的输出是否与当信号和模板相匹配时所期待的输出相一致。也就是说, 所谓  $\chi^2$  检验就是计算  $\chi^2$  统计值, 以检验“触发”信号与期望信号波形的符合程度。

$\chi^2$  统计检验法的基本思想是把我们使用的单个宽频带引力波探测器想象成由  $p$  个独立的且不同的窄频带子探测器组成的一个集合体, 集合中的每个子探测器工作在不同的窄频带内, 这组窄频带子探测器给出  $p$  个数据流。对于集合中的每个子探测器, 我们对信号构建一个最佳过滤器, 然后看看这些子过滤器的响应是否一致。比如说, 对  $p$  个独立的子探测器选择  $p$  个不同的基准时间  $t_0$ , 这  $p$  个不同的基准时间  $t_0$  分别使每个独立的子探测器的输出最大化。最后查看这  $p$  个基准时间是否为同一个值。

#### (1) 频率区间的划分

在引力波数据分析中,  $\chi^2$  检验的具体做法是把探测器覆盖的频率范围划分成多个 (如  $p$  个) 子频率间隔, 并且把总模板也分成多个子模板, 每个子模板分别覆盖对应的子频率间隔。

子频率间隔及对应的  $p$  个子模板以  $j$  来标识,  $j = 1, 2, \dots, p$ 。探测器获取的数据分别用由子模板构建的匹配过滤器进行过滤, 从而得到一组过滤器输出  $z_j(t)$ , 利用  $z_j(t)$  可以推导出进行  $\chi^2$  统计检验所需的一系列物理参量。

为了使统计结果具有经典的  $\chi^2$  分布, 子频率间隔的大小必须满足如下条件: 每个子频率间隔的触发输出  $z_j(t)$  的期待值对总信号输出  $z(t)$  的贡献是相等的, 即

$$\langle z_j(t) \rangle = \frac{z(t)}{p} . \quad (12)$$

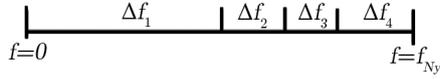


图 9 在  $p = 4$  时, 一组典型的子频率区间示意图<sup>[39]</sup>

对于密近双星旋绕产生的鸟鸣信号, 典型的一组子频率区间在图 9 中给出。可以看出, 在探测器最灵敏的地方频率区间最窄, 而在探测器最不灵敏的地方区间最宽。

### (2) $\chi^2$ 统计量的建立

为了使思路更加清晰, 我们先推导建立  $\chi^2$  统计所需的物理量  $\chi^2$ , 根据定义, 匹配过滤器的信噪比由下式给出:

$$Z \equiv \int \frac{\tilde{Q}^*(f)\tilde{s}(f)}{S_n(f)} df = (\tilde{Q}, \tilde{s}) . \quad (13)$$

它是对所有频率的一个积分。分成子频率间隔后, 它可以写成从  $p$  个不同子频率间隔而来的贡献之和:

$$z = \sum_{j=1}^p z_j , \quad (14)$$

其中

$$z_j \equiv (\tilde{Q}, \tilde{s})_j . \quad (15)$$

当信号不存在时, 我们得到:

$$\langle z_j \rangle = 0 , \quad \langle z_j^2 \rangle = \frac{1}{p} . \quad (16)$$

若在所有频率中测得的总信噪比为  $z$ , 摊派到各个子频率间隔  $\Delta f_j$  内预言的信噪比为  $\frac{z}{p}$ , 那么在子频率区间内的信噪比  $\frac{z}{p}$  与预言的信噪比之间的差  $\Delta z_j$  为:

$$\Delta z_j \equiv z_j - \frac{z}{p} . \quad (17)$$

这种物理量共有  $p$  个, 根据定义, 它们的和为零:

$$\sum_{j=1}^p \Delta z_j = 0 . \quad (18)$$

而且, 每个子频率间隔内这个差的期待值也是零:

$$\langle z_j \rangle = 0 . \quad (19)$$

根据

$$\langle z_j z \rangle = \frac{z^2}{p}, \quad \langle z_j^2 \rangle = \frac{1}{p}, \quad (20)$$

我们可以得到  $\Delta z_j$  平方的期待值:

$$\langle (\Delta z_j)^2 \rangle = \left\langle \left( z_j - \frac{z}{p} \right)^2 \right\rangle = \langle z_j^2 \rangle + \frac{\langle z^2 \rangle}{p^2} - \frac{2\langle z_j z \rangle}{p} = \frac{1}{p} \left( 1 - \frac{1}{p} \right). \quad (21)$$

根据以上导出的  $p$  个物理参量, 可以方便地定义  $\chi^2$  时间-频率甄别器的统计量为:

$$\chi^2 = \chi^2(z_1, z_2, \dots, z_p) = p \sum_{j=1}^p (\Delta z_j)^2. \quad (22)$$

(3)  $\chi^2$  时间-频率甄别器统计量的特性

a.  $\chi^2$  的期待值

根据公式  $\langle (\Delta z_j)^2 \rangle = \frac{1}{p} \left( 1 - \frac{1}{p} \right)$ , 我们可以推导出  $\chi^2$  的期待值:

$$\langle \chi^2 \rangle = \left\langle p \sum_{j=1}^p (\Delta z_j)^2 \right\rangle = p \left\langle \sum_{j=1}^p (\Delta z_j)^2 \right\rangle = p \left( p \cdot \frac{1}{p} \left( 1 - \frac{1}{p} \right) \right) = p - 1. \quad (23)$$

b.  $\chi^2$  的几率分布函数

在仪器噪声既稳定又有高斯分布的情况下,  $\chi^2$  分布函数几率是经典的, 自由度为  $p - 1$  的分布。 $\chi^2 < \chi_0^2$  的几率为<sup>[39]</sup>:

$$P_{\chi^2 < \chi_0^2} = \int_0^{\chi_0^2/2} \frac{u^{p/2-3/2} e^{-u}}{\Gamma(p/2-1/2)} du = \frac{\gamma(p/2-1/2, \chi_0^2/2)}{\Gamma(p/2-1/2)}, \quad (24)$$

$\gamma$  是不完全伽马函数。在探测器的噪声稳定且具有高斯分布的情况下,  $\chi^2$  值期待的分布是相当窄的, 它是:

$$\langle (\chi^2)^2 \rangle = p^2 - 1. \quad (25)$$

这意味着  $\chi^2$  分布的宽度为:

$$\langle (\chi^2)^2 - \langle \chi^2 \rangle^2 \rangle^{1/2} = \sqrt{2(p-1)}. \quad (26)$$

也就是说, 我们能够在区间  $\left[ p - 1 - \sqrt{2(p-1)}, p - 1 + \sqrt{2(p-1)} \right]$  内找到的值。由于该区间的相对宽度为:

$$\frac{\sqrt{2(p-1)}}{p-1} = \frac{\sqrt{2}}{\sqrt{p-1}}. \quad (27)$$

它随  $p$  的增加而减小, 我们可能认为大的  $p$  值是合乎理论的, 因为它好像能给出较高的甄别能力。但是实践和经验表明,  $p$  值并不是越大越好, 它的选择要适度。因为探测器的噪声既

不稳定也不是高斯分布,而且信号与模板的匹配并不完美。当采用的  $p$  值非常大时,不稳定的“差错”噪声会覆盖很多子频带,削弱了它在  $\chi^2$  上的作用。

#### (4) $\chi^2$ 时间-频率甄别器检验效果<sup>[39]</sup>

下面我们通过一个简单的例子说明  $\chi^2$  时间-频率甄别器的检验效果。如前所述,该方法的基本出发点是把探测器覆盖的频率范围划分成  $p$  个子频率间隔,并建立一组相应的匹配过滤器。用它们对获取的数据进行检验,看一看这些匹配过滤器输出信号的峰值是否都在正确的时间点出现。

在实施  $\chi^2$  检验时,我们首先用蒙特卡罗法产生两组数据:一组是把一个模拟的鸟鸣信号(即我们期待的双星旋转发出的信号)附加在探测器的噪声上产生的;另一组数据是模拟探测器噪声中有一个瞬发即逝的“差错”。为简单起见,我们把探测器覆盖的频率范围划分成 4 个子频率间隔,对每个子频率间隔都建立一个匹配过滤器,然后利用这 4 个匹配过滤器分别对两组数据进行过滤。

设信号  $z_1$  是由最低频带构成的,  $z_2$  是由其后的下一个频带构成的,以此类推。对每组模拟数据,都分别得到 4 个信号  $z_1, z_2, z_3, z_4$  (这里所说的信号指的是过滤器的输出  $z_1, z_2, \dots, z_p$ ) 图 10 给出了这四个输出信号的示意图。

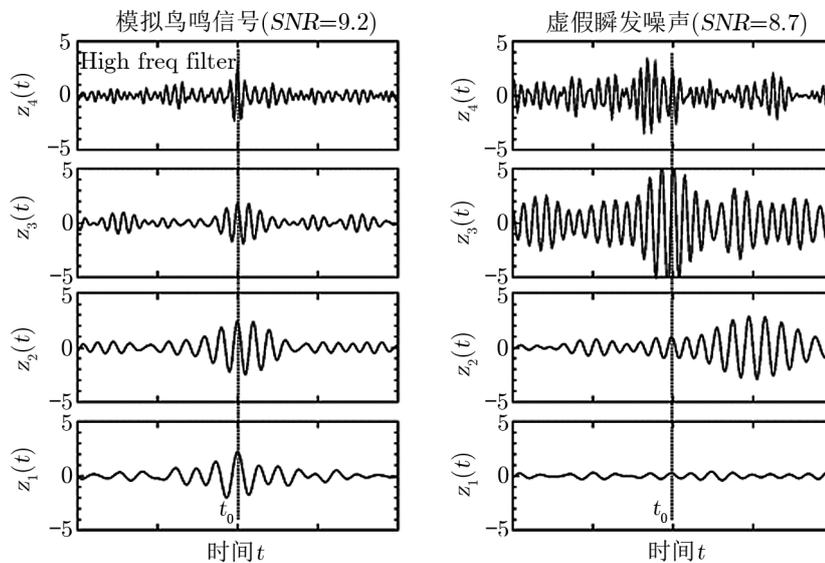


图 10 匹配过滤器的输出示意图<sup>[39]</sup>

左边的一组图是把一个模拟的“鸟鸣”信号(即 Chirp 信号)附加到探测器噪声数据流中的情况,右边一组图是探测器噪声数据流中存在一个路过的爆发性信号的情况。可以看出,对模拟的“鸟鸣”信号来说(左图),不同频带的匹配过滤器输出信号的峰值都出现在同一个时间偏移  $t_0$ ,在时间中的这一瞬间,所有的  $z_j$  都在同一个数值附近。然而,当过滤器被一个转瞬即逝的爆发性信号触发时(右图),不同频带中过滤器输出信号的峰值出现在不同的时间点。在时间  $t_0$  处,它们的值非常不同,有的很大,有的很小。下面我们通过具体数值对这两

种情况进行比较。

(1) 模拟的“鸟鸣”(chirp) 信号

从图 10 的左图可以看到, 模拟的“鸟鸣”信号的信噪比为  $z = 9.2$ , 在不同频带中匹配过滤器的输出为:

$$z_1 = 2.25, \quad z_2 = 2.44, \quad z_3 = 1.87, \quad z_4 = 2.64, \quad z_1 + z_2 + z_3 + z_4 = 9.2. \quad (28)$$

$\chi^2$  统计量的值为:

$$\chi^2 = p \sum_{j=1}^p \left( z_j - \frac{z}{p} \right)^2 = 4 \sum_{j=1}^p \left( z_j - \frac{9.2}{4} \right)^2 = 1.296. \quad (29)$$

该  $\chi^2$  值出现的几率为:

$$P_{\chi^2 \geq \chi_0^2} = P_{\chi^2 \geq 1.29} = \frac{\gamma(p/2 - 1/2, \chi_0^2/2)}{\Gamma(p/2 - 1/2)} = \frac{\gamma(3/2, 0.648)}{\Gamma(3/2)} = 73%. \quad (30)$$

这个结果与附加到高斯噪声中的“鸟鸣”信号所期待的  $\chi^2$  值及该值出现的几率相当一致。

(2) 欺骗性的噪声信号

从图 10 的右图可以看到, 欺骗性的、转瞬即逝的噪声信号的信噪比为  $z = 8.97$ , 它和“鸟鸣”信号的信噪比差不多, 在不同频带中配过滤器的输出为:

$$z_1 = 0.23, \quad z_2 = 0.84, \quad z_3 = 5.57, \quad z_4 = 2.33, \quad z = z_1 + z_2 + z_3 + z_4 = 8.97. \quad (31)$$

$\chi^2$  统计量的值为:

$$\chi^2 = p \sum_{j=1}^p \left( z_j - \frac{z}{p} \right)^2 = 4 \sum_{j=1}^p \left( z_j - \frac{8.97}{4} \right)^2 = 68.4. \quad (32)$$

该  $\chi^2$  值出现的几率为:

$$P_{\chi^2 \geq \chi_0^2} = P_{\chi^2 \geq 68.4} = \frac{\gamma(p/2 - 1/2, \chi_0^2/2)}{\Gamma(p/2 - 1/2)} = \frac{\gamma(3/2, 34.2)}{\Gamma(3/2)} = 9.4 \times 10^{-15}. \quad (33)$$

可以看到, 从信噪比的大小来看, 两种情况差不多。假设我们只根据信噪比的大小来选择我们需要的“真实事例候选者”, 就会把这种转瞬即逝的、爆发性的虚假信号当成“真实事例候选者”。但是, 如果从  $\chi^2$  检验结果来考虑, 我们就会发现, 两种情况下的  $\chi^2$  值及它们可能出现的几率有天壤之别。对于附加到高斯噪声中的“鸟鸣”信号来说, 出现这个  $\chi^2$  期待值的几率是极其微小的 (只有  $9.4 \times 10^{-15}$ ), 通过这两种情况的比较, 我们可以清楚地看到  $\chi^2$  检验的巨大威力。

## 4 总结与展望

引力波是广义相对论最重要的预言, 引力波探测是当代物理学重要的前沿领域之一, 以引力波探测为基础的引力波天文学是一门新兴的交叉学科, 是人类观测宇宙的一个新窗口, 对研究宇宙的起源和进化, 拓展天文学的研究领域都有极其重要的意义。

由于引力波信号非常弱,引力波天文学研究不但需要有灵敏度很高、频带很宽的探测器,需要高超的探测技术,更需要有科学的数据处理和分析方法。随着第二代激光干涉仪引力波探测器的建造、统调和投入运转,引力波数据分析的基本原理基本方法获得了广泛的应用,而更加先进的引力波数据处理方法及相应的通用软件的研发也同步开展起来,并取得了巨大进展。不同特色、不同研究方向、不同研究课题的专用和通用软硬件在世界各大实验室迅速建立起来,为引力波探测和引力波天文学的研究打下了坚实的基础。随着引力波天文学研究的快速发展,引力波数据分析必将迎来蓬勃发展的新阶段。

#### 参考文献:

- [1] Einstein A. Sitzungsberichte der Physikalisch-mathematischen Klasse. Berlin: Preuss. Akad Wiss, 1916: 688
- [2] Einstein A. Sitzungsberichte der Physikalisch-mathematischen Klasse. Berlin: Preuss. Akad Wiss, 1918: 154
- [3] Schutz B F. *Classical and Quantum Gravity*, 1999, 16(12A): 131
- [4] Marx J, Danzmann K, Hough J, et al. <http://arxiv.org/ftp/arxiv/papers/1111/1111.5825.pdf>, gr-qc/1111.5825, 2014
- [5] Christesen N, Meyer R. *Phys. Rev. D*, 2001, 64: 022001
- [6] Christensen N, Meyer R, Libson A. *Classical and Quantum Gravity*, 2004, 21: 317
- [7] Rover C, Meyer R, Christensen N. *Classical and Quantum Gravity*, 2006, 23: 4895
- [8] Rover C, Meyer R, Christensen N. *Phys. Rev. D*, 2007, 75: 062004
- [9] van der Sluys M V, Röver C, Stroeer A, et al. *ApJ* 2008, 688(2): L61
- [10] vanderSluys M, Raymond V, Mandel I C, et al. *Classical and Quantum Gravity*, 2008, 25(18): 4011
- [11] Raymond V, van der Sluys M V, Mandel I, et al. *Classical and Quantum Gravity*, 2009, 26: 114007
- [12] van der Sluys M, Mandel I, Raymond V, et al. *Classical and Quantum Gravity*, 2009, 26: 204010
- [13] Raymond V, van der Sluys M V, Mandel I, et al. *Classical and Quantum Gravity*, 2010, 27: 114009
- [14] Farr B, Kalogera V, Lujten E. *Phys. Rev. D*, 2014, 90(2): 4014
- [15] Jordan B, Cornish D. *Rev Nucl Part Sci*, 2004, 54: 525
- [16] Anderson W G, Brady P R, Creighton J D E, et al. *Phys. Rev. D*, 2001, 63(04): 2003
- [17] Alan Weinstein, "The physics of LIGO for SURF 2002" LIGO-G020007-00-R; LIGO-G000165-00-R
- [18] Cannon K, Cariou R, Chapman A, et al. *ApJ*, 2012, 748(2): 136
- [19] Chatterji S, Blackburn L, Martin G, et al. *Classical and Quantum Gravity*, 2004, 21(20): S1809
- [20] Waldman S J. <http://arxiv.org/pdf/1103.2728v1.pdf>, gr-qc/1103.2728, 2014
- [21] Saulson P. *Fundamentals of Interferometer Gravitational Wave Detectors*, Singapore, World Scientific, 1994
- [22] Davenport Jr W B, et al. *ibid*, pp154
- [23] Hooper S, Chung S, Luan J, et al. *Phys. Rev. D*, 2012, 86(02): 4012
- [24] Van Trees H L. *Detection, Estimation, and Modulation Theory*. UK: John Wiley, 2004: 3
- [25] Credico A D, the LIGO Scientific Collaboration. *Classical and Quantum Gravity*, 2005, 22: S1051
- [26] Ajith P, Hewitson M, Smith J R, et al. *Phys. Rev. D*, 2007, 76(04): 2004
- [27] Blackburn L, Cadonati L, Caride S, et al. *Classical and Quantum Gravity*, 2008, 25(18): 4004
- [28] The LIGO Scientific Collaboration, The Virgo Collaboration. *JPCS*, 2010, 243(01): 2005
- [29] Christensen N, the LIGO Scientific Collaboration, the Virgo Collaboration. *Classical and Quantum Gravity*, 2010, 27(19): 4010

- [30] Slutsky J, Blackburn L, Brown D A, et al. *Classical and Quantum Gravity*, 2010, 27(16): 5023
- [31] Smith J R, Abbott T, Hirose E, et al. *Classical and Quantum Gravity*, 2011, 28(23): 5005
- [32] The LIGO Scientific Collaboration, The Virgo Collaboration. *Classical and Quantum Gravity*, 2012, 29(15): 5002
- [33] Blackburn L. *Open Issues in the Search for Gravitational Wave Transients*, Ph.D. thesis, Massachusetts: MIT, 2010
- [34] Abbott B P, The LIGO Scientific Collaboration. *Phys. Rev. D*, 2009, 80(10): 2001
- [35] Biswas R, Blackburn L, Cao J, et al. *Phys. Rev. D*, 2013, 88(06): 2003
- [36] Smith A. *Biometrika*, 1975, 62: 407
- [37] Pai A, Dhurandha S, Bose S. *Data-analysis strategy for detecting gravitational-wave(2001)*
- [38] Cutler C, Schutz B. *Phys. Rev. D*, 2005, 72(06): 3006
- [39] Allen B. *Phys. Rev. D*, 2005, 71: 062001
- [40] Gürsel Y, Tinto M. *Phys. Rev. D*, 1989, 40: 3884
- [41] Chung S K, Wen L, Blair D, et al. *Classical and Quantum Gravity*, 2010, 27: 135009
- [42] Waistein L A, Zubakov V D. *Extraction of signals from noise*, *European Science Notes*, 1970
- [43] Liu Y, Du Z, Chung S K, et al. *Classical and Quantum Gravity*, 2012, 29(23): 235018.
- [44] Hecht-Nielsen R, in *Proceedings of International Joint Conference on Neural Networks*, IEEE, New York, Washington, 1989: 593
- [45] Breiman L. *Maching Learning*, <http://dx.doi.org/10.1023/A:1018054314350>, 2014
- [46] Breiman L. *Random Forests: Machine Learning*, <http://link.springer.com/article/10.1023%2FA%3A1010933404324>, 2014

## The Data Analysis in Gravitational Wave Detection

WANG Xiao-ge<sup>1</sup>, ERIC Lebigot<sup>1</sup>, DU Zhi-hui<sup>1</sup>, CAO Jun-wei<sup>1</sup>,  
WANG Yun-yong<sup>2</sup>, ZHANG Fan<sup>2</sup>, CAI Yong-zhi<sup>2</sup>, LI Mu-zi<sup>2</sup>, ZHU Zong-hong<sup>2</sup>,  
QIAN Jin<sup>3</sup>, YIN Cong<sup>3</sup>, WANG Jian-bo<sup>3</sup>, ZHAO Wen<sup>4</sup>, ZHANG Yang<sup>4</sup>,  
DAVID Blair<sup>5</sup>, JU Li<sup>5</sup>, ZHAO Chun-nong<sup>5</sup>, WEN Lin-qing<sup>5</sup>

(1. Tsinghua University, Beijing 100084, China; 2. Department of Astronomy, Beijing Normal University, Beijing 100875, China; 3. Chinese Academy of Metrology, Beijing 100013, China; 4. University of Science and Technology of China, Hefei 230026, China; 5. University of Western Australia, WA 6009, Australia)

**Abstract:** Gravitational waves are an important prediction of general relativity, made in 1916 by Einstein. Yet, the deformation that they imprint on space has never been observed directly. The first such observation should soon be made with pulsar timing arrays or large laser interferometers, which are emphasized in this paper. The advanced interferometers will be fully operational in a few years. We review many important data analysis techniques used for these instruments.

In the same way as the exploration of the electromagnetic spectrum opened new windows on the universe (radio waves, infrared, visible spectrum, ultraviolet, X-rays, gamma

rays, etc.), the ability to sense gravitational waves will open a new kind of astronomy. In fact, the shape of a gravitational wave encodes information about its physical origin, be it the coalescence of two very dense objects (neutron stars and/or black holes), a supernova explosion, or other phenomena.

Gravitational waves are very weak: the movement that the largest laser interferometers (LIGO and Virgo) are designed to observe has an amplitude of only about  $10^{-19}$  m, which is measured as the relative variation (strain) of the length of arms that measure 3 to 4 kilometers. Achieving this requires a host of physical measurements that flow from sensors to computers. These measurements are analyzed, displayed and stored through a complex hardware and software system that are presented in this paper.

Such a high precision measurement requires extremely sensitive experimental and data analysis techniques. The signal that continuously comes from a gravitational wave interferometer can be observed either as a function of time, or in the time-frequency domain, since this representation is adapted to the oscillatory nature of the gravitational waves. Different data analysis techniques are appropriate for each representation. For the time domain, we thus discuss matched filtering techniques, which can find a small signal of known form buried in noise. The time-frequency domain is presented in conjunction with techniques used for detecting bursts of gravitational waves, and unwanted glitches in the interferometer signal.

Since multiple large gravitational wave interferometers are being made available, it is natural to take advantage of their conjoint measurements: a gravitational wave should be seen on all running detectors (with an amplitude that depends on their orientation), whereas glitches in the interferometer signal should generally not happen at the same time in the same way. This forms the basis of the coincidence and coherent data analysis methods presented here. Other important computing techniques such as Monte-Carlo calculations and  $\chi^2$  for detecting gravitational waves and estimating their parameters (direction of their source, etc.) are also presented.

**Key words:** gravitational wave; laser; interferometer; data analysis