

# 引力波数据处理

胡一鸣

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$



git commitID: f6b8ca30230a13cf92bc7a3b235e3666e9ad3682

编译日期: Mon Mar 13 13:49:48 2023 +0800

任何建议及错误信息请发送至邮箱

[huyiming@sysu.edu.cn](mailto:huyiming@sysu.edu.cn)



本作品采用知识共享署名 4.0 国际许可协议进行许可。访问<https://creativecommons.org/licenses/by/4.0/>查看该许可协议。  
本作品  $\text{\LaTeX}$  模板采用<sup>a</sup>，遵守知识共享署名-非商业性使用 4.0 国际许可协议进行许可。访问<https://creativecommons.org/licenses/by-nc/4.0/>查看该许可协议。

<sup>a</sup><https://github.com/muzimuzhi/Qbook>



# 序

Joseph Weber 是引力波探测的先驱。他建造了两个相隔上千公里的探测器，这样就可以有效地隔绝环境噪声产生的误警（实际上，LIGO 也是遵从这样的宗旨选址的）——这是因为，如果两个探测器相隔很远，它们的噪声可以看作相互独立，而引力波以光速传播，所以，当两个探测器上同时测量到信号时，极有可能是来自引力波信号而非噪声。在 Weber 记录的数据中，就有几个信号，不同的探测器间所记录的时间差非常短。1969 年，Weber 在 *Physical Review Letters* 上发表论文 *Evidence for discovery of gravitational radiation*，宣称发现了引力波信号。一时间，他名声鹊起，成为了聚光灯前的宠儿，媒体将其捧为继 Einstein 之后最伟大的物理学家。

1993 年，两位天文学家因为引力波效应的间接证据而获得诺贝尔奖；2017 年，三位物理学家由于引力波的发现而获得诺贝尔奖。这些人里，都没有 Weber。当然，斯人已去，Weber 在世纪之交的 2000 年 9 月 30 日就已作古。然而，Weber 与诺奖失之交臂，并不是因为他活的不够长——这一点日渐成为获得诺奖的必要非充分条件——而是因为后续的研究无法确认 Weber 的结论。首先是 Weber 宣称的引力波信号在强度、频次上都与理论计算结果与相差悬殊；其次，其他团队建造的更灵敏的探测器宣布无法探测到引力波事件；最致命的是，有人发现，Weber 在相隔三个时区的两个信号中寻找同时信号的时候，竟然忘了考虑时差。

Weber 是真诚的，直到他去世时，他依然坚信他成功地探测到了引力波。在这里，并没有学术不端；有的，只是对数据处理的极度忽视。这显然是令人十分遗憾的。不过，设若 Weber 的数据处理地好一些，也许就不会有什么惊人的发现，也就不会吸引到这么多聪明人投身到引力波探测这个领域，也许 2015 年 9 月 14 日穿越地球而过的那一串小小涟漪，就会如它之前所有的双黑洞并合信号一般，悄悄地来，又悄悄地离开，不留下一丝痕迹。

2016 年，引力波探测的大门，已然被叩开，黑洞与中子星疯狂的舞蹈终于觅来了知音。在 2019 年的这个夏天，当我远眺未来，充满的是憧憬和期待，当天琴卫星上天，当宇宙的长波电台被天琴接收到信号时，又会上演什么样的一出好戏呢？科学的魅力往往就在于它的不可捉摸和不可预测。不管是怎样的发现，背后一定会有着有力的数据处理方法作为支撑。好在，卫星上会搭载原子钟用于时钟同步，不至于会忘记考虑时差。

历史不容假设，当我们回望过去的时候，我们也许会将 Weber 铭记为一位勇敢的先

驱，一位卓绝的工程师，甚或是一位天才的实验物理学家；但同时，不容否认，他在引力波数据处理这门课上，表现地糟糕透了。我相信，当未来的人们回望这段当前，回顾中国科学的崛起时，“引力波探测”、“天琴计划”，都会是绕不过去的、在历史的长河中熠熠发光的名词。我们有着国内顶尖的激光团队，国内顶尖的惯性基准团队，和国内顶尖的卫星系统团队。我希望，我们也将会有国内顶尖的引力波数据处理团队。我更希望，这本讲义能成为奔向这个远方的起点。

“夫江始出于岷山，其源可以滥觞”。这段历史画卷，现在就将由你谱写。愿这一本讲义，能成为这一切的波源。

# 目 录

序	i
前 言	ix
<b>第一章 相对论基础</b>	<b>1</b>
1.1 相对性原理 (Principle of relativity) . . . . .	2
1.1.1 Galilean 相对论 . . . . .	2
1.1.2 Maxwell 方程组 . . . . .	3
1.1.3 狹义相对论 . . . . .	4
1.2 微分几何初步 . . . . .	5
1.2.1 协变 (covariant) 与逆变 (contravariant) . . . . .	6
1.2.2 张量 (tensor) 初步 . . . . .	8
1.2.3 等效原理 . . . . .	10
1.2.4 协变导数 . . . . .	11
1.2.5 测地线 (geodesics) . . . . .	14
1.3 广义相对论 (General Relativity) 初步 . . . . .	15
1.3.1 曲率张量 . . . . .	15
1.3.2 Einstein 场方程 . . . . .	15
<b>第二章 引力波方程</b>	<b>19</b>
2.1 广义相对论的 Newtonian 极限 . . . . .	19
2.1.1 线性化引力 (linearized gravity) . . . . .	19
2.1.2 Newtonian 极限 . . . . .	22
2.2 引力波 . . . . .	23
2.2.1 引力波的传播 . . . . .	23
2.2.2 橫向无迹规范 (transverse-traceless gauge) . . . . .	24

2.3	引力波的物理性质 . . . . .	25
2.3.1	引力波的效果 . . . . .	25
2.3.2	探测原理初探 . . . . .	28
2.3.3	引力波携带的能量 . . . . .	30
2.3.4	引力波能流 . . . . .	32
<b>第三章</b>	<b>引力波源和波形</b>	<b>33</b>
3.1	双星系统的引力波 . . . . .	33
3.1.1	引力波与四极矩 . . . . .	33
3.1.2	双星系统引力波初探 . . . . .	35
3.2	双黑洞并合 . . . . .	37
3.2.1	双黑洞旋近 . . . . .	38
3.2.2	数值相对论与并合 . . . . .	43
3.2.3	QNM 与铃宕 . . . . .	44
3.2.4	EOB 波形与 IMR 波形 . . . . .	44
3.2.5	BBH 天文学 . . . . .	45
3.2.6	BBH 与基础物理 . . . . .	52
3.3	中子星双星并合 . . . . .	53
3.3.1	中子星 . . . . .	53
3.3.2	中子星双星天文学 . . . . .	56
3.4	EMRI . . . . .	62
3.4.1	天文学背景 . . . . .	62
3.4.2	波形 . . . . .	66
3.4.3	天琴探测能力 . . . . .	69
3.5	Continuous Wave . . . . .	69
3.5.1	变形的中子星 . . . . .	70
3.5.2	双白矮星 . . . . .	72
3.6	SGWB . . . . .	76
3.7	其他波源 . . . . .	79
3.7.1	Supernova . . . . .	79
3.7.2	引力波暴 . . . . .	79
<b>第四章</b>	<b>引力波探测手段</b>	<b>83</b>
4.1	棒状探测器 . . . . .	84

4.1.1	原理	84
4.1.2	天线响应函数	85
4.1.3	噪声	88
4.2	地面激光干涉探测器	90
4.2.1	原理	90
4.2.2	噪声	94
4.2.3	灵敏度	96
4.3	空间引力波探测	97
4.3.1	原理	97
4.3.2	噪声	102
4.4	脉冲星计时阵	108
4.4.1	原理	108
4.5	宇宙微波背景辐射	110
4.5.1	原理	110
4.6	其他探测方案	110
<b>第五章</b>	<b>“穷人版”引力波数据处理示例</b>	<b>111</b>
5.1	从观测数据内提取信息	112
5.1.1	白化 (Whitening)	112
5.1.2	带通滤波 (Bandpass)	113
5.2	论证引力波源为双黑洞系统	115
5.3	讨论假设的合理性以及对于结论的影响	117
5.3.1	非零轨道偏心率	118
5.3.2	不等质量	118
5.3.3	自旋效应	119
5.3.4	Newtonian 动力学	120
5.3.5	啁啾质量测量	120
5.3.6	估计波源的光度距离	121
5.4	结论	121
<b>第六章</b>	<b>引力波信号探测</b>	<b>123</b>
6.1	概率初步	124
6.1.1	概率的性质	124
6.1.2	频率学派 v.s.Bayesian	126

6.1.3	典型分布 . . . . .	127
6.2	时序列分析 . . . . .	128
6.2.1	相关函数 . . . . .	129
6.2.2	卷积 . . . . .	129
6.2.3	Fourier 变换 (Fourier Transform) . . . . .	129
6.2.4	功率谱密度 . . . . .	131
6.2.5	周期图 . . . . .	133
6.3	信号探测的统计学原理 . . . . .	133
6.3.1	假设检验 (hypothesis testing) . . . . .	134
6.4	连续引力波探测 . . . . .	140
6.4.1	$\mathcal{F}$ -统计 . . . . .	140
6.4.2	误警率和探测概率 . . . . .	146
6.4.3	次优滤波 . . . . .	146
6.5	啁啾信号探测 . . . . .	146
6.5.1	new SNR . . . . .	146
6.5.2	匹配滤波的统计学性质 . . . . .	148
6.5.3	时间未知的匹配滤波 . . . . .	149
6.5.4	匹配滤波模板库 . . . . .	149
6.5.5	同时性 . . . . .	150
6.5.6	显著度分析 . . . . .	151
<b>第七章</b>	<b>引力波信号测量</b> . . . . .	<b>155</b>
7.1	参数估计 . . . . .	155
7.1.1	测量精度 . . . . .	155
7.1.2	参数估计中的系统误差 . . . . .	157
7.1.3	置信区域 . . . . .	158
7.1.4	冗余参数 . . . . .	159
7.2	Markov 链蒙特卡洛 . . . . .	159
7.2.1	并行回火 MCMC (Parallel Tempering MCMC) . . . . .	162
7.2.2	延迟拒绝 MCMC (Delayed Rejection MCMC) . . . . .	162
7.2.3	可逆跳 MCMC(Reversible Jump MCMC) . . . . .	163
7.2.4	自动化 MCMC(Automated MCMC) . . . . .	163
7.2.5	仿射不变的 MCMC (Affine Invariant MCMC) . . . . .	164
7.2.6	组合 . . . . .	164

7.3	模型选择与 Occam 剃刀 . . . . .	164
7.3.1	Occam 剃刀 . . . . .	165
7.3.2	Odds ratio . . . . .	165
7.3.3	定量的 Occam 剃刀：简单推导 . . . . .	166
7.4	嵌套抽样 . . . . .	168
7.4.1	MultiNest . . . . .	170
<b>第八章 引力波数据处理实例</b>		<b>171</b>
<b>附录 A 常用物理量</b>		<b>173</b>
<b>参考文献</b>		<b>175</b>

2023-03-13

# 前 言

这份讲义是我为中山大学物理与天文学院研究生课程“引力波数据处理”课程所准备的。由于本人能力有限，准备时间仓促，一定包含了大量错误，我会力争在收到反馈之后进行更正。感谢谢思远、黄顺佳、刘丽璇、周立祥、张鑫、梅健伟、童晖、郑江锦、王晗、汪绍涵、陈洪昱、李知圆、杜宝诚、吴征、王璐、李亚南、王雨石、等指出部分错误。感谢李博宏等提出的宝贵意见。本书第3.2.5.2中借鉴了王海天的硕士论文、在第3.5.2节、第4.3.1节及第6.4.1.1节中借鉴了黄顺佳的硕士论文、在第3.6节借鉴了梁正程撰写的技术文档，在第3.2.3节借鉴了石常富博士的博士论文。第五章原为2019学年的课程大作业布置，现选取吕翔宇的高分作业作为蓝本。这里一并感谢。

讲义的电子版可以在[https://github.com/yiminghu-SYSU/GW\\_DA\\_notes](https://github.com/yiminghu-SYSU/GW_DA_notes)获得。

这本讲义的写作对象是对引力波数据处理感兴趣的高年级本科生或研究生。本书默认读者已经初步掌握了狭义相对论的基本概念，并有一定概率论基础。在讲义中，我会尽可能追求内容的完整性，以便尚未完成广义相对论等课程学习的同学也可以在脑中构建起足够的物理图景。

但限于篇幅和授课计划，本书肯定无法替代广义相对论等基础课程。因此，文中肯定会在数学逻辑的严谨性和课程主题的完备性之间做出倾向于后者的取舍，也敬请诸位谅解。

胡一鸣

2019年8月14日，于珠海唐家

2023-03-13

2023-03-13

# 1

## 相对论基础

本讲义的授课主题，是引力波数据处理，共分为两部分，即引力波与数据处理。如果脱离了引力波的物理图景，而直接空谈数据处理，未免空中楼阁。而在另一方面，引力波又是 Einstein 广义相对论的直接理论预言，因此，引力波的理论描述，无法跳脱广义相对论的框架。

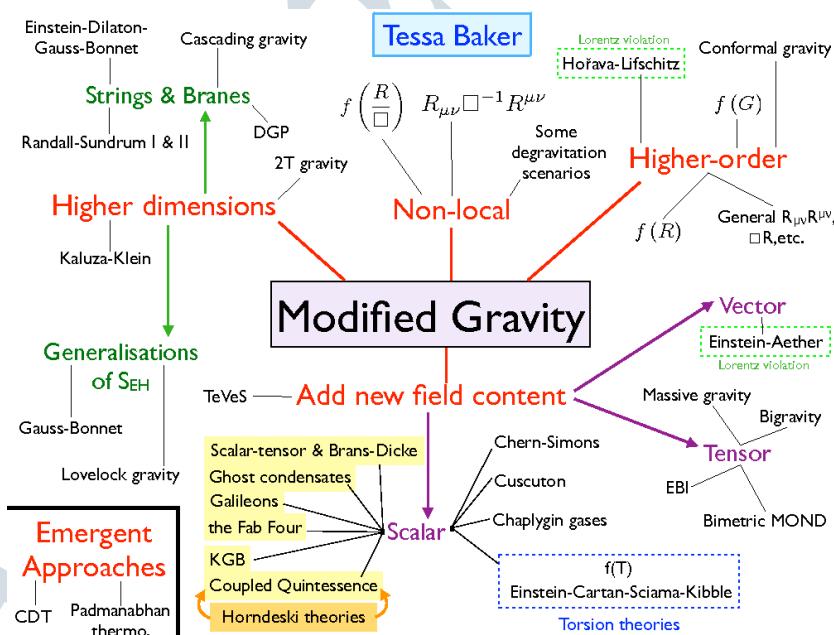


图 1-1 修改引力理论

Figure 1-1 Theories of modified gravity. Credit: [32]

从 Einstein 至今，引力理论已经有了长足的发展，如图1-1所示，仅基于广义相对论基础上发展起来的修改引力理论就已不计其数。由于和量子力学原理的深刻矛盾，有理由认为 Einstein 决定论性的广义相对论在某个地方一定背离了引力的物理本质。然而，时至今日，Einstein 昔日基于广义相对论所作出的诸多预言，一一被实验所验证；所有可靠的实验检验下，广义相对论均可以给出解释——而它通常是最简洁的那个理论。因此，即使将来的实验证明了广义相对论与引力的物理本质之间的偏离，对广义相对论的理解依然有着重要的意义。

## 1.1 相对性原理 (Principle of relativity)

### 1.1.1 Galilean 相对论

虽然在 20 世纪，相对论一词专指 Einstein 的理论，但是相对性原理 (Principle of relativity) 的思想在 Newtonian 力学中就有体现：两个服从 Newtonian 力学的、相对均匀运动的惯性参考系，是无法通过在内部开展的力学实验加以区分的。这一思想一般认为是 Galileo 在《关于 Ptolemaic 和 Copernican 两大世界体系的对话》中首先提出的 [100]：

#### 材料 1.1 Dialogo sopra i due massimi sistemi del mondo

把你和一些朋友关在一条大船的甲板下的主舱里，让你们带着几只苍蝇、蝴蝶和其他小飞虫，舱内放一只大碗，其中有几条鱼，然后，挂上一个水瓶，让水一滴一滴地滴到下面的一个宽口罐里。船停着不动时，你留神观察，小虫都以等速向舱内各方向飞行，鱼向各方向随便游动，水滴滴进下面的罐中。你把任何东西扔给你的朋友时，只要距离相等，向这一方向也不比向另一方向更多用力。你的双脚齐跳，无论向哪个方向跳过的距离都相等。当你仔细观察这些事情之后，再使船以任何速度前进，只要运动是均匀的，也不忽左忽右地摆动，你将发现，所有上述现象都没有丝毫变化，你无法从任何一个现象来确定，船是在运动还是在停着不动。即使船运动得相当快，在跳跃时，你也将和以前一样，你跳向船尾也不会比跳向船头更省力。

在 Galilean 相对性原理表明的这个表述中，对日常生活中所涉及到的物理学特性而言，在地球坐标系下（船是运动的）和船的坐标系下（船是不动的）没有任何区别。

用公式来表述的话，则可以设立两个坐标系，亦即“静止的” 地球坐标系  $S(t, x, y, z)$  和“运动的” 船坐标系  $S'(t', x', y', z')$ 。不妨令  $t = t' = 0$  时，两个坐标系重合，且船以速

度  $v$  沿  $x$  方向移动，则有

$$\begin{aligned} t' &= t \\ x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned} \tag{1.1}$$

这种变换通常被称为 Galilean 变换。不难得出，Galilean 变换下的速度转换公式是  $u'_x = u_x - v$

实际上，这种朴素的相对论性原理是非常直观的，在《尚书纬·考灵曜 (yào)》中，就有文字表达了相当类似的想法：

### 材料 1.2 尚书纬·考灵曜

地恒动不止而人不知，譬如人在大舟中，闭牖 (yǒu) 而坐，舟行而不觉也

### 1.1.2 Maxwell 方程组

Maxwell 总结了前人归纳的方程组，得到了

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} &= \mu \mathbf{J} \end{aligned} \tag{1.2}$$

其中  $\epsilon$  为介电常量， $\mu$  为磁导率。尤其，他注意到这组方程与连续性公式（如下）冲突

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{1.3}$$

在放弃实验总结公式和放弃连续性方程之间，Maxwell 作出了伟大的抉择：他相信连续性方程必须成立。由此，他总结了著名的 Maxwell 方程，用以描述电磁场的一般性质。

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} &= \mu (\mathbf{J} + \epsilon \frac{\partial \mathbf{E}}{\partial t}) \end{aligned} \tag{1.4}$$

而其中  $\epsilon \frac{\partial \mathbf{E}}{\partial t}$  这项被称为位移电流。

Maxwell 注意到，通过 Maxwell 方程组，利用真空中  $\rho = 0$ ,  $\mathbf{J} = \mathbf{0}$  的条件，可以推导出<sup>1</sup>

$$\nabla^2 \mathbf{B} - \varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} = 0. \quad (1.5)$$

不难看出，电磁场的变化以波动形式传播，其在真空中传播的速度  $c$  取决于：

$$c^2 = \frac{1}{\varepsilon_0 \mu_0} \quad (1.6)$$

从数值上看， $c = 1/\sqrt{\varepsilon_0 \mu_0}$  的取值与当时已经从实验上测得的光速极为接近，这使得他大胆假设：光就是一种电磁波。

然而，Maxwell 方程组与 Galilean 变换是不自洽的。考虑在运动的船  $S'$  上进行电磁学测量，根据 Galilean 变换<sup>1.1</sup>，真空中电磁场的传播方程<sup>1.5</sup>变为

$$c^2 \nabla^2 \mathbf{B} = \frac{\partial^2 \mathbf{B}}{\partial t^2} + (\mathbf{v} \cdot \nabla)^2 \mathbf{B} - 2\mathbf{v} \cdot \nabla \left( \frac{\partial \mathbf{B}}{\partial t} \right) \quad (1.7)$$

一个平庸的结论是，通过 Galilean 变换，船上的物理学家将测得光速变为  $c \pm v$ 。然而更深刻的问题是，这一结论意味着，如果 Galilean 相对论是正确的，那么 Maxwell 方程组只能对某个特定惯性参考系成立，而物理学家可以根据电磁场的测量来确定“船”的状态是静止还是运动。Newtonian 力学必须借助绝对空间的概念，在坚持 Galilean 相对论的前提下，似乎可以得出一个有趣的结论：满足<sup>1.5</sup>的参考系就是 Newtonian 力学概念中的绝对空间！

其时，人们认为电磁波传播需要介质，而这种依附于绝对空间而具有独特性质的参考系，具象化为“以太 (aether)” [99]。

### 1.1.3 狹义相对论

Lorentz 和 Poincaré 第一次意识到，如果  $S$  坐标系和  $S'$  坐标系之间的转换关系采用如

$$\begin{aligned} t' &= \gamma \left( t - \frac{vx}{c^2} \right) \\ x' &= \gamma (x - vt) \\ y' &= y \\ z' &= z \end{aligned} \quad (1.8)$$

的形式的话，那么 Maxwell 方程组（公式<sup>1.4</sup>）在所有的惯性系中都能成立。这里的  $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$  通常被称为 Lorentz 因子。注意由于  $\gamma \geq 1$  恒成立，对于固定在  $x = 0$  的时钟，

<sup>1</sup>注意  $\varepsilon_0$ 、 $\mu_0$  分别表示真空中的介电常量和磁导率。

运动的  $S'$  坐标系中的观测者会发现时钟会变慢,  $t' = \gamma t$ , 这就是所谓的时间膨胀 (time dilation)。考虑一根长度为  $L$  的尺子, 而对于  $S$  坐标系的观测者而言, 在  $t = 0$  时刻, 其一端 (首) 放置于  $x = 0$ , 另一端 (尾) 为  $x = L$ 。在  $S'$  坐标系中的观测者看来, 则其一端 (首) 为  $t' = 0, x' = 0$ 。为了保证“同时”测量, 则对于另一端 (尾), 也要保证  $t' = 0$ 。代入  $x = L$  后, 可以得到, 在  $S'$  系里的  $t' = 0$  时测量  $x = L$  的事件, 在  $S$  坐标系看来发生于  $t = \frac{vL}{c^2}$ , 这时对应的另一端 (尾) 为  $x' = L/\gamma$  处, 这就是所谓的 Lorentz 收缩 (Lorentz contraction)。

两个效应通常合称为狭义相对论的钟慢尺缩效应。

形如公式1.8的变换称为 Lorentz 变换 (Lorentz transformation), 我们可以说, Maxwell 方程组在不同的惯性系中, 通过 Lorentz 变换维持了不变性。可以发现, Lorentz 变换下的速度转换公式是

$$u'_x = \frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma(dt - \frac{vdx}{c^2})} = \frac{\frac{dx}{dt} - v}{1 - (\frac{v}{c^2})(\frac{dx}{dt})} = \frac{u_x - v}{1 - u_x v/c^2}. \quad (1.9)$$

在 Maxwell 方程组的有效性和 Galilean 相对论之间, Einstein 作出了伟大的抉择: 他相信光速不变必须成立, Lorentz 变换背后对应着深刻的物理原理。

从某种意义上说, Lorentz 变换就是狭义相对论的精髓。但物理学界今天达成共识, 认为是 Einstein 而非 Lorentz (或 Poincaré) 提出了狭义相对论, 这是因为 Einstein 第一次严肃地认为 Lorentz 变换体现了时间与空间的本质, 而非简单的数学工具。由此, 时间与空间并非完全独立, 而是交织在一起, 甚至可以互换。要标记一个事件, 必须同时标记其在某个惯性系下的空间坐标  $(x, y, z)$  和时间坐标  $t$ 。对于两个事件, 在  $S$  坐标系下看来可能是同时发生的 ( $\Delta t = 0$ ), 但在  $S'$  坐标系看来却可能发生在不同时间 ( $\Delta t \neq 0$ )。在不同的惯性系下, 通过 Lorentz 变换, 两个不同事件之间, 保持不变的, 是两个事件的时间间隔和空间间隔的某种组合, 称为时空间隔 (spacetime interval): <sup>1</sup>

$$\Delta s^2 = -(c\Delta t)^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \quad (1.10)$$

通常其微分形式使用起来更为方便:

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \quad (1.11)$$

特别的, 我们将符合这种描述的时空称为 Minkowski 时空。

## 1.2 微分几何初步

<sup>1</sup>注意, 时空间隔指  $\Delta s^2$ , 可正可负, 不能将其理解成某个  $\Delta s$  的平方。

### 1.2.1 协变 (covariant) 与逆变 (contravariant)

在相对论框架下，时间坐标  $t$  和空间坐标  $(x, y, z)$  联合起来形成一个统一的时空坐标

$$x^\alpha = (ct, x, y, z).^1 \quad (1.12)$$

如此处的  $\alpha$  一般出现在坐标上标上的希腊字母，会遍历  $\{0, 1, 2, 3\}$ 。 $x^0 = ct$  代表时间坐标，而  $x^1 = x, x^2 = y, x^3 = z$  则代表空间坐标。三维空间中的“点的位置”这一概念延伸到狭义相对论中就成了“事件”的时空坐标。

 **注 1.3.** 通常，约定俗成：采用希腊字母——如  $\alpha, \beta, \mu, \nu, \rho, \tau$  等——时，默认遍历  $\{0, 1, 2, 3\}$ ；采用英语字母——如  $i, j, k, m, n$  等——时，默认遍历  $\{1, 2, 3\}$ ，即只包含空间项。

这种用一组坐标来表征一个特定点的做法，大家在中学、大学的课程中学习过矢量的概念后，相信并不会感觉陌生。一个矢量，在选定一组坐标基底后，可以用不同基底上的投影分量去表示；而矢量相关的公式，也可以用不同分量展开来表示。例如，将电场 **E** 分解成不同正交坐标基底  $x, y$  和  $z$  上的分量  $E_x, E_y$  和  $E_z$ ，则 Maxwell 方程组（公式 1.4）中的第一个公式 Gauss 定律可以写成

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho}{\epsilon_0}. \quad (1.13)$$

将公式写成矢量计算的形式，不仅有简化表达式的功能，更重要的是它揭示了对应等式后面的一个重要结论：与矢量相关的一些结论，是矢量本身的固有性质，与具体的坐标选取没有关系。比如，两个矢量之间的点乘，和矢量之间的夹角有关，和计算过程中采用哪个坐标系并不相关。

不妨考虑一个在平直的四维时空中的矢量  $\vec{A}$  也可以写成

$$\vec{A} = \sum_{\alpha} A^{\alpha} \vec{e}_{\alpha} \equiv A^{\alpha} \vec{e}_{\alpha} \quad (1.14)$$

也就是说，矢量  $\vec{A}$  在基矢  $\vec{e}_{\alpha}$  上的投影分量为  $A^{\alpha}$ 。注意，根据 **Einstein 求和约定 (Einstein summation convention)**，当某个希腊字母同时出现在上下标的时候，自动对该字母求和，故此后文中将省去求和符号  $\Sigma$ 。在 Einstein 求和约定下，只要上下标相同，不管用哪个符号标记都没有区别，因为总会被消掉。所以形如 1.14 式中的  $\alpha$  通常被称为**哑指标**或**赝指标**。

<sup>1</sup>注意，此处指标记在上方和之后将基底的坐标标记在下方均有特殊意义，我们稍后会解释。

我们把原先的坐标系称为  $S$  坐标系，在新的坐标系  $\bar{S}$  中，坐标基底更换为  $\vec{e}_{\bar{\alpha}}$ ，则矢量  $\vec{A}$  在新的基矢  $\vec{e}_{\bar{\alpha}}$  上的新投影分量为  $A^{\bar{\alpha}}$ ，不难得到

$$\vec{A} = A^\alpha \vec{e}_\alpha = A^{\bar{\alpha}} \vec{e}_{\bar{\alpha}}. \quad (1.15)$$

换言之，选择不同的基底，不会改变矢量本身的特性。那么如何描述不同基底上的分量  $A^\alpha$  与  $A^{\bar{\alpha}}$  之间的关系呢？回顾 Lorentz 变换（公式1.8），可以意识到，在狭义相对论下，不同的参考系之间的坐标量的转换关系是线性变换，所以可以写成

$$A^{\bar{\alpha}} = \Lambda_{\beta}^{\bar{\alpha}} A^\beta \quad (1.16)$$

的形式。<sup>1</sup> 在一般情形下，有

$$\Lambda_{\beta}^{\bar{\alpha}} = \frac{\partial x^{\bar{\alpha}}}{\partial x^\beta} \quad (1.17)$$

 **注 1.4.** 可以看到，通过式1.17中， $\beta$  从  $x^\beta$  中的上指标变成  $\Lambda_{\beta}^{\bar{\alpha}}$  中的下指标。

特别地，我们把分量变换规则符合公式1.16的矢量称为**逆变 (contravariant)** 矢量。如果将公式1.16代入到公式1.15中，可以得到

$$A^\alpha \vec{e}_\alpha = A^{\bar{\alpha}} \vec{e}_{\bar{\alpha}} = \Lambda_{\beta}^{\bar{\alpha}} A^\beta \vec{e}_{\bar{\alpha}}. \quad (1.18)$$

其中，求和顺序可交换。由于  $\alpha$  和  $\beta$  都是哑指标，可以将其互换（这是一种在相对论计算中特别有用的技巧）我们有

$$A^\alpha \vec{e}_\alpha = A^\beta \Lambda_{\beta}^{\bar{\alpha}} \vec{e}_{\bar{\alpha}} = A^\alpha \Lambda_{\alpha}^{\bar{\beta}} \vec{e}_{\bar{\beta}}. \quad (1.19)$$

由此，我们得到恒等式

$$\vec{e}_\alpha = \Lambda_{\alpha}^{\bar{\beta}} \vec{e}_{\bar{\beta}}. \quad (1.20)$$

为了更好地与逆变矢量的变换关系（公式1.16）作比较，不妨将  $S$  系和  $\bar{S}$  系的地位互换，则有

$$\vec{e}_{\bar{\alpha}} = \Lambda_{\bar{\alpha}}^{\beta} \vec{e}_\beta. \quad (1.21)$$

不难发现，变换坐标系后，基底的转换公式和逆变矢量的转换公式是不同的。

除了符合逆变关系的矢量以外，还有一种“矢量”，其坐标分量在坐标系转变下不符合逆变关系（公式1.16）。类似地，我们可以通过引入基底  $\tilde{\omega}$ ，可以实现对这种“矢量”的分解：

$$\tilde{p} = \sum_{\alpha} p_\alpha \tilde{\omega}^\alpha \equiv p_\alpha \tilde{\omega}^\alpha \quad (1.22)$$

<sup>1</sup> 具体的表达式，请参照习题1.1自行推导。

如果在坐标系转换下，分量可以满足

$$p_{\bar{\alpha}} = \Lambda^{\beta}_{\bar{\alpha}} p_{\beta} \quad (1.23)$$

则我们将  $\tilde{p}$  称为**协变 (covariant)** 矢量。“协变”指的是其分量的变化规律与基底变化规律（公式1.21）一致。

 **注 1.5.** 有些时候，也把  $\tilde{p} = p_{\alpha} \tilde{\omega}^{\alpha}$  称为余向量 (*covector*) , 1-形式 (*one-form*)，或者对偶矢量 (*dual vector*)。

让我们考虑标量场  $\phi(\vec{x})$ ，可以定义  $\phi$  的梯度  $\tilde{d}\phi$ ，其分量为  $\left( \frac{\partial \phi}{\partial t}, \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial z} \right)$ 。不难发现，由于链式法则

$$\frac{\partial \phi}{\partial x^{\bar{\alpha}}} = \frac{\partial \phi}{\partial x^{\beta}} \frac{\partial x^{\beta}}{\partial x^{\bar{\alpha}}} = \Lambda^{\beta}_{\bar{\alpha}} \frac{\partial \phi}{\partial x^{\beta}} \quad (1.24)$$

因此，标量场的梯度是协变矢量。

## 1.2.2 张量 (tensor) 初步

公式1.11表明了，无论坐标基底如何选取， $ds^2$  总是不变的，我们称为标量。这种取值不依赖于坐标选取的特性，与矢量如出一辙。我们将这类本身性质不随坐标选取改变而发生变化的量，统称为**张量 (tensor)**。根据分量指标数的多寡，我们将标量（分量即  $s$ ）称为零阶张量，逆变矢量（分量即  $A^{\alpha}$ ）和协变矢量（分量即  $p_{\alpha}$ ）为一阶张量，而具有  $t_{\mu\nu}$  形式的分量具有两个指标，定义为二阶张量。在给定坐标基底后，可以将二阶张量的分量以矩阵形式表述。

在矢量空间中，两个矢量的点乘会得到一个与坐标系选取无关的数，所以矢量的点乘将得到标量。在 Minkowski 时空中设有位移矢量  $d\vec{x} = \vec{e}_{\alpha} dx^{\alpha}$ ，我们定义其与自身的内积为

$$ds^2 = d\vec{x} \cdot d\vec{x} = (\vec{e}_{\alpha} \cdot \vec{e}_{\beta}) dx^{\alpha} dx^{\beta}. \quad (1.25)$$

可以定义

$$\eta_{\alpha\beta} \equiv (\vec{e}_{\alpha} \cdot \vec{e}_{\beta}) \quad (1.26)$$

将两式结合，有

$$ds^2 = \eta_{\alpha\beta} dx^{\alpha} dx^{\beta} \quad (1.27)$$

通过与公式1.11比对，可以得到

$$\eta_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.28)$$

注意到  $\eta_{00} = -1$ , 而  $\eta_{11} = \eta_{22} = \eta_{33} = 1$ 。我们通常把  $\eta_{\alpha\beta}$  称为 Minkowski 时空中的度规张量 (metric tensor, 在数学语境中通常翻译为度量张量), 或者简称为 Minkowski 度规。

 **注 1.6.** 在引力研究领域通常将度规的号差选择为  $(-, +, +, +)$ , 而在粒子物理领域往往选为  $(+, -, -, -)$ , 请读者注意区分。

### 材料 1.7 Kronecker delta

有一个特殊的张量, 即所谓 Kronecker delta 张量

$$\delta_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.29)$$

该张量的表现形式在所有的坐标系中都一致。

在更一般的情形下, 可以写成

$$ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta \quad (1.30)$$

当坐标  $x^\alpha$  发生变化时, 度规张量  $g_{\alpha\beta}$  相应变化, 使得时空间隔  $ds^2$  维持不变量。也就是说, 通过公式 1.30, 我们可以把随坐标变换而改变的  $x^\alpha$  转化成不随坐标变换而改变的时空间隔  $ds^2$ 。<sup>1</sup>

 **注 1.8.** 注意, 平直时空 (Minkowski 时空) 并不一定意味着 Minkowski 度规。如, 若空间部分采用球坐标系, 则有坐标基底为  $\{ct, r, \theta, \phi\}$ , 这时时空间隔可以写为

$$ds^2 = -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \quad (1.31)$$

亦即, 对应的度规为

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix} \quad (1.32)$$

<sup>1</sup> 我们规定度规张量总是对称的, 也就是  $g_{\alpha\beta} = g_{\beta\alpha}$ 。

### 1.2.3 等效原理

Einstein 的狭义相对论虽然可以将 Newtonian 框架下的绝对空间观点移除, 但惯性系仍然在所有的坐标系中占据一个特殊的位置。<sup>1</sup> 由此探究了十年时间之后, Einstein 最终得到了广义相对论, 并最终将惯性系的特殊性也彻底移除。Einstein 观察到, 在 Newtonian 力学体系中, 质量  $m$  这个概念出现在两个完全截然不同的地方: Newton's second law of motion 指出, 物体所受的力  $\vec{F} = m\vec{a}$ , 这里  $\vec{a}$  为加速度, 由此可以定义出惯性质量  $m_I$ 。而在万有引力公式中, Newton 又指出, 物体所受的引力与其质量大小成正比, 由此定义了引力质量  $m_G$ 。两者的语境完全不同, 而实验可以证明在实验精度内  $m_I = m_G$  精确成立。两种质量的严格相等, 通常被称为“弱”等效原理 (weak equivalence principle, WEP)。<sup>2</sup>

$m_I$  与  $m_G$  的相等, 是物理性质的重合还是源于背后物理机理的等价性? 对此, Einstein 又作出了伟大的抉择: 他相信弱等效原理并非巧合, 而引力实际上等效于加速度。拥有相同质量的物体, 可以拥有完全不同的电量, 由此在电场中受到不同的电磁力作用。然而拥有相同惯性质量  $m_I$  的物体, 一定拥有相同的引力质量  $m_G$ , 这也就暗示了引力作用从本质上和其他相互作用的区别。在狭义相对论的语境中, 物理学家无法通过任何局部开展的实验了解自己所处的船舱是否处于均匀运动状态。而弱等效原理中, 物理学家无法通过实验分辨船舱是在加速 (由  $m_I$  确定, 服从 Newton's second law of motion) 还是受到了引力作用 (由  $m_G$  确定, 服从引力作用)。既然加速度和引力在局域上等价, 那么处于自由下落状态的实验室, 所开展的所有局域 (local) 实验结果, 都将完全一样, 这一点与实验室的速度、所处的位置都不相关。这一等价性, 被称为“强”等效原理 (strong equivalence principle, SEP)。[86]

 **注 1.9.** 严格来说, 上述原理实际上是 Einstein 等效原理 (*Einstein equivalence principle, EEP*), 因为只考虑了外界引力场对船舱的影响。而严格的强等效原理指的是即使考虑船舱中实验对象的引力效应 (自引力), 仍然无法局域地区分引力现象和加速度。

 **注 1.10.** 需要强调的是, 引力效应只能在局域上等效为加速度, 因此只要“船舱”足够大, 或者实验设备足够精确, 以至于可以检测出不同区域引力效应之差 (潮汐效应), 就可以将引力效应和加速度加以区分。

根据强等效原理, 所有在局域展开的物理实验, 只要转换到自由落体状态的参考系下看, 都是一致的。换言之, 可以将自由落体参考系下的结果, 通过合理的坐标变换, 得到有引力时的表达式。这其中最为关键的就是度规的转换。在自由落体状态时候, 局

<sup>1</sup> 扩展阅读: [81] 等著作中关于 Mach's principle 和 Isaac Newton's rotating bucket argument (also known as Newton's bucket) 的讨论。

<sup>2</sup> 需要注意, 弱等效原理的表述方式有很多, 但是几种表述互相之间均等效。

部 Lorentz 坐标系  $x^\mu$  遵从 Minkowski 度规  $\eta_{\mu\nu}$ 。假设一般的坐标系  $x'^\alpha$  与局部 Lorentz 坐标系之间符合转换关系  $x'^\alpha = x'^\alpha(x^\mu)$ , 则有  $dx'^\alpha = \frac{\partial x'^\alpha}{\partial x^\mu} dx^\mu$ 。可以得到

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = \left( \eta_{\mu\nu} \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} \right) dx'^\alpha dx'^\beta \quad (1.33)$$

与公式1.30比较, 可以得到,

$$g'_{\alpha\beta} = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} \quad (1.34)$$

### 材料 1.11 Raising and Lowering of index

张量指标可以通过度规  $g_{\mu\nu}$  进行升降。首先可以定义逆变度规张量  $g^{\mu\nu}$

$$g^{\alpha\mu} g_{\mu\beta} = \delta^\alpha_\beta \quad (1.35)$$

其中  $\delta^\alpha_\beta$  是 Kronecker delta 张量, 数值上与公式1.29一致。利用度规, 可以实现指标的升降:

$$A^\alpha = g^{\alpha\beta} A_\beta \quad (1.36)$$

类似地, 有

$$A_\alpha = g_{\alpha\beta} A^\beta. \quad (1.37)$$

## 1.2.4 协变导数

Minkowski 度规描述的几何是平直的, 与 Euclidean 几何差别不大<sup>1</sup>。在广义相对论的框架下, 要描述动力学, 就需要运用弯曲的时空对应的数学语言, 或者是非 Euclidean 几何 (non-Euclidean geometry)

### 材料 1.12 Parallel postulate

Euclidean 几何的五大公设中, 第五公设为:

**若两条直线都与第三条直线相交, 并且在同一边的内角之和小于两个直角, 则这两条直线在这一边必定相交。**

这一定义相较其他公设更为冗长, 并不是那么显然。晚近的研究表明, 将第五公设去除后, 也可以构成自洽的几何体系, 即所谓非 Euclidean 几何。根据曲率的不同, 可以将其分为椭圆几何 (又称 Riemannian 几何, 任意两条直线一定相交; 三角形内角和大于  $180^\circ$ ) 和双曲几何 (又称 Lobachevskian 几何, 至少可以引出两条平行线; 三角形内角和小于  $180^\circ$ )。

<sup>1</sup>有些人将其称为伪 Euclidean 几何

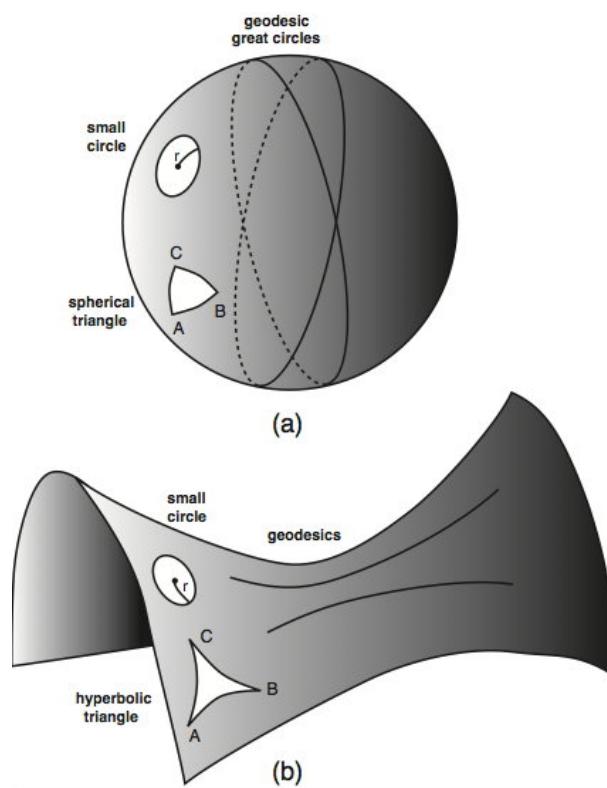


图 1-2 两个非 Euclidean 几何的例子

Figure 1-2 two examples of non-Euclidean geometry. Credit:

<https://blogs.futura-sciences.com/e-luminet/2017/09/22/non-euclidean-geometries/>

对于函数  $\vec{A}(x)$  的导数，我们通常定义为  $\lim_{dx \rightarrow 0} \frac{\vec{A}(x+dx)-\vec{A}(x)}{dx}$ 。然而在非 Euclidean 几何中，需要重新审视一些在 Euclidean 几何里习以为常的概念：比如，矢量的分量必须依赖局部的坐标系。如此一来， $\vec{A}(x+dx)-\vec{A}(x)$  的具体取值，就很值得玩味：如何对不在同一个位置处的两个矢量进行比较？

在 Euclidean 几何中这并不是一个问题，因为可以简单的平移矢量  $\vec{A}(x)$  至  $x + dx$  处，然后与  $\vec{A}(x + dx)$  比较。然而，在非 Euclidean 几何中，矢量的平移这个概念也需要被仔细检阅。具体来说，平移后的矢量，不仅取决于矢量的指向，也取决于平移的路径。

**例 1.13** 让我们想象一个生活在地球仪表面的二维生物（不妨假设这是一个没有厚度的蚂蚁，它只能沿着地球仪表面运动），假设这只蚂蚁从地球仪上的  $(0^\circ, 0^\circ)$  出发（也就是本初子午线和赤道的交点），决心一路向北。

在任意时刻，它的前进方向都是一个矢量，在起点处，这个方向指向正北。随着它的运动，这个矢量被不断的平移（注意，在这个二维球面空间里，矢量也只能在地球仪

表面上定义)。当它到达北极点时,这个平移后的矢量就称为指向东西经 $180^\circ$ 线。在我们的三维世界里去看的话,会发现,两个矢量的方向其实是垂直的(再继续走 $1/4$ 圈的话就变成了指向南极方向,与原方向相反了)。

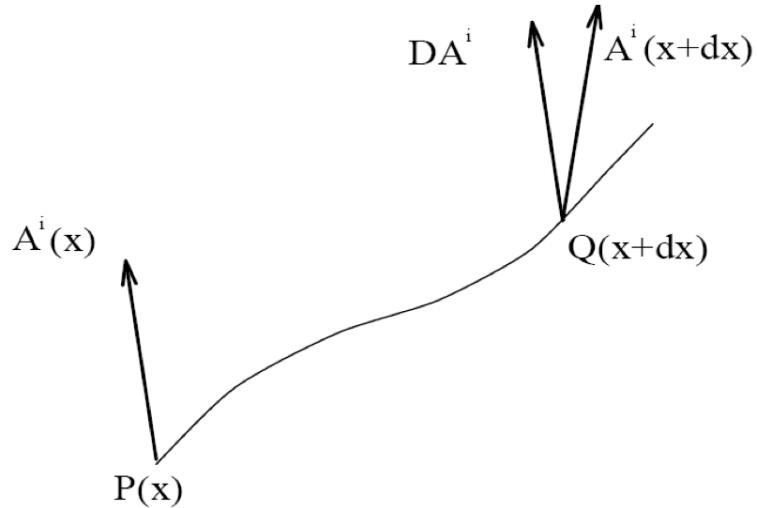


图 1-3 非 Euclidean 几何中矢量协变导数的图示

Figure 1-3 demonstration of covariant derivative of vectors in non-Euclidean geometry. Credit: Martin Hendry

如图1-3所示,在 $P$ 点处(坐标为 $x^\beta$ )有一矢量 $\vec{A}(x)$ ,其坐标分量为 $A^\alpha$ ,经过 $P \rightarrow Q$ 的路径到达 $Q$ 点(坐标为 $x^\alpha + dx^\alpha$ ),这时,如果 $\vec{A}$ 保持原来的坐标分量不变,则分量依然是 $A^\alpha$ 。与 Euclidean 几何不同,由于不同点处坐标基底也发生了改变,因此在 $x + dx$ 处的 $A^\alpha$ 并不一定与 $\vec{A}(x)$ 平行,真正平行的矢量为 $\vec{DA}(x + dx)$ ,其坐标分量为 $DA^\alpha$

$$DA^\alpha(x + dx) = A^\alpha(x) + \delta A^\alpha(x) \quad (1.38)$$

由坐标基底  $\vec{e}$  发生变化而引起的  $\delta A^\alpha$

$$\delta A^\alpha(x) = -\Gamma_{\mu\beta}^\alpha A^\mu dx^\beta \quad (1.39)$$

其中,  $\Gamma_{\mu\beta}^\alpha$  是 Christoffel 符号,通过它,可以得到不同位置处坐标基底之间的转换关系<sup>1</sup>

$$\frac{\partial \vec{e}_\mu}{\partial x^\beta} = \Gamma_{\mu\beta}^\alpha \vec{e}_\alpha \quad (1.40)$$

再次强调:我们在1.2.2节中提到过,在坐标基底发生变化时,张量的具体分量会变化,但是张量本身不会变。当我们定义了矢量的平移以后,我们终于可以合理地定义矢

<sup>1</sup>具体推导过程留作习题1.3。

量的（不随坐标基底变化的）导数了：可以在 Q 点  $x + dx$  处，将  $\vec{A}(x + dx)$  与从 P 点平移而来的  $\vec{D}A(x + dx)$  进行比较。这样定义出来的导数，我们称之为协变导数 (covariant derivative)  $A_{;\beta}^{\alpha}$ （见图 1-3）。

$$A_{;\beta}^{\alpha} \equiv \lim_{dx^{\beta} \rightarrow 0} \frac{\vec{A}^{\alpha}(x^{\beta} + dx^{\beta}) - \vec{D}A^{\alpha}(x^{\beta} + dx^{\beta})}{dx^{\beta}} = A_{,\beta}^{\alpha} + \Gamma_{\mu\beta}^{\alpha} A^{\mu} \quad (1.41)$$

其中， $A_{,\beta}^{\alpha}$  是普通的偏导  $\partial A^{\alpha} / \partial x^{\beta}$ ，有时也会写成  $\partial_{\beta} A^{\alpha}$ 。

通过度规张量  $g_{\mu\nu}$  的协变导数为 0，并通过交换指标、求和，可以得到

$$\Gamma_{\alpha\beta}^{\mu} = \frac{1}{2} g^{\mu\nu} \left( \frac{\partial g_{\nu\alpha}}{\partial x^{\beta}} + \frac{\partial g_{\nu\beta}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\nu}} \right), \quad (1.42)$$

这里可以看出，不同局部之间的联系，完全由局部的度规决定。

## 1.2.5 测地线 (geodesics)

考虑一个粒子的世界线<sup>1</sup>，在任意固有时间 (proper time)  $\tau$  的取值都对应一个坐标点  $x^{\alpha}$ ，同时有  $d\tau = \sqrt{-ds^2}/c$ ，从这一定义可以看出，固有时间的取值不依赖具体坐标系。可以定义，在  $x^{\alpha}$  处，粒子运动的四速度 (切矢) 为

$$u^{\alpha} = \frac{dx^{\alpha}}{d\tau}, \quad (1.43)$$

不难得到

$$g_{\alpha\beta} u^{\alpha} u^{\beta} = -c^2 \quad (1.44)$$

Newtonian 力学描述物体在惯性系中沿直线运动。然而，在弯曲时空中，直线的概念消失了；而和直线概念最为接近的，就是测地线。我们考虑一条曲线，它的每一个点上的切矢都与前一个点上的切矢平行，这样的操作可以唯一定义一条曲线，这样的曲线叫做测地线。根据定义，在任意点的局域惯性系中，有

$$\frac{du^{\alpha}}{d\tau} = \frac{du^{\alpha}}{dx^{\beta}} \frac{dx^{\beta}}{d\tau} = u_{,\beta}^{\alpha} u^{\beta} = 0 \quad (1.45)$$

由于在该点的局域惯性系中  $\Gamma_{\mu\nu}^{\alpha} = 0$ ，因此，上式又可以写成

$$u^{\beta} u_{,\beta}^{\alpha} = 0 \quad (1.46)$$

该方程是张量方程，在任意坐标系中均成立。

<sup>1</sup>世界线是三维空间中“轨迹”概念的延伸和对应。

代入速度定义式1.43和协变导数公式1.41，可以得到

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0 \quad (1.47)$$

这样一条测地线描述了在广义相对论框架下，Newton's first law of motion 的对应：不受外力作用的物体，沿测地线运动。需要特别说明的是，在推导过程中，我们默认了类时的测地线，这样，每一个点对应一个原时  $\tau$ 。然而，对于类光的测地线（即，测地线上每一点的原时  $\tau$  都相同），将  $\tau$  替换成仿射参量（affine parameter） $\lambda$  以后，公式1.46依然成立。特别的，这样的测地线称为零测地线 (null geodesics)。

## 1.3 广义相对论 (General Relativity) 初步

### 1.3.1 曲率张量

我们之前说过，在非 Euclidean 几何中，三角形内角和不等于  $180^\circ$ ，但显然三角形需要三个顶点，并不一定能在某个足够小的局域定义。能不能通过对某个点的时空性质观察，局域地确定时空是否弯曲？答案是肯定的，借助的工具就是 Riemann 张量。

$$R_{\beta\gamma\delta}^\alpha = \partial_\gamma \Gamma_{\beta\delta}^\alpha - \partial_\delta \Gamma_{\beta\gamma}^\alpha + \Gamma_{\mu\gamma}^\alpha \Gamma_{\beta\delta}^\mu - \Gamma_{\mu\delta}^\alpha \Gamma_{\beta\gamma}^\mu \quad (1.48)$$

它对应着一个矢量沿封闭路径平移后改变的量与封闭区域“面积”之间的比例关系。各类对称性可以消除独立分量的数目，在四维时空中，Riemann 张量一共有 20 个独立分量。

在 Riemann 张量基础上，还可以定义两个重要的张量：Ricci 张量

$$R_{\alpha\beta} \equiv R^\mu_{\alpha\mu\beta} \quad (1.49)$$

和 Einstein 张量

$$G^{\alpha\beta} \equiv R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R. \quad (1.50)$$

其中  $R = g^{\alpha\beta} R_{\alpha\beta}$  是 Ricci 标量。

### 1.3.2 Einstein 场方程

在 Newtonian 引力中，

$$\nabla^2 \left( -\frac{Gm}{r} \right) = 4\pi G\rho \quad (1.51)$$

等式的左边描述引力势（通常由  $\Phi$  表示），等式的右边描述的是质量分布。在广义相对论下，虽然对引力的描述改变了，但 Newton 对于引力的描述仍然应该近似成立。Einstein

认为引力是时空弯曲的表现，所以上述等式中的左边可以对应于某个与度规  $g^{\alpha\beta}$  有关的函数  $R^{\alpha\beta} + \mu g^{\alpha\beta}R + \Lambda g^{\alpha\beta}$ ，而质量分布在相对论中的对应是所谓能量-动量张量  $T_{\alpha\beta}$ <sup>1</sup>。可

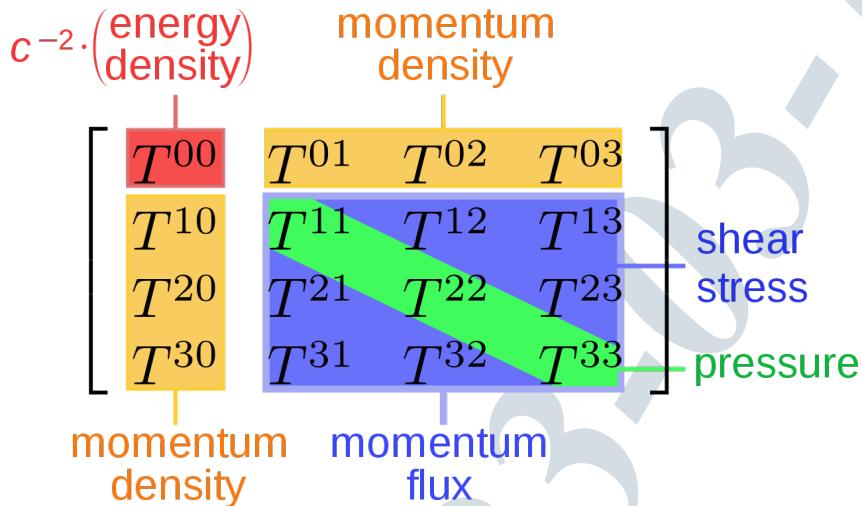


图 1-4 能量-动量张量的逆变分量 (注意上述表达式中默认了  $x^0 = t$  而非我们采用的  $x^0 = ct$ )

Figure 1-4 Contravariant components of the energy-momentum tensor (notice that instead of  $x^0 = ct$  that we adopted,  $x^0 = t$  is assumed in this form). Credit:

[https://en.wikipedia.org/wiki/File:StressEnergyTensor\\_contravariant.svg](https://en.wikipedia.org/wiki/File:StressEnergyTensor_contravariant.svg)

以将  $T^{\alpha\beta}$  理解为  $\alpha$  方向的动量流穿越等  $x^\beta$  线的分量。对于理想流体而言， $T^{00}$  是密度， $T^{ii}$  是压强，其他分量皆为零。

故此，可以猜测  $R^{\alpha\beta} + \mu g^{\alpha\beta}R + \Lambda g^{\alpha\beta} = kT^{\alpha\beta}$ 。注意到，由连续性要求可以得到

$$T^{\alpha\beta}_{;\beta} = 0. \quad (1.52)$$

又因为 Bianchi 恒等式，可以证明

$$G^{\alpha\beta}_{;\beta} = 0. \quad (1.53)$$

同时有  $g^{\alpha\beta}_{;\beta} = 0$  恒成立。因此 Einstein 猜测

$$G^{\alpha\beta} + \Lambda g^{\alpha\beta} = kT^{\alpha\beta} \quad (1.54)$$

取  $\Lambda = 0$ ，通过在弱场情形下与 Newtonian 引力的比较，可以确定常数  $k$  的取值，

$$R^{\alpha\beta} - \frac{1}{2}g^{\alpha\beta}R = \frac{8\pi G}{c^4}T^{\alpha\beta} \quad (1.55)$$

这就是广义相对论中描述时空与质量分布关系的公式——Einstein 场方程。

<sup>1</sup>有时候称为应力-能量张量，也称应力-能量-动量张量、能量-应力张量

 **注 1.14.** 在广义相对论中，通常默认公式 1.55 里的  $\Lambda = 0$ 。然而 Einstein 首次将广义相对论运用于宇宙时，为了使宇宙得到静态解，需要  $\Lambda$  不等于零，并将其称为宇宙学常数 (*cosmological constant*)。后来 Hubble 的观测证明宇宙并非静态，Einstein 由此将自己引入非零的  $\Lambda$  称为“我最大的错误 (*my biggest blunder*)”。然而近年来观测到宇宙有加速膨胀效应，恰好可以被这一宇宙学常数  $\Lambda$  所描述。不过其背后的物理机理（暗能量）尚属未解之谜。

### 材料 1.15 Natural units

在广义相对论相关的公式中，通常会出现常数  $G$  和  $c$  及其幂次。在一些文献中，会设定自然单位制 (**natural units**)，即将一些常数设置为无量纲数 1。具体到引力研究中，常会设置  $c = G = 1$ ，这又被称为几何单位制 (**Geometrized units**)。在本讲义中，为帮助物理量相应量级与量纲的计算和校验，我们不采用几何单位制。但部分公式推导中几何单位制更为方便，请同学们自行甄别。

### Exercise 1.1

用矩阵形式写出 Lorentz 变换公式，即显式展开公式 1.16。

### Exercise 1.2

试证，空间部分采用球坐标系时，坐标基底采取  $\{ct, r, \theta, \phi\}$ ，利用公式 1.34，Minkowski 时空的度规可以写成

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix} \quad (1.56)$$

或者等价地， $ds^2 = -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)$

### Exercise 1.3

对于平行移动的  $\vec{A} = A^\alpha \vec{e}_\alpha$ ，利用

$$\frac{\partial \vec{A}}{\partial x^\beta} = \frac{\partial A^\alpha}{\partial x^\beta} \vec{e}_\alpha + A^\alpha \frac{\partial \vec{e}_\alpha}{\partial x^\beta} = \vec{0}, \quad (1.57)$$

证明公式 1.40。

2023-03-13

# 2

## 引力波方程

### 2.1 广义相对论的 Newtonian 极限

由于在广义相对论下，“the matter tells spacetime how to curve, and the spacetime tells matter how to move”，因此 Einstein 场方程是高度耦合在一起的，我们说这样的系统是高度非线性的，因此它的求解是非常困难的一件事情。然而，我们可以通过在一些特殊情形下对其进行分析，进而得到一些有意义的结论。一个比较有用的特殊情形，就是弱场近似 (weak-field approximation)，这样可以将场方程线性化，因此这也被称为线性化引力近似 (linearized gravity approximation)。如果加上低速限制条件，我们就可以得到广义相对论的 Newtonian 极限。

#### 2.1.1 线性化引力 (linearized gravity)

让我们考虑如下情形：在原本平直的时空背景上，出现了一个小小的扰动，那么时空度规便偏离了原本的 Minkowski 度规  $\eta_{\alpha\beta}$

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} \quad (2.1)$$

当然，这样的扰动并不大，换句话说，<sup>1</sup>

$$|h_{\alpha\beta}| \ll 1, \quad (2.2)$$

<sup>1</sup>注意，只有  $g_{\alpha\beta}$  是张量，而  $h_{\alpha\beta}$  并不是。

我们可以在指标升降(见材料 1.11)时, 近似使用 Minkowski 度规  $\eta_{\alpha\beta}$  而不是真实的度规  $g_{\alpha\beta}$ 。<sup>1</sup>

由此可以计算在弱场近似下的 Riemann 张量

$$R_{\alpha\beta\gamma\delta} = \frac{1}{2} \left( -\frac{\partial^2 h_{\beta\delta}}{\partial x^\alpha \partial x^\gamma} + \frac{\partial^2 h_{\beta\gamma}}{\partial x^\alpha \partial x^\delta} + \frac{\partial^2 h_{\alpha\delta}}{\partial x^\beta \partial x^\gamma} - \frac{\partial^2 h_{\alpha\gamma}}{\partial x^\beta \partial x^\delta} \right) + O(h^2) \quad (2.3)$$

其中,  $h = h_\mu^\mu$  是  $h_\mu^\nu$  的迹 (trace)

对应地, 线性化的 Ricci 张量可以写为

$$\begin{aligned} R_{\alpha\beta} &= R_{\alpha\mu\beta}^\mu \\ &= \frac{1}{2} \left( -\frac{\partial^2 h}{\partial x^\alpha \partial x^\beta} + \frac{\partial^2 h_\beta^\mu}{\partial x^\alpha \partial x^\mu} + \frac{\partial^2 h_\alpha^\mu}{\partial x^\mu \partial x^\beta} - \eta^{\mu\nu} \frac{\partial^2 h_{\alpha\beta}}{\partial x^\mu \partial x^\nu} \right) + O(h^2) \end{aligned} \quad (2.4)$$

而 Ricci 标量则为

$$\begin{aligned} R &= g^{\alpha\beta} R_{\alpha\beta} = \eta^{\alpha\beta} R_{\alpha\beta} + O(h^2) \\ &= \frac{\partial^2 h^{\alpha\beta}}{\partial x^\alpha \partial x^\beta} - \eta^{\alpha\beta} \frac{\partial^2 h}{\partial x^\alpha \partial x^\beta} + O(h^2). \end{aligned} \quad (2.5)$$

将公式 2.4 和 2.5 代入公式 1.50, 可以得到线性化引力的 Einstein 张量  $G_{\alpha\beta}$ 。但是具体展开会发现, 其形式较为复杂。对此, 可以通过变量代换, 采用迹反转动度规微扰 (trace-reversed perturbation)  $\bar{h}_{\alpha\beta}$

$$\bar{h}_{\alpha\beta} \equiv h_{\alpha\beta} - \frac{1}{2} \eta_{\alpha\beta} h \quad (2.6)$$

 **注 2.1.** 注意,  $\bar{h}_{\alpha\beta}$  的迹为  $\bar{h} = \eta^{\alpha\beta} \bar{h}_{\alpha\beta} = \eta^{\alpha\beta} h_{\alpha\beta} - \frac{1}{2} \eta^{\alpha\beta} \eta_{\alpha\beta} h = -h$ , 其迹与  $h_{\alpha\beta}$  相反, 故此得名 *trace-reversed perturbation*。

在线性化近似下, 可以得到形式相对简化的 Einstein 张量

$$G_{\alpha\beta} = \frac{1}{2} \left( \frac{\partial^2 \bar{h}_\beta^\mu}{\partial x^\alpha \partial x^\mu} + \frac{\partial^2 \bar{h}_\alpha^\mu}{\partial x^\mu \partial x^\beta} - \eta^{\mu\nu} \frac{\partial^2 \bar{h}_{\alpha\beta}}{\partial x^\mu \partial x^\nu} - \eta_{\alpha\beta} \frac{\partial^2 \bar{h}^{\mu\nu}}{\partial x^\mu \partial x^\nu} \right) + O(h^2) \quad (2.7)$$

由此, 可以将线性化的 Einstein 场方程化为

$$-\eta^{\mu\nu} \frac{\partial^2 \bar{h}_{\alpha\beta}}{\partial x^\mu \partial x^\nu} - \eta_{\alpha\beta} \frac{\partial^2 \bar{h}^{\mu\nu}}{\partial x^\mu \partial x^\nu} + \frac{\partial^2 \bar{h}_\beta^\mu}{\partial x^\alpha \partial x^\mu} + \frac{\partial^2 \bar{h}_\alpha^\mu}{\partial x^\mu \partial x^\beta} + O(h^2) = \frac{16\pi G}{c^4} T_{\alpha\beta} \quad (2.8)$$

值得注意的是, 等式的第一项可以写为  $-\square \bar{h}_{\alpha\beta}$ , 其中  $\square$  符号为 d'Alembertian 算符, 也就是在平直时空中的波算符。但是, 这个等式依然过于冗长, 有没有方法可以将上式进一步简化?

<sup>1</sup>注, 计算度规的逆变分量时除外, 见 [37] 2.127 式

### 材料 2.2 Gauge transformations

可以注意到，在更一般的情形下，公式 1.34 可以写成

$$g'_{\alpha\beta} = g_{\mu\nu} \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} \quad (2.9)$$

并且有

$$ds^2 = g'_{\alpha\beta} dx'^\alpha dx'^\beta = g_{\mu\nu} dx^\mu dx^\nu. \quad (2.10)$$

我们把  $x'^\alpha(x^\mu)$  称为规范变换。这种在规范变化下公式形式保持不变的性质称为规范自由度 (gauge freedom)。

考虑基于“矢量”  $\xi^\alpha$  的如下形式的坐标变换：

$$x'^\alpha = x^\alpha + \xi^\alpha(x^\mu) \quad (2.11)$$

使得变换后依然可以依照公式 2.1 进行分解，并且有公式 2.2 仍然成立。当  $\frac{\partial \xi^\alpha}{\partial x^\beta}$  为小量时，通过定义  $\xi_\alpha \equiv \eta_{\alpha\beta} \xi^\beta$ ，可以得到

$$h_{\alpha\beta} \rightarrow h'_{\alpha\beta} = h_{\alpha\beta} - \frac{\partial \xi_\alpha}{\partial x^\beta} - \frac{\partial \xi_\beta}{\partial x^\alpha} \quad (2.12)$$

在特定的规范条件下，通过合理选取矢量  $\xi^\alpha$ ，可以在不改变公式物理本质的前提下，简化数学表达式。

相应地，有

$$\bar{h}_{\alpha\beta} \rightarrow \bar{h}'_{\alpha\beta} = \bar{h}_{\alpha\beta} - \frac{\partial \xi_\alpha}{\partial x^\beta} - \frac{\partial \xi_\beta}{\partial x^\alpha} + \eta_{\alpha\beta} \eta^{\mu\nu} \frac{\partial \xi_\mu}{\partial x^\nu} \quad (2.13)$$

通常，可以选取所谓 Lorenz 规范<sup>1</sup>，即，要求迹反转度规微扰的梯度为零

$$\bar{h}^{\mu\nu}_{,\nu} = 0 \quad (2.14)$$

符合上述条件的坐标系又称调和坐标条件 (harmonic coordinate conditions)

由于 Lorenz 规范意味着

$$0 = \frac{\partial}{\partial x'^\mu} \bar{h}'^\mu_\beta = \frac{\partial}{\partial x^\mu} \bar{h}^\mu_\beta - \square \xi_\beta + O(h^2), \quad (2.15)$$

可以令  $\xi_\beta$  满足

$$\square \xi_\beta = \frac{\partial \bar{h}^\mu_\beta}{\partial x^\mu} \quad (2.16)$$

从而得到 Lorenz 规范。

<sup>1</sup>注意，很多书本中错误地写作 Lorentz 规范，其实 Ludvig Lorenz (1829-1891) 和 Hendrik Lorentz (1853-1928, Lorentz 变换以其命名) 是两位科学家

 **注 2.3.** 需要注意，公式2.16的一条并不是唯一的，因为在原有的  $\xi_\alpha$  上只要加上一个满足  $\square \xi'_\alpha = 0$  的规范变换，所有结论依然保持不变。

在这样的形式下，公式2.8中，等式左边的后三项均变为零。我们也就得到了线性化引力在 Lorenz 规范下的 Einstein 场方程

$$-\square \bar{h}_{\alpha\beta} = \frac{16\pi G}{c^4} T_{\alpha\beta} \quad (2.17)$$

## 2.1.2 Newtonian 极限

在第2.1.1节中我们讨论了弱场情况下的 Einstein 场方程。在本节中，我们在弱场的前提下更进一步，引入低速 ( $v \ll c$ ) 要求，讨论对应的场方程。我们知道，在低速情形下，Newtonian 力学可以看做是狭义相对论的极限情形。类似地，在弱场 ( $h \ll 1$ )、低速 ( $v \ll c$ ) 条件下，Newtonian 引力也应该可以看做是广义相对论的极限情形。

在 Newtonian 极限下，我们可以得到如下条件

$$\begin{aligned} T_{00} &= \rho c^2 && (\text{mass energy density}) \\ |T_{0i}| &\sim \rho c^2 (v/c) \ll T_{00} && (\text{slow motion } v \ll c) \\ |T_{ij}| &\sim p && (\rho c^2 (v/c)^2 \ll T_{00}) \quad (\text{small internal stresses}) \end{aligned} \quad (2.18)$$

在低速情形下， $\frac{\partial}{c\partial t} \approx \frac{v}{c} \frac{\partial}{\partial x}$  是小量，由此 d'Alembertian 算符可以近似为空间 Laplace 算符  $\square \rightarrow \nabla^2$ 。因此，在 Newtonian 极限下，场方程退化为

$$\begin{aligned} \nabla^2 \bar{h}_{00} &= -16\pi G \rho / c^2 \\ \nabla^2 \bar{h}_{0i} &= 0 \\ \nabla^2 \bar{h}_{ij} &= 0 \end{aligned} \quad (2.19)$$

这个方程组可以得到平凡解  $\bar{h}_{0i} = 0$  和  $\bar{h}_{ij} = 0$ 。对应公式1.51，可以发现非平凡解  $\bar{h}_{00} = -4\Phi/c^2$ 。

更进一步，可以得到<sup>1</sup>

$$g_{\mu\nu} = \begin{pmatrix} -1 - 2\Phi/c^2 & 0 & 0 & 0 \\ 0 & 1 - 2\Phi/c^2 & 0 & 0 \\ 0 & 0 & 1 - 2\Phi/c^2 & 0 \\ 0 & 0 & 0 & 1 - 2\Phi/c^2 \end{pmatrix} + O(\Phi^2/c^4) \quad (2.20)$$

<sup>1</sup> 具体推导参照习题2.1

## 2.2 引力波

### 2.2.1 引力波的传播

从公式2.17中可以看到，在Lorenz规范下，由平直时空背景上的线性微扰引起的运动方程是波动方程，其中能量-动量张量是源项。在真空中，度规微扰有波动解，这就是**引力波 (gravitational wave)**。我们考虑平面波

$$\bar{h}^{\alpha\beta} = A^{\alpha\beta} \exp[i(\omega t + \mathbf{k} \cdot \mathbf{x})] = A^{\alpha\beta} \exp(ik_\mu x^\mu) \quad (2.21)$$

注意，其中  $k_\mu$  是实的协变矢量，而  $A^{\alpha\beta}$  是二阶逆变张量，其分量可以为复。所谓平面波，意味着在满足

$$k_\mu x^\mu = k_0 ct + \mathbf{k} \cdot \mathbf{x} = \text{const.} \quad (2.22)$$

的超曲面上， $\bar{h}^{\alpha\beta}$  是一个常数，换言之，符合公式2.22的波阵面是一个平面。这其中， $\mathbf{k} = k^i$ 。方便起见，可以定义  $k_0 = -\omega/c$ ，这样， $k_\alpha$  的四维分量可以写成  $(-\omega/c, \mathbf{k})$ 。

不难证明， $k_\alpha$  是一个零 (**null**) 协变矢量，换句话说，四维矢量  $k^\alpha$  是零矢量 (类光的)，它与光子的世界线相切，故此，引力波的传播速度即光速。也就是

$$\omega^2 = c^2 |\mathbf{k}|^2 \quad (2.23)$$

更进一步的，可以证明在广义相对论中，引力波传播过程中不包含色散。

#### 材料 2.4 Plane Wave

在物理学领域，经常采用类似  $A^{\alpha\beta} \exp(ik_\mu x^\mu)$  形式的平面波解，这其中， $A^{\alpha\beta}$  可以是复的，而我们通常仅选取解的实部。

当然，现实情况下我们感兴趣的波很可能更为复杂，而不是单频的平面波。然而，根据 Fourier 定理，可以证明，任何满足波动方程 (公式2.17) 和 Lorenz 规范 (公式2.14) 的解，都可以看作是平面波解的叠加。因此，本章中引力波部分的讨论均基于平面波。

在推导公式2.17的过程中，用到了规范条件  $\bar{h}^{\alpha\beta}_{;\beta} = 0$  (公式2.14)。同时注意，通过对波动解公式2.21求偏导，有  $\bar{h}^{\alpha\beta}_{,\mu} = ik_\mu \bar{h}^{\alpha\beta}$ 。将上述两者结合，可以发现

$$k_\beta A^{\alpha\beta} = 0 \quad (2.24)$$

这给了  $A^{\alpha\beta}$  限制条件，它必须与  $k_\beta$  正交。在后面的学习中我们会发现，上述条件意味着引力波的传播方向与其作用方向垂直，因此，引力波是横波。

## 2.2.2 横向无迹规范 (transverse-traceless gauge)

可以注意到, Lorenz 规范并不是唯一的。实际上, 通过引入额外的满足  $\square \xi_\alpha = 0$  条件的  $\xi_\alpha$ , 公式2.21将依然成立(注2.3)。我们总可以选取  $\xi_\alpha = B_\alpha \exp(i k_\mu x^\mu)$  (这样的选择自然满足  $\square \xi_\alpha = 0$ )。可以证明, 依照公式2.13, 对于一个任意的 Lorenz 规范下的  $\bar{h}_{\alpha\beta}$ , 总可以选取横向无迹规范 (transverse-traceless gauge, TT gauge), 使得  $h^{\text{TT}} = \eta^{\mu\nu} h_{\mu\nu}^{\text{TT}} = 0$  (无迹), 以及  $h_{0\alpha}^{\text{TT}}(x) = 0$  (仅空间分量非零), 其转换关系满足

$$h_{\alpha\beta}^{\text{TT}} = \bar{h}_{\alpha\beta} - \frac{\partial \xi_\alpha}{\partial x^\beta} - \frac{\partial \xi_\beta}{\partial x^\alpha} + \eta_{\alpha\beta} \eta^{\mu\nu} \frac{\partial \xi_\mu}{\partial x^\nu} \quad (2.25)$$

注意等号左边是  $h_{\alpha\beta}^{\text{TT}}$  而非  $\bar{h}_{\alpha\beta}^{\text{TT}}$ 。这是因为根据公式2.6, 由于迹  $h = 0$ , 所以  $\bar{h}_{\alpha\beta} = h_{\alpha\beta}$ 。

总结起来, 在横向无迹规范下, 我们有如下条件

$$h_{0\alpha}^{\text{TT}} = 0, \quad \eta^{ij} h_{ij}^{\text{TT}} = 0, \quad \partial^j h_{ij}^{\text{TT}} = 0, \quad (2.26)$$

共计 8 个限制条件。注意到, 由于  $h_{\mu\nu}$  对称, 整个  $4 \times 4$  的度规张量一共只有 10 个自由度。在采用 Lorenz 规范后, 又消减了 4 个自由度, 剩下 6 个自由度。通过选定  $\xi_\alpha = B_\alpha \exp(i k_\mu x^\mu)$ , 进一步消减, 只剩下 2 个自由度。这两个自由度, 就对应了引力波的两个极化 (polarisation, 或称偏振)。

从简化表达式的角度考虑, 不妨将坐标基矢  $e_3$  方向选为平面波的波矢  $\mathbf{k}$  方向。换言之, 四维波矢  $k_\beta$  可以写作形如  $(-\omega/c, 0, 0, \omega/c)$  的形式 (公式2.23)。由于公式2.24, 可以得到  $\omega/c (A^{\alpha 0} - A^{\alpha 3}) = 0$ 。另一方面, 由于横向无迹规范意味着  $h_{0\alpha}^{\text{TT}} = h_{\alpha 0}^{\text{TT}} = 0$ , 或者说  $A_{\text{TT}}^{\alpha 0} = 0$ , 不难得出  $A_{\text{TT}}^{\alpha 3} = 0$ 。这样, 引力波振幅张量  $A_{\alpha\beta}$  中所有非零的量只剩下  $A_{11}$ ,  $A_{12}$ ,  $A_{21}$ ,  $A_{22}$ , 由于无迹要求导致了  $A_{11} + A_{22} = 0$ , 又由于对称性要求  $A_{12} = A_{21}$ , 因此通常可以将横向无迹规范下的引力波张量表达为

$$h_{\mu\nu}^{\text{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{11} & A_{12} & 0 \\ 0 & A_{12} & -A_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \exp[-i\omega(t - z/c)] \quad (2.27)$$

我们这里定义  $h_+ = A_{11} \exp[-i\omega(t - z/c)]$ , 以及  $h_\times = A_{12} \exp[-i\omega(t - z/c)]$ , 将其称为在横向无迹规范下引力波的加模 (+ 模, plus mode) 和乘模 ( $\times$  模, cross mode)<sup>1</sup>。这

<sup>1</sup>我们在图2-1会解释这样命名的原因。

样，可以将引力波张量表示简化为

$$h_{\mu\nu}^{\text{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_x & 0 \\ 0 & h_x & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (2.28)$$

注意到，引力波的效果仅体现在  $x^1 - x^2$  平面上，与传播方向  $x^3$  垂直，因此引力波是横波。

## 2.3 引力波的物理性质

在本节中，我们探讨引力波的各种物理效应，从而更深入地理解引力波。我们将描述当引力波作用到物体时，其坐标如何变化；同时作用到两个物体时，其距离如何变化。这些描述可以帮我们从原理上构造出一种引力波的探测手段，并且可以给出具体的引力波效应表达式。最后，我们将描述引力波携带的能量与能流，这对之后的章节中描述引力波源的演化规律是十分重要的。

### 2.3.1 引力波的效果

由于广义相对论说任何局域坐标系在自由落体状态下都可以由 Minkowski 度规描述，因此，作用在一个点上时，引力波的效果无法被探测到。考虑一个检验粒子，遵从测地线方程（公式1.47），在引力波到来之前，其相对参考系静止 ( $dx^i/d\tau = 0$ )。故此，其加速度为

$$\begin{aligned} \frac{d^2x^i}{d\tau^2} &= -\Gamma_{\alpha\beta}^i \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} \\ &= -\Gamma_{00}^i \left( \frac{dx^0}{d\tau} \right)^2 \\ &= -\frac{1}{2}\eta^{i\alpha}(h_{\alpha 0,0} + h_{0\alpha,0} - h_{00,\alpha}) \left( \frac{dx^0}{d\tau} \right)^2 \\ &= 0 \end{aligned} \quad (2.29)$$

这其中，第二行用到了条件  $dx^i/d\tau = 0$ ，第四行选取横向无迹规范，则根据公式2.28，有  $h_{0\alpha}^{\text{TT}} = 0$ 。公式2.29表明，在引力波作用之前处于静止状态的粒子，当引力波经过时，粒子的加速度为零，粒子依然处于静止状态。因此，在横向无迹规范下，对于单个粒子而言，引力波没有可观测效应。

接下来，我们考虑两个在  $x^1$  方向上间隔小量  $\varepsilon$  的粒子。选取坐标系，使得沿  $x^3$  方向传播的平面引力波到达两个粒子的事件对应于  $(0, 0, 0, 0)$  和  $(0, x^1 = \varepsilon, 0, 0)$ 。我们在前

面已经展示，引力波并不会改变两个粒子的坐标。但引力波会改变两者间的**固有距离 (proper distance)**

$$\begin{aligned}\Delta l &= \int |ds^2|^{1/2} = \int |g_{\alpha\beta}dx^\alpha dx^\beta|^{1/2} \\ &= \int_0^\varepsilon |g_{11}dx^1 dx^1|^{1/2} \\ &\approx \varepsilon |g_{11}(x^1 = 0)|^{1/2} \\ &\approx \varepsilon [1 + \frac{1}{2}h_+^{TT}(x^1 = 0)]\end{aligned}\tag{2.30}$$

在横向无度规下计算引力波效应，静止的粒子坐标刻度不变，但是粒子之间的固有距离会发生变化。为了更直观地理解引力波效应，可以在**固有探测器系 (proper detector frame)**下讨论引力波对**检验质量 (test mass)**的影响。在固有探测器系中，检验质量至少在特定方向上作自由落体运动，近似可以采用 Minkowski 度规。引力波传播过来后，检验质量的位置会发生变化。如果物理学家手里有一把不受引力波效果影响的刚性“尺子”进行测量，就可以测量出引力波的效果来。换言之，概念上，可以将固有距离的变化视作坐标的变化。这样的物理图景更符合直觉，但与横向无度规下得到的结论并无本质区别。<sup>1</sup>

下面在更一般的语境下讨论引力波的效应。假设引力波波前到达两个粒子的事件对应间隔为  $\xi^\alpha$ ，且两个粒子的四速度均为  $\frac{dx^\alpha}{d\tau}$ 。如果两者的空间间隔远小于引力波的特征波长，就可以利用**测地偏离 (geodesic deviation)** 公式来计算  $\xi^\alpha$  随时间的演化

$$\frac{D^2\xi^\alpha}{D\tau^2} = -R^\alpha_{\beta\gamma\delta} \frac{dx^\beta}{d\tau} \xi^\gamma \frac{dx^\delta}{d\tau}\tag{2.31}$$

这其中， $D\xi^\alpha/D\tau = \frac{dx^\mu}{d\tau} \xi^\alpha_{;\mu}$  是  $\xi^\alpha$  沿  $x^\alpha(\tau)$  方向的协变微分，而  $R^\alpha_{\beta\gamma\delta}$  是 Riemann 张量。在低速情形下，由于  $\frac{dx^j}{d\tau} \ll \frac{dx^0}{d\tau}$ ，方程可以简化为：

$$\frac{d^2\xi^j}{d\tau^2} = -R^j_{0k0} \xi^k \left( \frac{dx^0}{d\tau} \right)^2\tag{2.32}$$

代入 Riemann 张量 (公式1.48)，并且考虑到在局域惯性系下：1.  $\Gamma^\alpha_{\beta\gamma} = O(h)$ ，直接贡献项为  $\Gamma^\alpha_{\beta\gamma}$  二次方的形式，因此是二阶小量而可忽略；2. 两个事件间小间隔 (即对度规的空间偏导为零)，有

$$\frac{d^2\xi^j}{dt^2} = \frac{1}{2} \eta^{jk} \frac{\partial^2 h_{kl}^{TT}}{\partial t^2} \xi^l\tag{2.33}$$

这表明，在固有探测器坐标系下，质量为  $m$  的检验质量会受到一个力：

$$F_j = ma_j = \frac{m}{2} \frac{\partial h_{jk}^{TT}}{\partial t^2} \xi^k\tag{2.34}$$

<sup>1</sup>这里体现了相对论的特点：选择不同的参考系，看到的物理图像很不相同，但其实通过坐标转换，其实质又并未变化。

因此，实验学家不需要成为广义相对论的专家，也可以利用 Newtonian 力学的语言，描述引力波作用下的探测器的效果。

考虑引力波  $h_{jk}^{\text{TT}}$  很小的前提下，可以直接积分，得到

$$\xi^j = \xi^j|_{t=0} + \frac{1}{2} \eta^{jk} h_{kl}^{\text{TT}} \xi^l|_{t=0} \quad (2.35)$$

可以看到，引力波对于两个事件的间隔的影响，同时正比于引力波  $h_{kl}^{\text{TT}}$  和初始间距  $\xi^l|_{t=0}$ 。

不妨设一个事件处于原点  $(0, 0, 0, 0)$ ，另一个事件则为  $x^\alpha = (0, \xi \sin \theta \cos \phi, \xi \sin \theta \sin \phi, \xi \cos \theta)$ ，代入公式2.33，可以得到：

$$\begin{aligned} a_1 &= \frac{d^2 x^1}{dt^2} = \frac{1}{2} \frac{\partial^2}{\partial t^2} h_+ \xi \sin \theta \cos \phi + \frac{1}{2} \frac{\partial^2}{\partial t^2} h_\times \xi \sin \theta \sin \phi \\ a_2 &= \frac{d^2 x^2}{dt^2} = -\frac{1}{2} \frac{\partial^2}{\partial t^2} h_+ \xi \sin \theta \sin \phi + \frac{1}{2} \frac{\partial^2}{\partial t^2} h_\times \xi \sin \theta \cos \phi \\ a_3 &= \frac{d^2 x^3}{dt^2} = 0 \end{aligned} \quad (2.36)$$

注意到，两个事件的距离在  $z$  轴上的分量保持不变，这是因为引力波是横波，而我们选用的横向无迹规范中，引力波的传播方向为  $z$  轴方向。因此，在引力波传播方向上 ( $z$  轴)，引力波不会产生作用。

**例 2.5** 让我们讨论设两个事件的连线位于  $x^1 - x^2$  平面上的情形，也就是说  $\theta = \pi/2$ ，我们可以简化得到  $\xi^j = (\xi \cos \phi, \xi \sin \phi, 0)$ 。这种情况下，不妨考虑两个特例：

如果  $h_\times = 0$  且  $\frac{\partial^2}{\partial t^2} h_+ > 0$ ，可以得到

$$\begin{aligned} \frac{d^2 x^1}{dt^2} &\propto x^1 \\ \frac{d^2 x^2}{dt^2} &\propto -x^2 \end{aligned} \quad (2.37)$$

反之，如果  $h_+ = 0$  且  $\frac{\partial^2}{\partial t^2} h_\times > 0$ ，可以得到

$$\begin{aligned} \frac{d^2 x^1}{dt^2} &\propto x^2 \\ \frac{d^2 x^2}{dt^2} &\propto x^1 \end{aligned} \quad (2.38)$$

如图2-1所示，在垂直于引力波传播方向的平面上，一圈自由放置的粒子将会由于引力波影响而变形。当  $h_+ \neq 0$  而  $h_\times = 0$  时，这圈粒子的轨迹如同加号 + 般，故称  $h_+$  为 **加模 (plus mode)**；当  $h_\times \neq 0$  而  $h_+ = 0$  时，这圈粒子的轨迹如同乘号  $\times$ ，故称  $h_\times$  为 **乘模 (cross mode)**。

下面考虑两点间距离  $\xi$  随时间  $t$  的变化  $\xi(t)$ 。对于  $\xi^k$  而言，其在连线方向上所受的

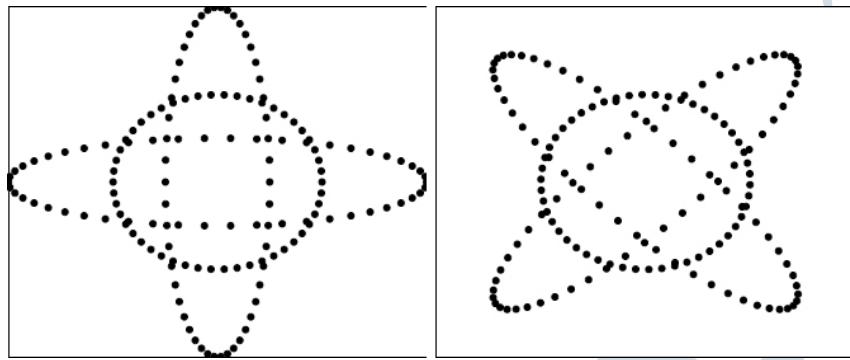


图 2-1 广义相对论下引力波的两种偏振

Figure 2-1 Two polarisation modes under general relativity. Credit: [84]

加速度  $a$  是加速度  $a^j$  在单位矢量  $\xi^k/\xi$  上的投影，即

$$\begin{aligned} a &= \eta_{jk} a^j \xi^k / \xi = \frac{c^2}{\xi} R_{0j0k} \xi^j \xi^k \\ &= \frac{d^2 x_1}{dt^2} \sin \theta \cos \phi + \frac{d^2 x_2}{dt^2} \sin \theta \sin \phi \\ &= \xi \left( \frac{1}{2} \frac{\partial^2}{\partial t^2} h_+ \sin^2 \theta \cos 2\phi + \frac{1}{2} \frac{\partial^2}{\partial t^2} h_\times \sin^2 \theta \sin 2\phi \right) \end{aligned} \quad (2.39)$$

若初始时两粒子静止，两重积分后，就可以得到

$$\xi(t) = \xi(0) \left( 1 + \frac{1}{2} h_+ \sin^2 \theta \cos 2\phi + \frac{1}{2} h_\times \sin^2 \theta \sin 2\phi \right) \quad (2.40)$$

**注 2.6.** 注意，公式 2.40 中， $\xi$  与  $h_+$  和  $h_\times$  之间依赖关系是取决于两倍的方位角  $2\phi$ 。

### 2.3.2 探测原理初探

在固有探测器系下，可以将引力波的效果视为对检验质量的加速度。在这里，我们通过一个简单的原理示意，初步探究利用电磁波测量引力波的基本原理。<sup>1</sup> 如图 2-2 所示，位于  $x^1 - x^3$  平面内的两点  $\mathcal{A}$  与  $\mathcal{B}$ 。考虑  $\mathcal{B}$  处于原点，并接收由  $\mathcal{A}$  发射的单频电磁波信号，假设出射频率为  $\nu_{\mathcal{A}}$ ，而接收频率为  $\nu_{\mathcal{B}}$ ，两者会因为引力波作用产生的相对运动而不同，会导致 Doppler 红移  $z = \frac{\nu_{\mathcal{A}} - \nu_{\mathcal{B}}}{\nu_{\mathcal{B}}}$ 。由  $\mathcal{B}$  到  $\mathcal{A}$  的方向矢量为  $p^i$ ，而引力波沿  $z$  轴方向  $n^i = (0, 0, 1)$  传播，两者夹角为  $\theta$ 。为简化计算，我们考虑引力波只含加模成分，因此有在横向无迹规范下， $h_{11} = -h_{22} = h_+(x^0 - x^3)$ <sup>2</sup> 由  $\mathcal{A}$  发出的电磁信号沿零测地线

$$k^\alpha = \frac{dx^\alpha}{d\lambda} = \nu \left[ 1, -\left( 1 - \frac{1}{2} h_+ \right) \sin \theta, 0, -\cos \theta \right] \quad (2.41)$$

<sup>1</sup> 在第 4.2.1 节中我们会讨论更为实际的情形。

<sup>2</sup> 注意公式 2.27 中写作一般读者更熟悉的  $t + z/c$  的形式。但在本节中  $z$  指代红移，请读者注意甄别。另外，因为定义了引力波波矢方向在空间中指向  $x^3$  轴正方向，因此相位部分为  $x^0 - x^3$  而非  $x^0 + x^3$ 。

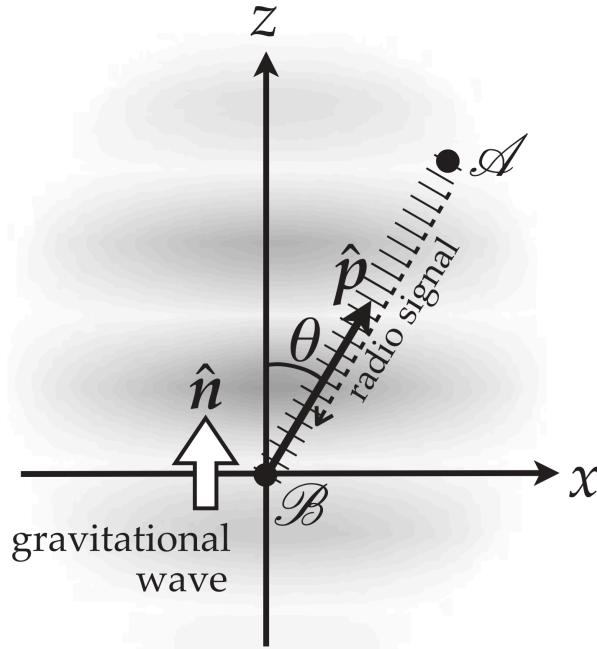


图 2-2 引力波探测原理示意

Figure 2-2 Illustration of detection of gravitational wave. Credit: [37]

其中  $\lambda$  为仿射参量，并且有  $k^0 = \frac{dx^0}{d\lambda} = \nu$ 。

由于该电磁波沿零测地线传播，由测地方程得

$$\frac{dk^\alpha}{d\lambda} = -\Gamma_{\beta\gamma}^\alpha k^\beta k^\gamma \quad (2.42)$$

为了计算  $\mathcal{A}$ 、 $\mathcal{B}$  两点之间的 Doppler 频移，需要沿测地线对  $\frac{d\nu}{d\lambda}$  进行积分。考虑加模引力波的影响，有

$$\begin{aligned} \frac{d\nu}{d\lambda} &= -\Gamma_{11}^0 k^1 k^1 = -\frac{1}{2} \frac{\partial h_+}{\partial x^0} \left[ \nu \left( 1 - \frac{1}{2} h_+ \right) \sin \theta \right]^2 \\ &\approx -\frac{1}{2} \frac{\partial h_+}{\partial x^0} \nu^2 \sin^2 \theta \end{aligned} \quad (2.43)$$

其中  $\Gamma_{11}^0 = \frac{1}{2} \frac{\partial h_+}{\partial x^0}$ 。又因为  $h_+(x^0 - x^3)$  对时间的偏导可以用其对仿射参量  $\lambda$  的全微分表示：

$$\begin{aligned} \frac{dh_+}{d\lambda} &= \frac{dx^0}{d\lambda} \frac{\partial h_+}{\partial x^0} + \frac{dx^3}{d\lambda} \frac{\partial h_+}{\partial x^3} \\ &= \frac{dx^0}{d\lambda} \frac{\partial h_+}{\partial x^0} - \frac{dx^3}{d\lambda} \frac{\partial h_+}{\partial x^0} \\ &= (k^0 - k^3) \frac{\partial h_+}{\partial x^0} \\ &= \nu (1 + \cos \theta) \frac{\partial h_+}{\partial x^0} \end{aligned} \quad (2.44)$$

由此可以推算出

$$\frac{1}{\nu} \frac{d\nu}{d\lambda} = \frac{d \ln \nu}{d\lambda} \approx -\frac{d}{d\lambda} \left[ \frac{1}{2} (1 - \cos \theta) h_+ \right] \quad (2.45)$$

将其沿  $\mathcal{A} - \mathcal{B}$  积分，则有

$$\begin{aligned}\ln(\nu_{\mathcal{A}}/\nu_{\mathcal{B}}) &= - \int_{\mathcal{A}}^{\mathcal{B}} d\lambda \frac{d \ln \nu}{d\lambda} \\ &\approx \int_{\mathcal{A}}^{\mathcal{B}} d\lambda \left[ \frac{1}{2}(1 - \cos \theta) h_+ \right] \\ &\approx \frac{1}{2}(1 - \cos \theta) [h_+(\mathcal{B}) - h_+(\mathcal{A})]\end{aligned}\quad (2.46)$$

又因为  $\ln \nu_{\mathcal{A}}/\nu_{\mathcal{B}} = \ln(1 + z) \approx z$ ，因此有

$$z = \frac{\nu_{\mathcal{A}} - \nu_{\mathcal{B}}}{\nu_{\mathcal{B}}} \approx \frac{1}{2}(1 - \cos \theta) [h_+(\mathcal{B}) - h_+(\mathcal{A})] \quad (2.47)$$

我们看到，由于引力波的存在，在  $\mathcal{B}$  点处接收到的电磁波会发生 Doppler 频移，对于平面波形式的引力波而言，这一效应仅依赖于收发两端所受的引力波取值。在一般情况下，引力波引起的电磁波 Doppler 频移可以表示为

$$z = \frac{\nu_{\text{em}} - \nu_{\text{rec}}}{\nu_{\text{rec}}} \approx \frac{1}{2} \frac{p^j p^k}{1 + n_i p^i} [h_{jk}^{\text{TT}}(t_{\text{rec}}, x_{\text{rec}}^l) - h_{jk}^{\text{TT}}(t_{\text{em}}, x_{\text{em}}^l)] \quad (2.48)$$

### 2.3.3 引力波携带的能量

Feynman 已经用粘珠实验定性地证明了引力波携带能量。如何将其定量描述？这一点，对研究引力波源的演化至关重要。然而，一个略为尴尬的事实是：在广义相对论中，无法将引力波能量局域地定义成一个标量。这是因为，我们总可以局域地转换出一个惯性参考系，从而将引力效应消除，也就无所谓引力波能量的概念了。好在，我们还可以在一个足够大的尺度上讨论引力波的能量（空间尺度至少包含数倍波长），等效地计算出这个时空范围内的引力波平均能量。

依旧考虑度规微扰  $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$ ，对应地可以计算出爱因斯坦张量  $G_{\alpha\beta}$ 。在真空中，可以将爱因斯坦场方程关于  $h$  展开：

$$0 = G_{\alpha\beta} = \overset{0}{G}_{\alpha\beta} + \overset{1}{G}_{\alpha\beta}[h_{\mu\nu}] + \overset{2}{G}_{\alpha\beta}[h_{\mu\nu}] + \dots \quad (2.49)$$

这里  $\overset{0}{G}_{\alpha\beta}$  其实就是平直背景  $\eta_{\alpha\beta}$  对应的爱因斯坦张量，自然可以满足  $\overset{0}{G}_{\alpha\beta} = 0$ 。而  $\overset{1}{G}_{\alpha\beta}[h_{\mu\nu}], \overset{2}{G}_{\alpha\beta}[h_{\mu\nu}]$  等分别是  $h$  的一阶项、二阶项、等等。

同时，也可以将度规对形式参量  $\lambda$  展开：

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} = \eta_{\alpha\beta} + \lambda \overset{1}{h}_{\alpha\beta} + \lambda^2 \overset{2}{h}_{\alpha\beta} + O(\lambda^3) \quad (2.50)$$

因此，公式 2.49 可以按照  $\lambda$  展开为

$$0 = G_{\alpha\beta} = \lambda \overset{1}{G}_{\alpha\beta}[h_{\mu\nu}] + \lambda^2 \left( \overset{1}{G}_{\alpha\beta}[h_{\mu\nu}] + \overset{2}{G}_{\alpha\beta}[h_{\mu\nu}] \right) + O(\lambda^3). \quad (2.51)$$

并且等式对于  $\lambda$  的每一阶都成立，也就是

$$0 = \overset{1}{G}_{\alpha\beta}[\overset{2}{h}_{\mu\nu}] \quad (2.52)$$

以及

$$\overset{1}{G}_{\alpha\beta}[\overset{2}{h}_{\mu\nu}] = -\overset{2}{G}_{\alpha\beta}[\overset{1}{h}_{\mu\nu}] \quad (2.53)$$

在之前的讨论中，我们都只考虑到度规的一阶修正项。然而，根据公式2.53在原先一直被忽略的  $O(h^2)$  阶项会成为一个等效的能量动量张量

$$\overset{1}{G}_{\alpha\beta}[\overset{2}{h}_{\mu\nu}] = \frac{8\pi G}{c^4} T_{\alpha\beta}^{\text{GW}} \quad (2.54)$$

其中  $T_{\alpha\beta}^{\text{GW}} = -\frac{c^4}{8\pi G} \overset{2}{G}_{\alpha\beta}[\overset{1}{h}_{\mu\nu}]$ 。要定义一个规范不变的能量动量张量，必须在一个足够大的范围内平均

$$T_{\alpha\beta}^{\text{GW}} = -\frac{c^4}{8\pi G} \left\langle \overset{2}{R}_{\alpha\beta} - \frac{1}{2} \eta_{\alpha\beta} \overset{2}{R} \right\rangle \quad (2.55)$$

其中  $\langle \dots \rangle$  表示在足够大的范围内平均。可以证明<sup>1</sup>

$$T_{\alpha\beta}^{\text{GW}} = \frac{c^4}{32\pi G} \left\langle \frac{\partial h_{\text{TT}}^{ij}}{\partial x^\alpha} \frac{\partial h_{ij}^{\text{TT}}}{\partial x^\beta} \right\rangle \quad (2.56)$$

若考虑向  $x^3$  方向传播的平面引力波

$$h_{\text{TT}}^{ij} = h_{\text{TT}}^{ij}(x^0 - x^3) = h_+(x^0 - x^3) e_{ij}^+ + h_\times(x^0 - x^3) e_{ij}^\times, \quad (2.57)$$

可以得到

$$T_{00}^{\text{GW}} = -T_{03}^{\text{GW}} = -T_{30}^{\text{GW}} = T_{33}^{\text{GW}} = \frac{c^2}{16\pi G} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle \quad (2.58)$$

其中  $\dot{h} = \frac{\partial h}{\partial t}$ 。更进一步地，如果将平面波按偏振展开

$$h_{\text{TT}}^{ij} = \mathcal{A}_+ \cos \left[ \frac{\omega}{c} (x^0 - x^3) + \phi_+ \right] e_{ij}^+ + \mathcal{A}_\times \cos \left[ \frac{\omega}{c} (x^0 - x^3) + \phi_\times \right] e_{ij}^\times, \quad (2.59)$$

则由于三角函数的平均值为  $1/2$ ，可以得到

$$T_{00}^{\text{GW}} = -T_{03}^{\text{GW}} = -T_{30}^{\text{GW}} = T_{33}^{\text{GW}} = \frac{c^2}{32\pi G} \omega^2 \langle \mathcal{A}_+^2 + \mathcal{A}_\times^2 \rangle \quad (2.60)$$

<sup>1</sup>详细过程可参考 [69]1.4 节

### 2.3.4 引力波能流

既然可以定义引力波的能量，也可以定义引力波的能流，即单位时间内、通过单位面积的引力波所携带的能量  $\frac{dE}{dAdt}$ 。由引力波贡献的能量动量张量同样满足连续性方程  $T_{,\mu}^{\mu\nu} = 0$ ，这就意味着，在距离源足够远的区域  $V$ ，

$$\int_V d^3x \left( \frac{\partial T_{\text{GW}}^{00}}{\partial x^0} + \frac{\partial T_{\text{GW}}^{0i}}{\partial x^i} \right) = 0 \quad (2.61)$$

同时，可以定义区域  $V$  中的引力波能量  $E_V = \int_V d^3x T_{\text{GW}}^{00}$ 。根据 Gauss 公式，可以将公式2.61转换为

$$\begin{aligned} \frac{dE_V}{dx^0} &= \frac{dE_V}{cdt} = - \int_V d^3x \frac{\partial T_{\text{GW}}^{0i}}{\partial x^i} \\ &= - \int_S dA n_i T_{\text{GW}}^{0i}. \end{aligned} \quad (2.62)$$

其中  $S$  是区域  $V$  的边界， $n^i$  是边界的单位外法向。可以证明<sup>1</sup>，

$$\frac{dE_V}{dt} = -c \int_S dA T_{\text{GW}}^{00}. \quad (2.63)$$

由于  $\frac{dE_V}{dt}$  指的是区域  $V$  内的能量变化率，因此引力波的能流为

$$\frac{dE}{dAdt} = c T_{\text{GW}}^{00} = \frac{c^3}{32\pi G} \langle \dot{h}_{TT}^{ij} \dot{h}_{ij}^{TT} \rangle \quad (2.64)$$

### Exercise 2.1

利用公式2.6，结合  $\bar{h}_{0i} = 0$ ,  $\bar{h}_{ij} = 0$ , 以及  $\bar{h}_{00} = -4\Phi$ , 尝试推导  $h_{\mu\nu}$ 。

### Exercise 2.2

在真空中，有  $-\square \bar{h}_{\alpha\beta} = 0$ ，考虑平面波解  $\bar{h}_{\alpha\beta} = A_{\alpha\beta} \exp(ik_\mu x^\mu)$ 。试证明， $k_\mu$  是一个零矢量 (null vector)，即， $\eta^{\mu\nu} k_\mu k_\nu = k^\mu k_\mu = 0$

### Exercise 2.3

一圈粒子围成圆型，在垂直于其平面方向上，有引力波作用。假设引力波为正弦函数  $h(t) = h_0 \sin t$ (建议采用  $h_0 = 0.1$ )。采用长波近似，即粒子间距相对于引力波波长可以忽略不计。请利用公式2.36，分别作图画出： $h_x = 0$ ,  $t = \pi/2$ ;  $h_x = 0$ ,  $t = -\pi/2$ ;  $h_+ = 0$ ,  $t = \pi/2$ ;  $h_+ = 0$ ,  $t = -\pi/2$  时粒子的分布。

(加分题) 作 gif 图，分别展示出  $h_x = 0$  和  $h_+ = 0$  时引力波的效应（交作业时请附上程序源代码，否则成绩无效。）

<sup>1</sup>详细过程可参考 [69]1.4 节

# 3

## 引力波源和波形

引力波数据处理的最终目标是通过测量引力波源的物理参数，更好地理解引力波源的物理性质。因此，我们有必要对引力波源的物理性质，以及广义相对论所预言的引力波波形有足够的了解。

引力波可以由加速的质量四极矩产生，最常见的情形即对应于绕转的非轴对称系统。通常而言，双星系统可以产生较强的时变质量四极矩；而在宇宙中，双星系统又是较为容易出现的，因此我们可以看到，在天文学起源的引力波源中，我们主要讨论的都是双星系统。我们首先在第3.1章中从理论角度出发，描述一个双星系统的引力波。接下来，我们结合真实的天文学模型，讨论更加实际的情形，可以按照质量大小来对各类双星系统分类：两个重的天体（如大质量双黑洞3.2.5.2）、两个轻的天体（如恒星级双黑洞3.2.5.1、中子星双星3.3、双白矮星3.5.2等）、一个重-一个轻的天体（如极端质量比旋近3.4）组成的双星。除此以外，单个天体也能产生引力波信号（如变形的中子星3.5.1、超新星爆发3.7.1等）；许多信号的叠加也能产生引力波信号（随机引力波背景3.6）。在本章中，我们根据引力波源的物理特性开展讨论。

### 3.1 双星系统的引力波

#### 3.1.1 引力波与四极矩

在有引力波传播情况下，可以定义等效能量动量张量  $\tau^{\alpha\beta}$  不仅包括了源场的能量动量张量  $T^{\alpha\beta}$ ，也包含了  $O(h^2)$  的引力波项。

$$\square \bar{h}^{\alpha\beta} = -\frac{16\pi G}{c^4} \tau^{\alpha\beta} \quad (3.1)$$

考虑到我们关心的情况总是发生在远场，即位于  $\mathbf{x}'$  的观测者与引力波源的距离  $D$  远大于波长。这时， $D$  可以近似视作常数，可以得到解

$$\bar{h}^{\alpha\beta} \simeq \frac{4G}{c^4 D} \int \tau^{\alpha\beta}(t - D/c, \mathbf{x}') d\mathbf{x}' \quad (3.2)$$

在远场讨论的另一个好处是我们可以采用横向无迹规范，也就是只需要讨论空间分量  $\tau^{ij}$ 。从守恒律（连续性方程） $\tau_{,\mu}^{\mu\alpha} = 0$  出发，可以将空间分量  $\tau^{ij}$  与 00 分量联系起来：

$$\tau^{ij} = \frac{1}{2} \frac{\partial^2}{c^2 \partial t^2} (x^i x^j \tau^{00}) + \frac{\partial}{\partial x^k} (x^i \tau^{jk} + x^j \tau^{ki}) - \frac{1}{2} \frac{\partial^2}{\partial x^k \partial x^l} (x^i x^j \tau^{kl}) \quad (3.3)$$

远场条件意味着  $T^{00} = 0$  且  $O(h^2)$  的贡献也趋近于零，因此公式 3.3 中包含对空间偏导的部分都是高阶小量可以消掉，可以得到

$$\bar{h}^{ij}(t, \mathbf{x}) \simeq \frac{2G}{c^4 D} \frac{\partial^2}{c^2 \partial t^2} \int x'^i x'^j \tau^{00}(t - D/c, \mathbf{x}') d\mathbf{x}' \quad (3.4)$$

我们可以定义四极矩张量

$$I^{ij}(t) = \int x^i x^j \frac{\tau^{00}(t - D/c, \mathbf{x})}{c^2} d\mathbf{x}, \quad (3.5)$$

从而可以将公式 3.4 简化为

$$\bar{h}^{ij}(t, \mathbf{x}) \simeq \frac{2G}{c^4 D} \ddot{I}^{ij}(t) \quad (3.6)$$

在选取横向无迹规范后，可以得到

$$h_{\text{TT}}^{ij}(t, \mathbf{x}) \simeq \frac{2G}{c^4 D} \ddot{I}_{\text{TT}}^{ij}(t) \quad (3.7)$$

**例 3.1** 通过公式 3.6，我们可以对引力波进行量级估算。考虑一个疯狂旋转的跷跷板：长度  $R = 10\text{m}$  的杆两端分别绑定质量  $M = 10^3\text{kg}$  的重物。<sup>a</sup> 将杆绕质心以  $10\text{Hz}$  的频率快速旋转。这个系统的四极矩  $I \sim MR^2 \sim 10^5\text{kg} \cdot \text{m}^2$ 。四极矩对时间的二阶导数为  $\ddot{I} \sim Mv^2 \sim 10^7\text{kg} \cdot \text{m}^2/\text{s}^2$ （注意表达式和动能的相似性）。

为了计算引力波强度，还需要确定与源的距离。对于上述波源，其引力波的波长为  $1.5 \times 10^7\text{m}$ （考虑到引力波频率是系统频率的两倍），与地球的直径大致相当，根据远场条件要求，引力波描述公式在引力波波长外才成立。假设在北极点设置上述疯狂旋转的“跷跷板”，则在南极处测量的引力波为  $h \sim 5 \times 10^{-43}$ 。

<sup>a</sup>想一想：如果考虑一个直径为十米的环，在环所在的平面上绕圆心快速旋转，其引力波的强度如何？

### 3.1.2 双星系统引力波初探

我们之后将会对各类双星系统开展详细的描述。在这里，我们对最简单的双星系统开展分析，研究以绕转双星作为波源时，引力波的波形与频率、距离、质量、轨道倾角等物理参数之间的关系。

我们考虑两个质量分别为  $m_1$  和  $m_2$  的质点，假设其围绕双星质心做圆周运动。为方便计算，不妨将其轨道面定义为  $x^1 - x^2$  平面，坐标系与双星质心重合，而轨道角动量方向定义为  $x^3$  方向，如图3-5所示。

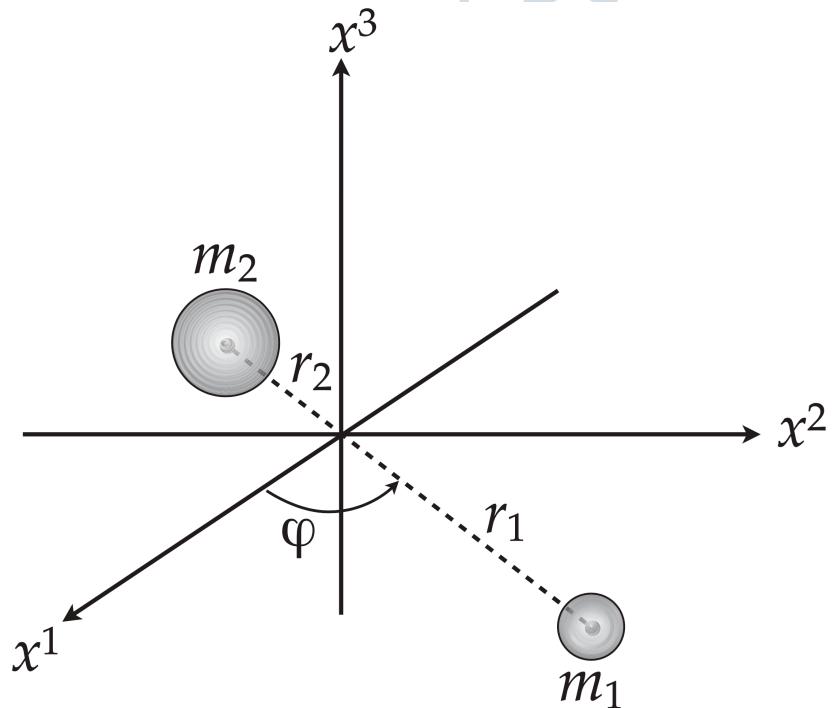


图 3-1 双星系统示意图

Figure 3-1 Illustration of a binary system. Credit: [37]

两星与质心间的距离  $r_1$ ,  $r_2$ , 双星间距离为  $r = r_1 + r_2$ 。对于这样的双星系统，可以通过定义总质量  $M = m_1 + m_2$  和约化质量 (reduced mass)  $\mu = m_1 m_2 / M$  将问题简化。首

先，可以将两星与质心的距离表示为  $r_1 = rm_2/M$  和  $r_2 = rm_1/M$ 。不难计算出有

$$\begin{aligned} I_{11} &= m_1(r_1 \cos \varphi)^2 + m_2 [r_2 \cos(\varphi + \pi)]^2 \\ &= \mu r^2 \cos^2 \varphi = \frac{1}{2} \mu r^2 (1 + \cos 2\varphi) \end{aligned} \quad (3.8)$$

$$\begin{aligned} I_{22} &= m_1(r_1 \sin \varphi)^2 + m_2 [r_2 \sin(\varphi + \pi)]^2 \\ &= \mu r^2 \sin^2 \varphi = \frac{1}{2} \mu r^2 (1 - \cos 2\varphi) \end{aligned} \quad (3.9)$$

$$\begin{aligned} I_{12} &= I_{21} = m_1(r_1 \cos \varphi)(r_1 \sin \varphi) + m_2 [r_2 \cos(\varphi + \pi)][r_2 \sin(\varphi + \pi)] \\ &= \mu r^2 \cos \varphi \sin \varphi = \frac{1}{2} \mu r^2 \sin 2\varphi \end{aligned} \quad (3.10)$$

对于以角速度  $\omega$  作圆周运动的双星而言，其四极矩对时间的二阶偏导为

$$\ddot{I}_{11} = -2\mu r^2 \omega^2 \cos 2\varphi \quad (3.11)$$

$$\ddot{I}_{22} = 2\mu r^2 \omega^2 \cos 2\varphi \quad (3.12)$$

$$\ddot{I}_{12} = \ddot{I}_{21} = -2\mu r^2 \omega^2 \sin 2\varphi \quad (3.13)$$

这意味着，对于处于  $x^3$  轴上距离源  $D$  处的观测者而言，在横向无迹规范下的引力波信号为

$$h_{ij}^{TT} = -\frac{4G\mu r^2 \omega^2}{c^4 D} \begin{pmatrix} \cos 2\varphi & \sin 2\varphi & 0 \\ \sin 2\varphi & -\cos 2\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.14)$$

换言之，两个偏振模式分别为

$$h_+ = -\frac{4G\mu r^2 \omega^2}{c^4 D} \cos 2\varphi \quad (3.15)$$

$$h_\times = -\frac{4G\mu r^2 \omega^2}{c^4 D} \sin 2\varphi \quad (3.16)$$

上述表达式依然是比较特殊的情况，我们假设了观测者正好处于双星系统的角动量方向上。为了得到更一般性的结论，我们考虑观测者的位置并不在  $x^3$  轴的情况。不妨假设观测者位于  $x^2 - x^3$  平面上，且波矢方向与  $x^3$  轴的夹角为  $\iota$ 。可以通过旋转矩阵  $\mathbf{R}_1(\iota)$ ，

得到新的坐标系下沿  $x^{3'}$  轴（观测者方向）的四极矩二阶时间偏导：

$$\begin{aligned}\ddot{\mathbf{I}}' &= \mathbf{R}_1(\iota)\ddot{\mathbf{I}}\mathbf{R}_1^{-1}(\iota) \\ &= -2\mu r^2 \omega^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \iota & \sin \iota \\ 0 & -\sin \iota & \cos \iota \end{pmatrix} \begin{pmatrix} \cos 2\varphi & \sin 2\varphi & 0 \\ \sin 2\varphi & -\cos 2\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \iota & -\sin \iota \\ 0 & \sin \iota & \cos \iota \end{pmatrix} \quad (3.17)\end{aligned}$$

$$= -2\mu r^2 \omega^2 \begin{pmatrix} \cos 2\varphi & \cos \iota \sin 2\varphi & -\sin \iota \sin 2\varphi \\ \cos \iota \sin 2\varphi & -\cos^2 \iota \cos 2\varphi & \sin \iota \cos \iota \cos 2\varphi \\ -\sin \iota \sin 2\varphi & \sin \iota \cos \iota \cos 2\varphi & -\sin^2 \iota \cos 2\varphi \end{pmatrix} \quad (3.18)$$

注意到，经过旋转以后的  $\ddot{\mathbf{I}}'$  并不符合横向无迹规范，因此需要将上述张量投影。首先可以投影到横向张量，即，将第三行/列的分量取零

$$\ddot{\mathbf{I}}'_{\text{transverse}} = -2\mu r^2 \omega^2 \begin{pmatrix} \cos 2\varphi & \cos \iota \sin 2\varphi & 0 \\ \cos \iota \sin 2\varphi & -\cos^2 \iota \cos 2\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.19)$$

注意上述张量的迹为

$$\text{Tr}(\ddot{\mathbf{I}}'_{\text{transverse}}) = -2\mu r^2 \omega^2 (1 - \cos^2 \iota) \cos 2\varphi \quad (3.20)$$

通过对对角元对称地扣除上述迹，可以得到横向无迹规范下的四极矩二阶时间导数张量

$$\ddot{\mathbf{I}}'_{\text{TT}} = 2\mu r^2 \omega^2 \begin{pmatrix} -\frac{1}{2}(1 + \cos^2 \iota) \cos 2\varphi & -\cos \iota \sin 2\varphi & 0 \\ -\cos \iota \sin 2\varphi & \frac{1}{2}(1 + \cos^2 \iota) \cos 2\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.21)$$

对应地，两个偏振分量为

$$h_+ = -\frac{4G\mu r^2 \omega^2}{c^4 D} \frac{1 + \cos^2 \iota}{2} \cos 2\varphi \quad (3.22)$$

$$h_\times = -\frac{4G\mu r^2 \omega^2}{c^4 D} \cos \iota \sin 2\varphi \quad (3.23)$$

不难看出，在  $\iota = 0$  或者  $\iota = \pi$  时，上述双星系统的引力波为圆偏振。而  $\iota = \pi/2$  时，则为线偏振。

## 3.2 双黑洞并合

对于真空中的双黑洞并合过程而言，一般理论并不会预期产生电磁波或高能粒子（如中微子等）的机制。因此，引力波手段是针对双黑洞并合过程几乎唯一的探测手段。

当然，我们总可以讨论处于吸积盘等环境中的双黑洞并合过程以及可能的多信使联合探测。但到目前为止，关于双黑洞的并合过程，仍然没有可信的电磁波等其余信使的观测证据<sup>1</sup>。

这种描述双星并合过程的相对论性两体问题非常复杂，在本节中，主要讨论质量比接近 1 的双星系统。<sup>2</sup>更进一步地，物质的存在会引起更进一步的复杂性，因此我们讨论更简化的问题，这种最简化的问题就是双黑洞并合的问题。

### 3.2.1 双黑洞旋近

双黑洞并合的整个过程可以分为三个阶段：inspiral（旋近），merger（并合）和 ringdown（铃宕），图3-2所示。在旋近阶段，可以对 Einstein 场方程进行 post-Newtonian 近似，得到所谓 post-Newtonian(PN) 波形。在并合阶段，必须利用数值相对论 (numerical relativity) 才能准确计算系统的演化和相应波形。在铃宕阶段，可以利用微扰法去计算 Kerr 黑洞辐射的准正则模 (quasi-normal mode, QNM)。

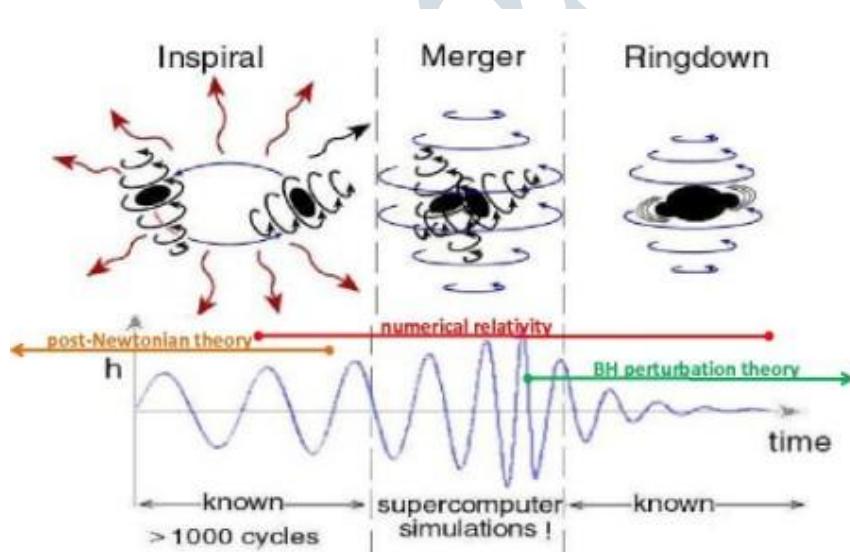


图 3-2 双黑洞并合过程的三个阶段：旋近、并合和铃宕

Figure 3-2 Three stages of binary black hole merger: inspiral, merger and ringdown. Credit: [51]

<sup>1</sup>注意，在 LVC 报告的 GW150914 探测之后约 0.4 秒后，Fermi 的伽马射线暴检测器 (Gamma-ray Burst Monitor, GBM) 报告了一个持续时间约 1 秒，置信度约  $2.9\sigma$  的电磁信号成协，[36]，但是如 INTEGRAL 等设备都没能确认这一发现 [85]。另外，Graham et al. 在 [47] 中汇报了 GW190521 的可能的电磁对应体的发现，但是对应的红移与 GW190521 相差较大。

<sup>2</sup>质量相差悬殊的双星系统的演化和引力波波形尤为复杂，因此将极端质量比旋近 (EMRI) 单独列出于3.4。

### 3.2.1.1 旋近与 Newtonian 近似

在旋近阶段，两个黑洞距离较远，这时，可以用第3.1.2节的公式描述这一双星系统的引力波。当然，由于系统持续辐射引力波，系统的能量不断衰减，因此黑洞的轨迹并非封闭轨道，而是半径不断衰减的螺旋线。但在两黑洞距离较远时，轨道衰减的时标远长于双星绕转时标，这时双星的轨道是准圆 (quasi-circular) 轨道。<sup>1</sup> 在之后的计算中，我们会发现，虽然引力波的描述来自广义相对论，但对双黑洞系统的引力波特性分析，除了双星系统的引力波光度需要用到前面章节推导的知识外，大部分推导仅需大一水平的物理和数学知识即可完成。

对于一个引力波探测器而言，其对 + 模和 × 模的响应强度各不相同。因此，探测器记录到的引力波信号是两种极化的线性叠加

$$h(t) = F_+(\text{angles})h_+(t) + F_\times(\text{angles})h_\times(t). \quad (3.24)$$

具体的天线响应函数  $F_{+/\times}$  表达式及其推导，我们将在第四章具体展开 (见公式4.41, 4.70-4.73)。在这里，不妨把与响应强度有关的部分统一归到  $Q(\text{angles})$  中，表示引力波探测器对来自不同方位、偏振、倾角的双星系统的响应强度。因此，探测器实际记录到的信号为

$$h(t) = \Re\{Q(\text{angles}) \frac{G\mu r^2(t)\omega^2}{c^4 D} \exp(i\Phi(t))\} \quad (3.25)$$

注意到，在公式3.25中，随时间变化的量只有  $r(t)$  和  $\Phi(t)$ 。下面，我们针对这两个量展开推导。

首先， $r(t)$  的随时变化，根本原因来源于引力波的辐射带走了系统能量。因此，可以通过分析轨道衰减率  $\frac{dr}{dt}$ ，并对其积分，获得双星间距随时间变化的函数  $r(t)$ 。可以利用链式法则推导出双星间距的变化率  $\frac{dr}{dt} = \frac{dr}{dE} \frac{dE}{dt}$ 。注意到第二项是双星系统引力波光度的负数  $L_{\text{GW}} = -\frac{dE}{dt}$ 。下面展开具体推导。

可以证明，在准圆轨道下，利用 Newtonian 引力，可以推导出 Kepler 第三定律

$$\omega = \sqrt{\frac{GM}{r^3}} \quad (3.26)$$

并且由此可以得到双星系统的能量

$$E = -\frac{1}{2}\mu\omega^2 r^2. \quad (3.27)$$

<sup>1</sup>由于轨道的圆化效应 circularization，在并合时的双星轨道偏心率几乎都为零。

因此，双星能量  $E$  与双星间距  $r$  的关系符合  $\frac{dr}{dE} = -\frac{r}{E}$ 。另一方面，利用引力波能流（公式2.64），可以推算出上述双星系统的光度为

$$\begin{aligned} L_{\text{GW}} &= -\frac{dE}{dt} = \frac{1}{5} \frac{G}{c^5} \langle \ddot{I}_{ij} \ddot{I}^{ij} \rangle \\ &= \frac{32}{5} \frac{G^4}{c^5} \frac{\mu^2 M^3}{r^5} \end{aligned} \quad (3.28)$$

不难发现，随着时间的推移，双星系统的轨道不断衰减，引力波光度不断增加，因此引力波的振幅随着时间推移而增强。

$$\frac{dr}{dt} = -\frac{r}{E} \frac{dE}{dt} = -\frac{64G^3}{5c^5} \frac{\mu M^2}{r^3} \quad (3.29)$$

对3.29式积分，可以得到

$$r = \left( \frac{256G^3}{5c^5} \mu M^2 \right)^{1/4} (t_c - t)^{1/4} \quad (3.30)$$

其中  $t_c$  是  $r \rightarrow 0$  时对应的时刻，我们称之为并合时刻。

 **注 3.2.** 从形式上看，旋近阶段的 *Newtonian* 近似似乎可以描述一直到并合的整个过程，但是当双星系统靠得足够近时，广义相对论效应开始体现，*Newtonian* 及至 *post Newtonian* 描述都会失效。一个较为合适的分界线可以选取为最内稳定圆轨道 *innermost stable circular orbit, ISCO*。其对应的引力波频率是  $f_{\text{ISCO}} = \frac{c^3}{6^{3/2} \pi G M}$ 。关于 *ISCO* 的具体推导，可以参考 [80] 第 5.6 章。

接下来，我们分析引力波的相位演化  $\Phi(t)$ 。实际上，在我们的推导中，我们从圆周运动的公式出发得出了一些结果，但准圆轨道的本质意味着这种定义并不是十分严谨的。比如，圆频率  $\omega$  与双星间距  $r$  有关，而  $r$  又是一个时变函数，所以  $\omega$  其实是随时间变化的。如何理解这一点的？实际上，我们不妨反过来，对于圆周运动而言（不一定是匀速圆周运动），总可以定义相位

$$\Phi = \int \frac{d\Phi(t)}{dt} dt \quad (3.31)$$

由此，我们可以定义瞬时频率  $\omega_{\text{GW}}(t) = 2\pi f_{\text{GW}}(t) = \frac{d\Phi(t)}{dt}$ 。注意，其中  $\omega_{\text{GW}}$  和  $\omega$  存在两倍关系，这是因为在双星系统转过一圈时，根据注2.6，引力波的相位转过了两圈，也就是  $\omega_{\text{GW}} = 2\omega$  和  $f_{\text{GW}} = 2f_{\text{orb}}$ 。

结合式3.26与式3.29，可以得到引力波频率的演化

$$\begin{aligned} \frac{df}{dt} &= \frac{df_{\text{GW}}}{dt} = \frac{df_{\text{GW}}}{d\omega_{\text{orb}}} \frac{d\omega_{\text{orb}}}{dr} \frac{dr}{dt} \\ &= \frac{1}{\pi} \sqrt{GM} \left( -\frac{3}{2} \right) r^{-5/2} \left( -\frac{64G^3}{5c^5} \right) \frac{\mu M^2}{r^3} \\ &= \frac{96}{5} \frac{G^{7/2}}{c^5 \pi} \frac{\mu M^{5/2}}{r^{11/2}} \end{aligned} \quad (3.32)$$

更进一步地，继续将公式3.26代入，可以有，

$$\begin{aligned}\frac{df}{dt} &= \frac{96}{5} \frac{G^{7/2}}{c^5 \pi} \frac{\mu M^{5/2} \pi^{11/3} f^{11/3}}{G^{11/6} M^{11/6}} \\ &= \frac{96}{5} \frac{G^{5/3}}{c^5} \mu M^{2/3} \pi^{8/3} f^{11/3} \\ &= \frac{96}{5} \left(\frac{GM}{c^3}\right)^{5/3} \pi^{8/3} f^{11/3}\end{aligned}\quad (3.33)$$

在最后一行，我们定义了**啁啾质量 chirp mass**  $\mathcal{M} = \mu^{3/5} M^{2/5}$ 。可以发现，伴随着时间的推移，双星轨道衰减，而双星频率增加。

从  $t$  时刻演化到并合时刻  $t_c$  的过程中，双星系统的频率从  $f$  增长到  $\infty$ 。因此，可以对式3.33变形，并积分得

$$\begin{aligned}\int_f^\infty df' f'^{-11/3} &= \int_t^{t_c} dt' \frac{96}{5} \left(\frac{GM}{c^3}\right)^{5/3} \pi^{8/3} \\ -\frac{3}{8} f'^{-8/3} \Big|_f^\infty &= \frac{96}{5} \left(\frac{GM}{c^3}\right)^{5/3} \pi^{8/3} (t_c - t) \\ f^{-8/3} &= \frac{256}{5} \left(\frac{GM}{c^3}\right)^{5/3} \pi^{8/3} (t_c - t) \\ f &= \frac{5^{3/8}}{8\pi} \left(\frac{c^3}{GM}\right)^{5/8} (t_c - t)^{-3/8}\end{aligned}\quad (3.34)$$

最终，利用  $\frac{d\Phi}{dt} = 2\pi f_{\text{GW}}$ ，可以积分得到引力波相位。

$$\begin{aligned}\int_\Phi^{\Phi_c} d\Phi' &= \int_t^{t_c} dt' 2\pi \frac{5^{3/8}}{8\pi} \left(\frac{c^3}{GM}\right)^{5/8} (t_c - t)^{-3/8} \\ \Phi_c - \Phi &= \frac{5^{3/8}}{4} \left(\frac{c^3}{GM}\right)^{5/8} \int_t^{t_c} dt' (t_c - t')^{-3/8} \\ \Phi_c - \Phi &= \frac{5^{3/8}}{4} \left(\frac{c^3}{GM}\right)^{5/8} \frac{8}{5} \int_0^{t_c-t} d(t_c - t') (t_c - t')^{-3/8} \\ \Phi &= \Phi_c - 2 \left(\frac{c^3}{5GM}\right)^{5/8} (t_c - t)^{5/8}\end{aligned}\quad (3.35)$$

回顾公式3.25，我们发现，其中与时间相关的两个量  $r(t)$  和  $\Phi(t)$  现在均已解析给出。感兴趣的读者可以据此画出引力波的时间域波形了。

需要注意的是，在上述推导过程中，只考虑了质量四极矩的引力波效应。在近场（引力波源附近）情形下，采取低速近似 ( $v/c \ll 1$ )，则可以对  $v/c$  作小量展开。<sup>1</sup> 纳入高极矩项的贡献，同时在引力波辐射功率  $\frac{dE}{dt}$  也截断到  $v/c$  的更高阶，则可以得到 post Newtonian 修正的引力波波形。通常，选择以  $(v/c)^2$  的阶数作为 post Newtonian 展开的阶数，故此有 1.5PN, 2.5PN 等。<sup>2</sup> 在第七章中可以看到，引力波的相位效应可以较为准确地被测量，但其振幅效应的测量精度较差。因此，一些文献中会讨论所谓 restricted PN 波形，即对相位采用更为准确的高阶表达式，而对振幅采用较为粗糙的低阶表达式。

<sup>1</sup>为了更清楚地看出这一点，我们可以引入速度  $v = \omega r = (\pi GMf)^{1/3}$ ，则引力波光度（公式3.28）可以重新表述为  $L_{\text{GW}} = \frac{32}{5} \frac{c^5}{G} \frac{\mu^2}{M^2} \left(\frac{v}{c}\right)^{10}$

<sup>2</sup>具体的展开方式篇幅较长且与课程关系不紧密，请有兴趣的同学参阅 [80] 的第 8-11 章，或 [69] 第 5 章。

### 3.2.1.2 旋近信号的频域波形

在第**四**章、**六**章等章节中我们会看到，在频率域开展引力波数据处理往往有着巨大的优势。因此，我们有必要研究双黑洞旋近信号的 Fourier 变换。我们定义引力波波形的 Fourier 变换为

$$\tilde{h}(f) = \int_{-\infty}^{\infty} dt \exp(i2\pi ft) h(t) \quad (3.36)$$

解析地对波形积分是较为困难地，但是，利用稳相近似 stationary phase approximation (SPA)，则可以得到近似的频率域波形。对于  $B(t) = A(t) \cos \phi(t)$ ，如果有  $d \ln(A)/dt = d(A)/Adt \ll d\phi(t)/dt$ ，并且  $d^2\phi(t)/dt^2 \ll (d\phi(t)/dt)^2$ ，可以得到

$$\tilde{B}(f) \approx \frac{1}{2} A(t) \left( \frac{df}{dt} \right)^{-1/2} \exp \{ i[2\pi ft - \phi(f) - \pi/4] \} \quad (3.37)$$

其中  $\phi(f) = \phi(t(f))$ 。为了得到具体的表达式，需要先将频率作为自变量表示出时间  $t$ 。将公式**3.34**的倒数第二行代入到公式**3.35**，可以得到

$$\begin{aligned} \Phi &= \Phi_c - 2 \left( \frac{c^3}{5GM} \right)^{5/8} \left[ \frac{5}{256} \left( \frac{c^3}{GM} \right)^{5/3} \pi^{-8/3} f^{-8/3} \right]^{5/8} \\ &= \Phi_c - 2 \left( \frac{c^3}{8G\pi M f} \right)^{5/3} \end{aligned} \quad (3.38)$$

对应到引力波波形中，引力波相位的变化远快于其振幅的变化，稳相近似适用，因此可以得到

$$\tilde{h}(f) = \sqrt{\frac{5}{384}} \frac{Q}{D} G^{5/6} c^{-3/2} M^{5/6} \pi^{-2/3} f^{-7/6} \exp[i\Psi(f)] \quad (3.39)$$

对于正频率部分  $f > 0$ ，相位项  $\Psi(f)$  为

$$\Psi(f) = 2\pi f t_c - \Phi_c - \frac{\pi}{4} + \frac{3}{4} \left( \frac{8\pi GM f}{c^3} \right)^{-5/3} \quad (3.40)$$

#### 材料 3.3 Stationary Phase Approximation

让我们看一下稳相近似的具体推导过程。将引力波信号分为振幅  $A$  和相位  $\Phi$  两部分，即， $h(t) = A(t_{\text{ret}}) \cos \Phi(t_{\text{ret}})$ ，其中  $t_{\text{ret}} = t - \frac{r}{c}$  是推迟时间。

$$\begin{aligned} \tilde{h}(f) &= \int_{-\infty}^{\infty} dt A(t_{\text{ret}}) \cos \Phi(t_{\text{ret}}) \exp(2\pi ift) \\ &= \frac{1}{2} \exp(i2\pi fr/c) \int_{-\infty}^{\infty} dt_{\text{ret}} A(t_{\text{ret}}) \left( \exp[i\Phi(t_{\text{ret}})] + \exp[-i\Phi(t_{\text{ret}})] \right) \exp(2\pi ift_{\text{ret}}) \\ &\approx \frac{1}{2} \exp(i2\pi fr/c) \int_{-\infty}^{\infty} dt A(t) \exp \left[ i(2\pi ft_{\text{ret}} - \Phi(t)) \right] \end{aligned} \quad (3.41)$$

在第三行，我们略去了快速震荡的小项。

我们假设在  $t_s$  时刻，有  $\dot{\Phi}(t_s(f)) = 2\pi f$  成立，那么围绕  $t = t_s$  对  $2\pi ft - \Phi(t)$  作 Taylor 展开，就有

$$2\pi ft - \Phi(t) = 2\pi ft_s - \Phi(t_s) - \frac{1}{2}\ddot{\Phi}(t_s)(t - t_s)^2 + O(t - t_s)^3 \quad (3.42)$$

进而有

$$\tilde{h}(f) \approx \frac{1}{2}A(t_s(f))e^{i2\pi fr/c}e^{i(2\pi ft_s(f)-\Phi(t_s(f)))}\left(\frac{2}{\ddot{\Phi}(t_s(f))}\right)^{1/2}\int_{-\infty}^{\infty}dx e^{-ix^2} \quad (3.43)$$

又因为有 Fresnel integral

$$\int_{-\infty}^{\infty}dx e^{-ix^2} = \sqrt{\pi}e^{-i\pi/4} \quad (3.44)$$

最后化简为

$$\tilde{h}(f) \approx \frac{\sqrt{\pi}}{2}A(t(f))e^{i2\pi fr/c}e^{i\Psi(f)}\left(\frac{2}{\ddot{\Phi}(t_s(f))}\right)^{1/2} \quad (3.45)$$

而其中

$$\Psi(f) = 2\pi ft(f) - \Phi(t(f)) - \pi/4 \quad (3.46)$$

### 3.2.2 数值相对论与并合

在双黑洞并合阶段，低速、弱场等近似条件纷纷失效，故此无法用解析方法得到可靠的引力波波形。而这时最可靠的方法就是对场方程进行数值求解。然而，这种利用数值方法求解广义相对论方程的思路在过去的半个世纪里艰难前行，在本节中我们简单回顾数值相对论的过往。

我们在第2.2.2节中看到，虽然度规张量  $g_{\mu\nu}$  有 10 个分量，但是其中的部分自由度是冗余的。通过寻找合适的坐标系，可以简化问题。较为常用的做法是，首先选取类时坐标  $t$ ，然后确定剩余三个类空的坐标  $x^i$ 。在这类对时空的所谓  $3+1$  分解中，最常用的有 Arnowitt-Deser-Misner (ADM) 分解。<sup>1</sup>

这样，Einstein 的 10 个场方程就可以被拆分成 6 个动力学方程和 4 个针对初始条件的限制方程。由于截断误差的存在，初始条件可能无法精确满足限制方程，而随着动力学系统的演化，这种差异可能被放大。近年来，数值相对论取得突破的一大要素，就是发展了一套新的场方程表达式，使得违背限制方程的项的生长速率得到控制。

<sup>1</sup>感兴趣的读者可以参阅 [37] 的第 4.3.1 节。

对于引力波研究而言，仅仅研究动力学系统在广义相对论方程下的演化是不够的，还需要从系统中提取引力波波形。可以根据 Weyl 张量  $C_{\alpha\beta\gamma\delta}$ ，定义出 Weyl 标量  $\Psi_4 = -k^\mu m^{*\nu} k^\rho m^{*\sigma} C_{\mu\nu\rho\sigma} = -\ddot{h}_+ + i\ddot{h}_\times$ 。在远场下， $\Psi_4$  就包含了引力波波形的完整描述。

### 3.2.3 QNM 与铃宕

受扰动黑洞所产生的铃宕阶段引力波信号可以由多个随时间振荡衰减的准正则模 (quasinormal mode, QNM) 叠加构成，即有复本征值 (频率) 的线性微分方程的形式解。这些准正则模的振荡频率和衰减时间可以通过求解在边界上附加了合适边界条件的扰动方程得到。广义相对论认为：黑洞的最终性质都仅由几个物理量唯一确定，即黑洞都可以通过少数几个“毛”(荷)来描述，这就是黑洞的无毛定理。若广义相对论预言的无毛定理在强场下依然有效，则每一个准正则模的振荡频率和衰减时间都由黑洞的质量和自旋唯一确定。

在广义相对论框架下，Kerr 黑洞的准正则模通常由  $l, m, n$  三个指标来表示，其中  $n = 0, 1, 2, 3, \dots$  是 overtone 指标， $l, m$  是谐和指标。 $n = 0$  被称为基模，通常相较于  $n \geq 1$  的模式幅度更大且有更长的衰减时间，更容易被探测器探测到。QNM 的波形可以表示为：

$$\begin{aligned} h_{+, \times}(t) &= \frac{M_z}{D_L} \sum_{l, m > 0} A_{lm} Y_{+, \times}^{lm}(\iota) \Psi_{lm}^{+, \times}(t) \\ \Psi_{lm}^+(t) &= \exp\left(-\frac{t}{\tau_{lm}}\right) \sin(\omega_{lm} t - m\phi_0) \\ \Psi_{lm}^\times(t) &= -\exp\left(-\frac{t}{\tau_{lm}}\right) \cos(\omega_{lm} t - m\phi_0) \end{aligned} \quad (3.47)$$

这其中  $M_z$  是红移后的质量， $\phi_0$  是初始相位； $A_{lm}, \omega_{lm}, \tau_{lm}$  分别是  $(l, m)$  模式下的幅度、振荡频率和衰减时间；而  $Y_{+, \times}^{lm}(\iota)$  可以表达成多个自选权重为 2 的球谐函数的和

$$\begin{aligned} Y_+^{lm}(\iota) &= {}_2Y^{lm}(\iota, 0) + (-1)^l {}_{-2}Y^{l-m}(\iota, 0) \\ Y_\times^{lm}(\iota) &= {}_2Y^{lm}(\iota, 0) - (-1)^l {}_{-2}Y^{l-m}(\iota, 0) \end{aligned} \quad (3.48)$$

### 3.2.4 EOB 波形与 IMR 波形

对于质量比接近的双黑洞引力波波形模板的研究，主要有 EOB 波形和 IMR 波形。

在 Newtonian 力学中，双星系统的演化可以通过研究有心力场下约化质量  $\mu = \frac{m_1 m_2}{m_1 + m_2}$  而得到，本质上这是利用了系统的对称性，消除冗余自由度实现的。在广义相对论下，

类似的对称性依然存在，Damour 和 Buonanno 等人由此发展了一套等效单体（effective one body, EOB）体系，来描述双星系统的引力波波形。

EOB 方法的思路是，寻找一个单体系统，使得其 Hamiltonian 与两体系统的一致。自然，在 Newtonian limit 下，这一思路是可行的。经过一系列复杂的计算以后，可以证明在 2PN 下，双星的演化可以由一个约化质量  $\mu$  在变形的 Schwarzschild 度规<sup>1</sup> 描述。并且，可以证明，这种从双星系统到等效单体问题的映射是严格的，可以将其拓展到更高阶的 PN 项。

这对于引力波数据处理是十分重要的，因为 PN 波形只能作为双星系统演化到并合之前的近似，只能较为有效地近似描述旋近波形。反过来，EOB 方法对于双星系统的描述可以一直贯穿旋近、并合、铃宕的全过程。当然，EOB 方法中等效的 Schwarzschild 度规也同样有视界等概念，在约化质量接近视界范围时，描述系统演化的公式发生变化，但是这种 EOB-双星系统的对应关系依然成立。现在，通过将并合阶段与数值相对论波形、准正则模波形相比对、校准，即可由 EOB 方法给出双星并合全过程的引力波波形。这种研究的意义在于，目前数值相对论的计算极为耗时，而通过校准后，EOB 方法可以以极快的速度产生相对准确的波形，这为快速开展引力波数据处理提供了保障。

与 EOB 波形类似的还有 IMR 波形，也是试图通过拟合方法快速产生旋近、并合、铃宕全过程波形。事实上，IMR 即得名于双星并合的三个阶段 (Inspiral-Merger-Ringdown)。该方法以解析波形为基础，乘以一个形如  $1 + \sum_k a_k (\pi M f)^{k/3}$  的修正项，并通过与数值相对论的结果比对，确定不同阶段下的系数  $a_k$ 。

### 3.2.5 BBH 天文学

英国天文学家 John Michell 和法国博物学家 Pierre-Simon Laplace 分别于 1783 年和 1796 年独立提出“黑星”的概念，认为当星体引力场极大时，其逃逸速度将超过光速。Oppenheimer 和 Snyder 在 1939 年利用广义相对论推导得出结论：一个球对称的物体，在引力作用下会塌缩，并且该天体会与远处的观测者断绝一切联系，只有引力效应可以保留。然而，由于上述推导假设了球对称条件，包括 Einstein 本人都认为真实宇宙中不会包含黑洞。Lifshitz 和 Khalatnikov 甚至论证认为在一切实际情况下，物质都不会准确地塌缩到一个点上而产生黑洞。Wheeler 也怀疑量子效应会阻止奇点的形成。

恒星究其一生都在通过核聚变产生的辐射压和气体压强抵抗引力作用。在核球区的核燃料燃烧殆尽后，恒星的核聚变反应终将迎来终结，而恒星也不可能避免的迎来引力塌缩的命运。根据残余物质质量的不同，恒星塌缩后分别可以形成白矮星、中子星和黑洞。

在 Advanced LIGO (aLIGO) 运作之前，人们对于双黑洞存在的观测证据是相对较

<sup>1</sup>Schwarzschild 度规是球对称黑洞的广义相对论解。

弱的，因为黑洞本身难以产生可观测的效应。不过，在由黑洞和恒星形成的双星系统中，黑洞可以通过对伴星的吸积，产生可观测的 X 射线辐射。如天鹅座 X-1(Cygnus X-1) 就是如此被发现，并最终被证实是黑洞的。围绕着这个发现，Stephen Hawking 还曾和 Kip Thorne 打赌并输掉了赌局。(见图3-3)



图 3-3 Hawking 与 Thorne 关于 Cyg X-1 的赌约

Figure 3-3 Bet over Cyg X-1 between Stephen Hawking and Kip Thorne. Credit:

<https://www.star.le.ac.uk/sav2/blackholes/xrays.htm>

近年来，各种观测手段的成熟与应用，使得与黑洞有关的研究逐渐成为研究热点，黑洞存在的各类证据越来越丰富。2020 年的诺贝尔物理学奖，一半授予了 Roger Penrose (“for the discovery that black hole formation is a robust prediction of the general theory of relativity”), 另一半由 Reinhard Genzel 和 Andrea Ghez 分享 (“for the discovery of a supermassive compact object at the centre of our galaxy” )

20 实际 60 年代对类星体的发现促进了对大质量致密天体的研究，Penrose 借助俘获面 (trapped surface) 这一工具，论证在一般性的条件下，物质塌缩可以产生黑洞。Penrose 还通过保角变换 (conformal transformation)，引入了 Penrose 图，更方便理解物理图景。

他还提出了 Penrose 机制，可以从旋转黑洞中提取能量。

要在观测上确认黑洞的存在，需要证明质量包含在极小尺度内，在不考虑相对论修正的前提下，距离黑洞  $r$  的天体，其运动速度应该与  $r^{-1/2}$  成正比。若是质量分布较为弥散，则观测结果应当与上述结论矛盾。Genzel 和 Ghez 就是利用大口径望远镜对银心进行长期观测，确认了上述关系。通过积累长时间的观测，更是可以观测到天体的运行轨道，确认了在银心处有数百万太阳质量的天体。而已知的所有稳定多体系统是球状星团，其密度大约  $10^5 M_\odot pc^{-3}$ ，无法解释观测到的至少  $5 \times 10^{15} M_\odot pc^{-3}$  的密度。

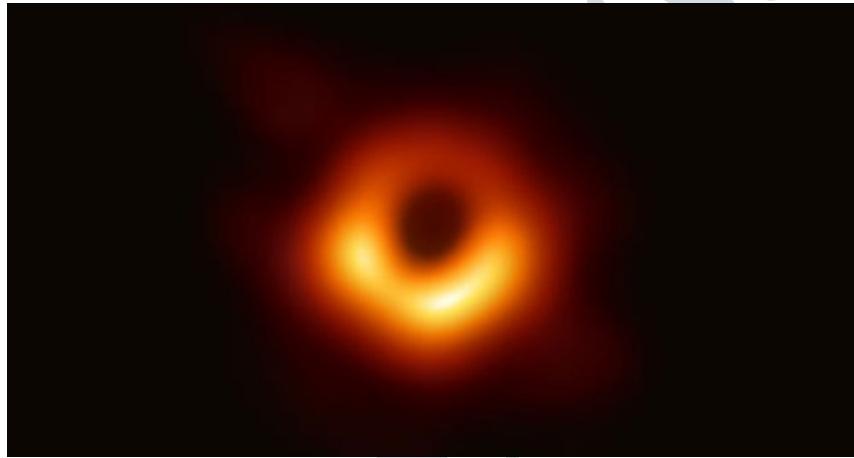


图 3-4 EHT 合作组拍摄到的 M87 星系中心的大质量黑洞。

Figure 3-4 The black hole image in the middle of the M87 galaxy observed by EHT collaboration. Credit: [15]

事件视界望远镜（Event Horizon Telescope, EHT）通过甚长基线干涉阵的观测，合成了一个等效口径与地球表面积相近的高分辨率射电望远镜，进而可以观测黑洞视界附近的现象。[\[15\]](#) 2019 年 4 月 10 日，EHT 项目合作组发表了“黑洞照片”，如图3-4所示。对 EHT 在 2017 年 4 月份的四个夜晚的联合观测的数据分析显示，在 M87 星系中心，EHT 观测到了一个非对称的直径为  $42 \pm 3 \mu\text{as}$  的致密射电源。这一现象与质量为  $(6.5 \pm 0.7) \times 10^7 M_\odot$  的 Kerr 黑洞所产生的阴影一致。其周围非对称的亮环可以由等离子体辐射的相对论性射束（relativistic beaming）效应解释。

### 3.2.5.1 恒星级

LIGO 和 Virgo 等地面引力波探测器在过去的几年时间里取得了令人瞩目的成就，其对 GW150914 等引力波事件的探测，使得 LIGO 科学合作组成员获得了基础物理突破奖，更使得 Rainer Weiss, Barry C. Barish 和 Kip S. Thorne 三位 LIGO 的重要创始人获得

了 2017 年的 Nobel 奖，“for decisive contributions to the LIGO detector and the observation of gravitational waves”。以第一例引力波事件 GW150914 为例的双黑洞并合也让我们对于恒星级双黑洞有了更深入的理解。

一般定义质量  $M \lesssim 100M_{\odot}$  的黑洞为恒星级黑洞。在 LIGO 实现引力波探测之前，通过电磁手段对黑洞的研究主要认为黑洞的质量小于约  $20M_{\odot}$ ，大部分集中于  $5 - 10M_{\odot}$ 。这也与之前的恒星演化模型一致，由于金属<sup>1</sup>的存在，会增大恒星中光子与物质作用的散射界面，增强辐射场到物质的动量转移，进而导致强的星风过程，使最终残余的致密星质量偏小。然而，以 GW150914 为例的引力波事件表明，黑洞双星的质量可以大大超过之前认为的上限，这为黑洞双星的形成机制带来了挑战。一个可能的解释是，这些双黑洞的前身星拥有极低的金属丰度，因此残余致密天体的质量可以很大；另外一个是，这些黑洞并非由恒星产生，而是所谓的原初黑洞 (primordial black hole)，这是由于宇宙早期过程中局部密度涨落导致部分区域引力塌缩产生的黑洞。

LIGO 等地面引力波探测器对双黑洞并合辐射的引力波信号的探测，表明了黑洞不仅存在，还会成对存在，并在足够短的时间内并合。恒星级双黑洞的形成，主要有三种可能性：1. 共同演化；2. 动力学俘获；3. 原初黑洞双星。在第一种机制中，黑洞的前身星就是独立的双恒星系统，并经过动力学和化学演化，最终形成黑洞双星。由于双星系统经历过公共包层阶段 (common envelope)，预期双黑洞的自旋方向一致，并与轨道角动量一致。在动力学俘获模型中，双黑洞处于密集星体环境，如球状星团、星系核区等区域中，两个黑洞分别单独形成，通过动力学俘获形成双星，并通过三体相互作用等机制进一步加强引力束缚。这种模型下，黑洞的自旋方向可以任意，近乎各向均匀。第三种机制就是两个原初黑洞分别独立产生，并形成双星系统。

双星共同演化存在一些机制，可能会阻止双黑洞的形成。比如，双星系统中，首先有一颗恒星通过超新星爆发形成黑洞，这个过程可能会通过超新星爆发的不对称性，引入较大的 natal kick，有可能打破双星间的引力束缚。较强的 natal kick 甚至能将黑洞踢出星团的引力势，所以也可能会影响动力学俘获的双黑洞形成机制。即使双星可以在第一个黑洞形成时存活，第二颗恒星末期会膨胀，经历所谓公共包层阶段，将包裹业已形成的黑洞。因为黑洞通过摩擦可能会与恒星核区并合，所以可能就无法形成第二颗黑洞，也就无所谓双黑洞系统了。

目前对于双黑洞并合信号的探测，一定程度上排除了这些阻碍双黑洞形成、演化的机制的普遍存在性。另一方面，现有的测量表明，GW170104 中黑洞的自旋与轨道角动量有非零夹角，倾向于认为双黑洞并非来自双星协同演化。

到目前为止，恒星级双黑洞的观测，均来自 LIGO/Virgo 等地面引力波探测的结果。

<sup>1</sup> 天文学中，“金属”指除了氢、氦以外的所有元素

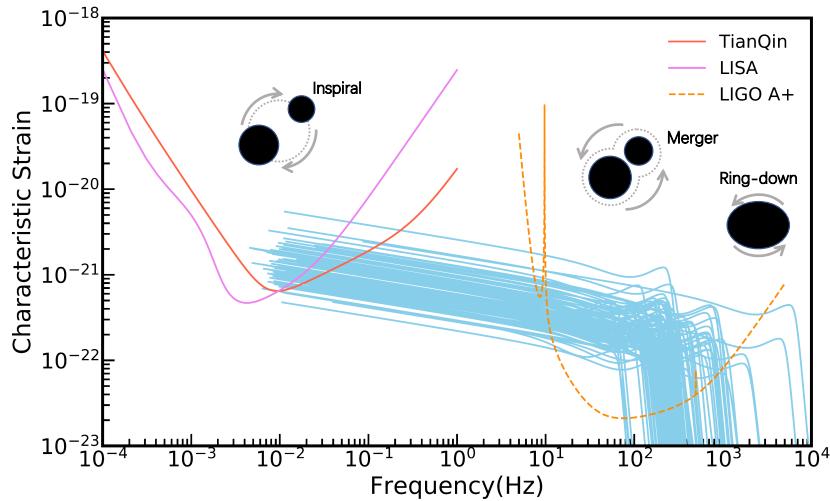


图 3-5 多波段引力波探测示意

Figure 3-5 Illustration of the multi-band gravitational wave detection. Credit: Shuai Liu

虽然恒星级双黑洞的并合发生在百赫兹的高频频段，但在其并合前数年，其早期旋近引力波信号位于毫赫兹频段，可以被天琴等空间引力波探测器探测到。由此，这类引力波信号构成了一类理想的多波段引力波探测源。

地面观测结果给出了恒星级双黑洞的族群性质，包括并合率和质量分布等信息。根据这些信息，刘帅等人评估了天琴对恒星级双黑洞信号的预期探测数和参数估计精度。研究结果显示，如果把探测阈值选为信噪比  $\rho \geq 12$ ，则天琴在五年时间内可以探测到数起恒星级双黑洞的旋近信号。若将探测阈值降低，或者采用多探测器网络，都可以提高探测数。对于并合时刻，典型的限制精度可以达到 1 秒；对于初始偏心率，其相对精度可以达到  $10^{-4}$ ；对于质量参数，可以限制到  $10^{-6} - 10^{-7}$  的相对精度 [65]。特别注意到，对于一个典型的恒星级双黑洞旋近信号，天琴预期可以获得较高精度的三维空间定位。在这么小的空间范围内，统计上来说预期只会包含其宿主星系。结合天琴观测得到的光度距离，和宿主星系给出的红移信息，有望获得较高精度的宇宙学参数限制能力。

#### 材料 3.4 Mass Gap

通过 X 射线观测发现的黑洞，其质量通常大于约  $5M_{\odot}$ ，而中子星的质量上限为约  $3M_{\odot}$ 。这里就形成了一个质量间隙 (mass gap)。另一方面，理论预言恒星演化无法产生  $45 - 150M_{\odot}$  区间内的恒星级黑洞。恒星演化模型认为，随着前身星质量的增加，最终残余的致密星质量也越大，但这一关系只会持续到质量小于约  $45M_{\odot}$  的黑洞

为止。更大质量的恒星会经历所谓生对不稳定性超新星 (pair-instability supernova) 爆发，由于恒星内核温度极高，高能光子可以产生正负电子对。这种光子-正负电子对的平衡在一定条件下是不稳定的，一旦被打破，会导致恒星收缩形成热核爆炸，以超新星的形式将整个恒星炸开，不剩下任何残余。不过，当前身星质量超过约  $260M_{\odot}$  时，恒星核区会产生光致蜕变，抑制高能光子和正负电子对之间的不稳定平衡，从而避免灾变性的超新星爆发，并在核反应停止后产生质量大于约  $150M_{\odot}$  的黑洞。通常把  $3 - 5M_{\odot}$  的质量间隙称为低质量间隙，而把约  $45 - 150M_{\odot}$  的质量间隙称为高质量间隙。

GW200210\_092254 的次星质量测量值为  $2.83^{+0.47}_{-0.42}M_{\odot}$ ，有一定几率落在低质量间隙中。2019 年，刘继峰等人发表论文认为 LB-1 系统包含一颗质量约为  $70M_{\odot}$  的恒星级黑洞，但该结论尚未完全确认 [64]。LIGO/Virgo 探测到的 GW190814 事件中包含一个质量在  $2.50 - 2.67M_{\odot}$  的天体，处于低质量间隙中 [10]。另外，GW190521 事件中，包含两个质量分别为  $85^{+21}_{-14}M_{\odot}$  和  $66^{+17}_{-18}M_{\odot}$  的黑洞并合，均处于高质量间隙中 [9]。特别注意 GW190521 并合后生成的黑洞质量大于  $100M_{\odot}$ ，处于传统定义的中等质量黑洞范围内，是人类第一次确切探测到的中等质量黑洞。地面引力波探测器的后续探测认为 GW200220\_061928，GW191109\_010717，GW200208\_222617，及 GW191127\_050227 等事件均有不小的概率包含高质量间隙内的黑洞 [92]。◆

### 3.2.5.2 大质量

主流模型认为，在宇宙形成之后，在小尺度上的密度涨落使冷暗物质凝聚起来，并进一步促使重子物质掉入引力势井形成星系。根据主流的“自上而下”的层级理论，星系之间会发生并合，由小长大。一般定义，质量  $\gtrsim 10^5 M_{\odot}$  的黑洞即为大质量黑洞。观测表明星系的中心普遍存在大质量黑洞，并且其与宿主星系密切相关。因此，星系并合也会触发对应的中心大质量黑洞的并合。对这一过程的引力波观测，将提供一个宝贵的观测窗口，供我们一窥大质量黑洞、星系、甚至宇宙演化的奥妙。

对于大质量黑洞的形成过程，我们的理解并不完美。当前主要的形成模型有两种，即“轻种子模型”与“重种子模型”。轻种子模型中，气体云降落入原初暗物质晕形成的引力势中，这样就形成了第一代大质量恒星，它们的质量可以达到  $100M_{\odot}$  甚至更大，这些初代恒星被称为“星族 III(Population III)” 恒星。大质量的星族 III 恒星在红移约为 20 处可以塌缩成为质量约为  $100M_{\odot}$  的种子黑洞。接下来，这些种子黑洞会通过吸积生长，这个过程中，它会和其宿主星系以及背后的暗物质晕协同演化。这种星系中心大质量黑洞的生长模型被称为“轻种子”模型。

模型	事件率 ( $yr^{-1}$ )	天琴		两组天琴	
		探测率 ( $yr^{-1}$ )	探测占比	探测率 ( $yr^{-1}$ )	探测占比
L – seed	2.57	0.08	3.1%	0.162	6.3%
H – seed	2.57	1.055	41.1%	1.642	63.9%
popIII	174.70	10.58	6.1%	22.60	12.9%
Q3_d	8.18	4.42	54.0%	8.06	98.5%
Q3_nod	122.44	58.96	48.2%	118.12	96.5%

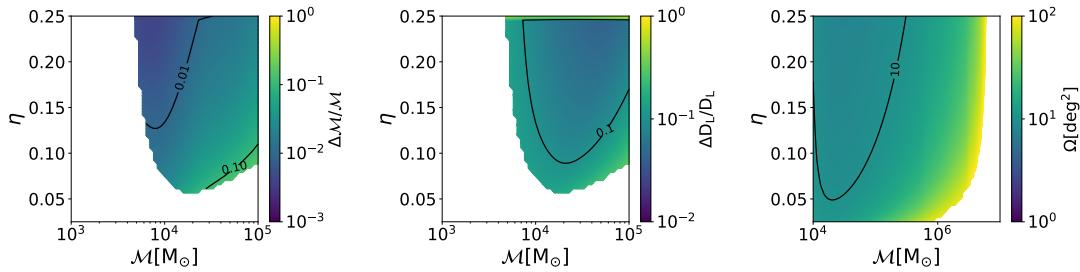
表 3-1 不同天琴构型、不同天文学模型下，天琴对大质量双黑洞并合的探测率汇总。

Table 3-1 Summary of detection rate of TianQin for massive black hole binaries under different configuration and different astronomical models.

然而，近年来在高红移处观测到了许多明亮类星体，意味着在宇宙早期就已存在较高质量的黑洞。这一观测事实为轻种子黑洞模型带来了一定挑战，这是因为如果黑洞诞生时质量较小，是很难在短时间内成长为高质量黑洞的。为了解决这一问题，有研究者提出了重种子模型。在重种子模型下，由于宇宙早期的背景温度较高，强紫外背景阻止了气体云的冷却，有效地阻止了气体云的分块。在引力作用下，质量达到  $10^4 M_{\odot}$  大质量气体云会直接坍塌，在红移约为 10 时形成质量约为  $10^4 \sim 10^6 M_{\odot}$  的重种子黑洞。

基于不同的种子黑洞模型，可以预测引力波探测器的探测率。对此，王海天等人的研究指出（如表3-1所示），在不同构型下，对不同天文学模型预言出的黑洞并合历史，天琴预期的探测率各不相同 [95]。不同模型的具体含义可以参考 [95] 及相关硕士论文。需要注意第一二行对应的星系并合模型基于 Millennium simulation I，由于数值模拟的计算力有限导致质量分辨率有限，因此其预言出的事件率是极为保守的估计。天琴的真实探测率预期将高于此预期。总体来说我们相信天琴运行五年后可以探测到大质量双黑洞并合事件，且利用质量分布、事件率等信息可以反推其种子模型。

值得强调的是，对于高红移（红移为 15）的大质量双黑洞并合，天琴可以以 10% 的精度确定其质量和光度距离等信息。理论上，通过对高红移信号的单事例分析，就可以限制种子模型。对于红移为 2 处的大质量双黑洞并合，天琴则可以在它们并合 24 小时之前以足够高的信噪比对其实现探测（空间定位精度好于  $100 \text{ deg}^2$ ），从而实现早期预警，并触发多信使联合观测。



(a)  $z = 15$  处的并合事件, 天琴对  $M$  的测量不确定度  
 (b)  $z = 15$  处的并合事件, 天琴对  $D_L$  的测量不确定度  
 (c)  $z = 2$  处的并合事件, 天琴的空间定位能力

图 3-6 针对大质量双黑洞并合事件, 天琴对某些特定情形下的参数限制能力。

Figure 3-6 TianQin's ability of constraining physical parameters of binary massive black hole mergers.

### 3.2.5.3 中等质量

通常认为, 质量处于  $10^2 M_\odot \lesssim M \lesssim 10^5 M_\odot$  范围内的黑洞为中等质量黑洞。以往对于中等质量黑洞的观测都只有非常间接的证据, 如通过动力学测量显示中等质量黑洞存在的迹象。不过, 观测表明, 星系质量与星系中心黑洞的质量有很好的相关关系。由此类推, 可以预期在球状星团中心, 或者矮星系中心, 存在中等质量黑洞。超亮 X 射线源 (ultra-luminous X-Ray source, ULX) 可以用中等质量黑洞来解释。

GW190521 并合后生成的黑洞是人类第一次确切探测到的中等质量黑洞。<sup>[9]</sup> 刘帅等人也研究了天琴对类 GW190521 的旋近信号的探测能力。研究表明, 对于轨道偏心率较小的源, 天琴可以探测到十多个信噪比大于 8 的源。即使信噪比阈值增加到 12, 天琴仍然可以检测类 GW190521 的旋近信号。对于几年时间内并合的信号, 其时间精度可以达到 1 秒, 方位精度可以达到 1 平方度。这表明天琴的探测可以为地基引力波探测器和电磁望远镜提供预警, 为多信使天文学、多波段引力波天文学的开展提供有力保障。此外, 天琴对这类信号的轨道偏心率可以用  $10^{-4}$  的相对精度进行限制, 从而有望这些系统的形成通道。最后, 如果类 GW190521 信号的偏心率极高, 则其引力波信号可能无法被天琴直接探测到, 但这样的零探测仍然可以帮助我们更好地理解一些基础问题<sup>[66]</sup>。

## 3.2.6 BBH 与基础物理

### 3.2.6.1 超越 GR

虽然广义相对论到目前为止通过了所有实验检验, 但是, 与黑洞奇点和黑洞信息有关的理论问题以及关于暗物质和暗能量的观察证据都表明, 广义相对论可能不是终极的引力理论。目前, 在广义相对论的基础上, 许多研究者已经提出了许多修改引力理论。

通过对黑洞并合过程中引力波信号进行深入分析，可以实现对引力理论的检验。[25]

### 3.2.6.2 检验 Kerr 本质

引力波提供了一种检验黑洞无毛定理的实验方法，基于此可进一步检验强场下广义相对论的有效性。铃宕阶段辐射的引力波可以通过多个振荡衰减的准正则模式（quasi-normal mode）叠加构成。若广义相对论预言的无毛定理在强场下依然有效，则每一个准正则模的振荡频率和衰减时间都由黑洞的质量和自旋唯一确定。可以通过不同模式所预言的黑洞质量和自旋之间的一致性进行无毛定理的检验。[88]

### 3.2.6.3 与暗物质的联系

到目前为止，我们对暗物质的理解仍然十分有限，仅知道其具有引力作用，并且占据整个宇宙质量的重要部分，其总质量甚至是我们日常研究的普通物质的数倍。引力波的研究方法自然就是十分理想的手段。如，有研究者认为 LIGO 探测到的恒星级质量双黑洞可能源于原初黑洞，并且这类黑洞起码构成了部分的暗物质。同时，不同的暗物质模型也会间接地影响引力波测量，比如，由于暗物质的自引力效应，Einstein 场方程的具体解可能与广义相对论下的解发生偏离。黑洞处在暗物质环境中，可能也会产生吸积和拖曳效应，这种效应通过轨道的长期演化，可能会在引力波波形中体现出来。如果考虑轴子/类轴子粒子作为暗物质的候选对象，当他们出现在 Kerr 黑洞附近时，会通过超辐射 (superradiance) 过程提取黑洞角动量，并生成更多的轴子/类轴子粒子，在黑洞附近形成类似于氢原子核附近电子云般的云状结构，而当粒子湮灭时，会产生连续引力波信号，目前已有工作通过 LIGO/Virgo 数据开展相关搜索。[30]

## 3.3 中子星双星并合

### 3.3.1 中子星

早在上世纪 30 年代，Landau 等理论学家就曾经预言中子星的存在。由于中子星引力场极强，电子间的简并压无法抵抗引力，因此纷纷形成中子（也包括一小部分质子）。因此，可以将中子星形象地比喻为一个巨大的原子核。

在上世纪 60 年代，射电天文学和 X 射线天文学的进展为天文观测带来了飞跃。1967 年，Antony Hewish 为了观测行星际射电源闪烁，建立了一个射电望远镜。而当时还是博士生的 Jocelyn Bell 在处理仪器记录的数据时发现在数据中存在一个周期为 1.337s 的信号，从而第一次发现了脉冲星 PSR B1919+21。其实这种规律的周期信号前人也发现

过，但很多时候是被认为人类活动产生的噪声而被弃之不理的。Jocelyn 的贡献在于她仔细检查了持续时间非常长的信号，并确认了这一周期信号每隔 23 小时 56 分钟出现一次，由此可以确认这一信号是天文学起源。Pacini 和 Gold 首先指出快速旋转的中子星磁场可以解释蟹状星云的能量来源。如果中子星在偏离自转轴的某个  $\theta$  角度方向上辐射，则中子星就好像一个灯塔一般。脉冲星，就是“灯塔”的光会扫过地球观测者的中子星。

在中子星的研究领域， $P$ - $\dot{P}$  图 ( $P - \dot{P}$  图) 的应用非常广泛。然而另一方面，我们会看到一些重要物理量（如下面介绍的制动指数）的定义是基于角速度  $\omega = 2\pi/P$  的，因此我们接下来的讨论中将同时讨论角速度  $\omega$  和周期  $P$ 。将中子星简化为磁偶极矩  $\mu$ ，并以  $\omega$  的角速度旋转，对应的其周期为  $P = 2\pi/\omega$  对应的磁偶极矩辐射会将中子星的自转动能提取，从而导致中子星自转速度减慢，系统动能减慢的速率

$$\begin{aligned} -\frac{dE_{\text{rot}}}{dt} &= -\frac{d}{dt} \frac{1}{2} I \omega^2 \\ &= -I \omega \dot{\omega} \\ &= 4\pi^2 I \frac{\dot{P}}{P^3} \end{aligned} \quad (3.49)$$

考虑中子星典型的转动惯量  $I = 10^{45} \text{ g cm}^2$ ，有

$$-\frac{dE_{\text{rot}}}{dt} \sim 3.95 \times 10^{31} \text{ erg s}^{-1} \left( \frac{I}{10^{45} \text{ g cm}^2} \right) \left( \frac{\dot{P}}{10^{-15}} \right) \left( \frac{1 \text{ s}}{P} \right)^3 \quad (3.50)$$

设  $\dot{\omega} = -\kappa \omega^n$ ，由于磁偶极矩辐射的功率为

$$L_{\text{EM}} = \frac{2}{3c^3} \mu_\perp^2 \omega^4 \quad (3.51)$$

若脉冲星的减速完全由磁偶极矩辐射引起，则可以得到  $\kappa = 2\mu_\perp^2/(3Ic^3)$  且  $n = 3$ 。通常把  $n$  称为制动指数 (braking index)，也可以由  $n = \omega \ddot{\omega} / \dot{\omega}^2$  得到。实际上，不同的脉冲星中，其制动指数差别巨大，在为数不多的可以测量制动指数的脉冲星中，制动指数大多均小于 3。

我们在第3.5.1节将针对中子星的引力波展开详细讨论。结合公式3.74和3.28，可以证明，旋转中子星的引力波光度为

$$L_{\text{GW}} = \frac{1}{5} \frac{G}{c^5} \langle \ddot{I}_{ij} \ddot{I}^{ij} \rangle = \frac{32}{5} \frac{G}{c^5} \epsilon^2 I_{zz}^2 \omega^6 \quad (3.52)$$

可以得到

$$\dot{\omega} = -\frac{32}{5} \frac{GI^{zz}\epsilon^2}{c^5} \omega^5, \quad (3.53)$$

不难发现，对于完全由引力波主导的系统而言，其制动指数  $n = 5$ 。2016 年，Archibald 等人宣布测量了 PSR J1640-4631 的制动指数  $n = 3.15 \pm 0.03$ ，这可能可以由引力波辐射所解释 [20]。

$P$  和  $\dot{P}$  对于中子星研究有着至关重要的作用，比如，可以从量级上估计出中子星年齡

$$\begin{aligned}\tau &\sim -\frac{1}{n-1} \frac{\omega}{\dot{\omega}} \\ &= \frac{1}{n-1} \frac{P}{\dot{P}}\end{aligned}\quad (3.54)$$

在假设中子星减速过程由磁偶极辐射主导 ( $n = 3$ ) 以后，可以得到典型年龄时标  $\tau_c \equiv \frac{P}{2\dot{P}}$ 。另一方面，由于中子星表面磁场强度  $B \propto \mu_\perp$ ，通过公式3.49和3.51，可以得到  $B \propto (\dot{P}P)^{1/2}$ 。

通常会将脉冲星在对数的  $P - \dot{P}$  上展示，如图3-7所示

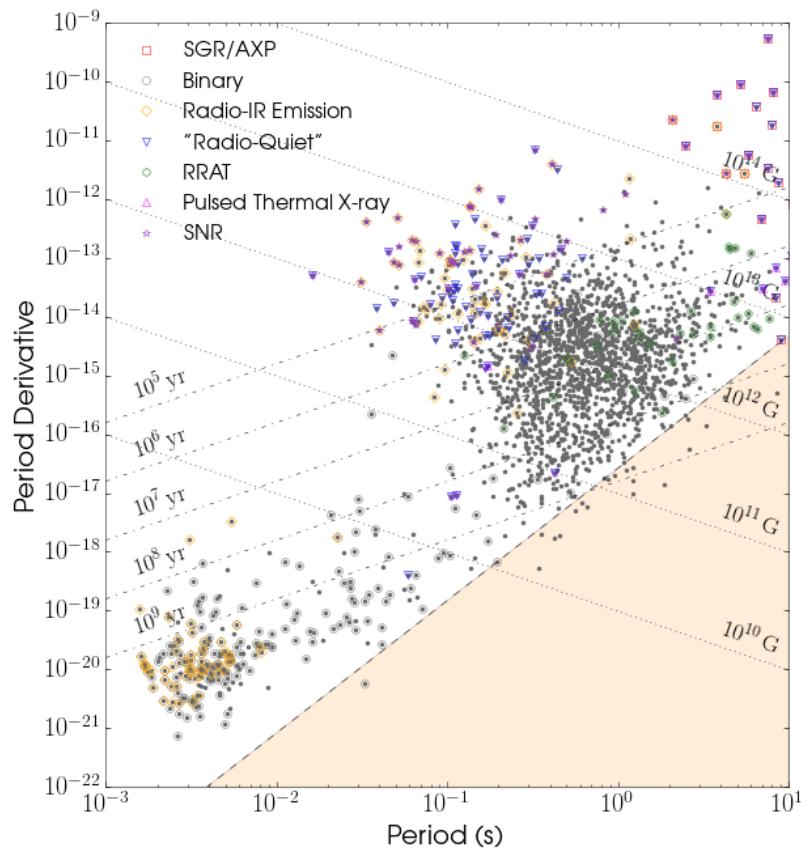


图 3-7 ATNF 脉冲星星表的  $P - \dot{P}$  图。其中  $\dot{P} \propto P^3$  的线代表着等光度线，阴影区域内的脉冲星射电活动关闭（对应于光度小于  $1.1 \times 10^{30} \text{ erg/s}$ ），不再有效地进行辐射，分界线称为“death line”。 $\dot{P} \propto P$  的线代表着等年龄线。 $\dot{P} \propto 1/P$  的线代表着等磁场强度线。图中最右上角的区域磁场强度最强，称为磁星（magnetar）；中部大量的普通脉冲星；左下角的脉冲星磁场强度小，年龄大，通常是老年的毫秒脉冲星。[79]

Figure 3-7  $P - \dot{P}$  plot for pulsars within the catalogue of ATNF. Credit: [79]

### 3.3.2 中子星双星天文学

#### 3.3.2.1 BNS

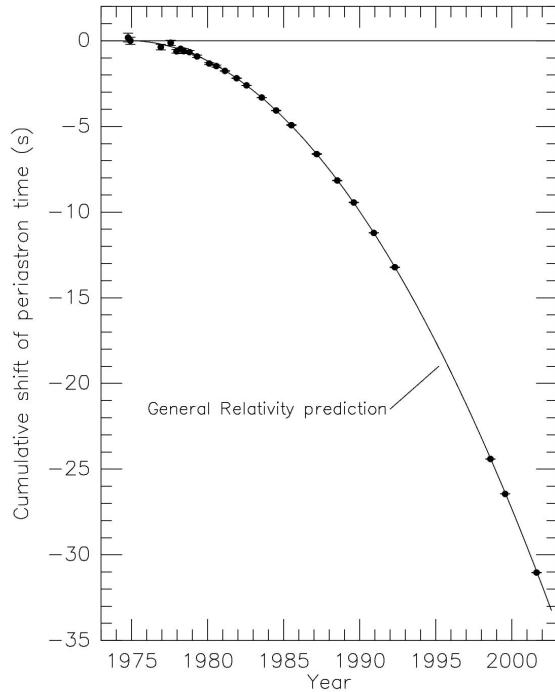


图 3-8 来自双中子星系统 PSR B1913 + 16 的轨道衰减。数据点对应具有不确定性的实际观测值，而实线是广义相对论的理论预测。

Figure 3-8 Orbital decay from the binary pulsar system PSR B1913+16. The points are actual observations with uncertainty, while the line is predicted by GR. Credit: [96]

虽然 LIGO 在 21 世纪才第一次直接观测到引力波，但引力波的间接效应在 1982 年就已经被探测到。在 1974 年，Russell Hulse 和 Joseph Taylor 两人开展了一项脉冲星搜寻工作，并在 1975 年发现了 PSR B1913+16 这一双中子星系统，由于这一双星系统的重要意义，也被称为 Hulse-Taylor binary。1982 年，通过数年间对 Hulse-Taylor binary 系统的观测，Russell Hulse 和 Joel Weisberg 公布了其公转周期演化，并与理论预言进行了比较，特别是确认了轨道周期的缓慢缩减，与广义相对论推导的引力波预期高度一致（相对误差 0.2%）。由于这一发现，Hulse 和 Taylor 两人获得了 1993 年的 Nobel 奖，“for the discovery of a new type of pulsar, a discovery that has opened up new possibilities for the study of gravitation”。之后发现的双中子星系统 PSR J0737-3039 对广义相对论作出了更强的限制，其观测结果与理论预言在 0.05% 的误差范围内一致。

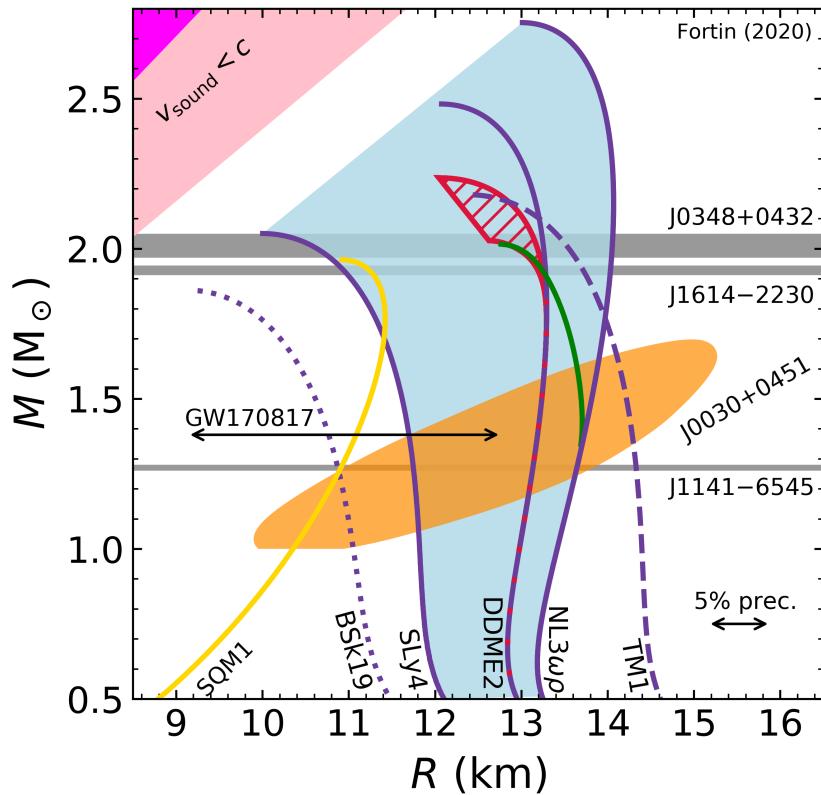


图 3-9 中子星的状态方程及观测给出的限制。

Figure 3-9 Equations of state and constraints from observations. Credit: [44]

在中子星极强引力场、极强压强的物理条件下，核子的状态方程无法在地面实验室确定，当前理论模型很多，存在很大的不确定度，也导致了中子星的质量上限无法精确测量。一些大质量中子星系统的发现，如 PSR J0348+0432，可以排除掉许多中子星物态方程。另一方面，不同的中子星物态方程会决定中子星的潮汐形变能力。中子星具有一定尺寸，因此在双星系统中会受引潮力  $\mathcal{E}_{ij} = \partial_i \partial_j U_{\text{ext}}$  作用而产生形变，导致质量四极矩会产生响应变化  $Q_{ij} = -\lambda \mathcal{E}_{ij}$ 。形变的大小由 Love number  $k_2$  决定

$$k_2 = \frac{3}{2} \frac{G\lambda}{R^5} \quad (3.55)$$

其中， $R$  是中子星半径。对于双中子星系统中的第  $i$  颗中子星而言，可以构造无量纲量  $\Lambda_i = \frac{2}{3} k_2 \left( \frac{R_i}{Gm_i/c^2} \right)^5$ ，更大的  $\Lambda_i$  对应于更松散的中子星（更容易形变）；更小的  $\Lambda_i$  对应于更紧致的中子星（更不易形变）。对于双星系统而言，可以进一步构造出

$$\tilde{\Lambda} = \frac{16}{13} \frac{(m_1 + 12m_2)m_1^4\Lambda_1 + (m_2 + 12m_1)m_2^4\Lambda_2}{(m_1 + m_2)^5} \quad (3.56)$$

中子星形变会对形如  $\tilde{h}(f) = \mathcal{A}(f) \exp(i\Psi(f))$  的引力波带来 5PN 阶的波形修正量

$$\Delta\Psi_{\text{5PN}}^{\text{tidal}} = -\frac{117}{256} \frac{M^2}{m_1 m_2} \tilde{\Lambda} \left(\frac{v}{c}\right)^5 \quad (3.57)$$

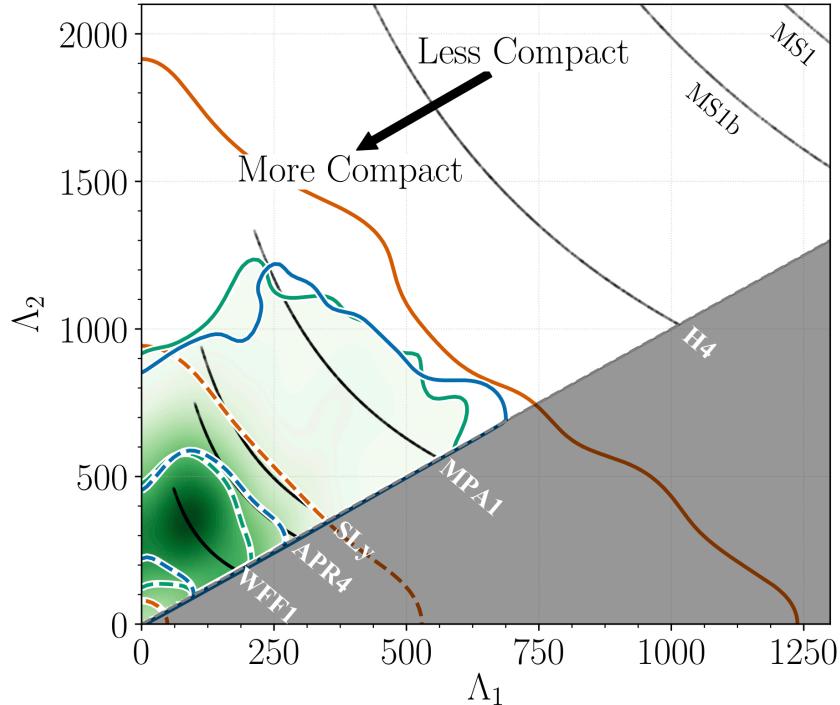


图 3-10 GW170817 显示的双中子星的潮汐变形性  $\Lambda$ ，同时画出了几种可能的状态方程。

Figure 3-10 The tidal deformability of the two neutron stars from GW170817. Credit: [5]

2017 年 8 月 17 日，LIGO 和 Virgo 观测到了一起双中子星并合事件，GW170817，同时电磁波段望远镜也观测到了这一事件的电磁对应体。[3, 4, 2, 1]。通过对引力波信号的数据分析，可以限制两个中子星的潮汐形变，如图 3-10 所示。图中的绿色阴影区域为 GW170817 给出的限制。黑线表示某些状态模型方程的预测。虚线（实线）围住了潮汐变形能力真实值所在的区域，概率为 50% (90%)。绿线和蓝线是指此分析中使用的两种方法，而橙线是指对源性质的假设较少的分析。

这一多信使观测确定了短伽玛暴的双中子星并合起源，并确认了双中子星并合导致的  $r-$  过程元素增丰是宇宙中重元素产生的重要渠道。这一双中子星并合发生于 S0 星系 NGC 4993 中，由此估算出的局部双中子星并合率为  $1540^{+3200}_{-1220} \text{ Gpc}^{-3} \text{ yr}^{-1}$ 。同样，可以通过双星演化模型解释观测到的双中子星系统。但目前的理论预言给出的中子星并合事件率低于探测值。在乐观估计下，考虑不同种类的星系均能产生双中子星并合，则可以调和两者的矛盾。但需要注意的是，由短伽马暴和  $r-$  过程重元素增丰预言的双种子

Pulsar Name	$M_T (M_\odot)$	$m_r (M_\odot)$	$m_s (M_\odot)$	$\mathcal{M}_c (M_\odot)$	$q$	$P_b$ (day)	$T_c$ (Gyr)
Systems will merge within a Hubble time							
J1946+2052	2.50(4)	< 1.35	> 1.17	(1.05, 1.11)	(0.68, 1)	0.078	0.046
J1756-2251	2.56999(6)	1.341(7)	1.230(7)	1.1178(3)	0.92(1)	0.320	1.656
J0737-3039A/B	2.58708(16)	1.3381(7)	1.2489(7)	1.1253(1)	0.933(1)	0.102	0.086
J1906+0746	2.6134(3)	1.322(11)	1.291(11)	1.1372(2)	(0.956, 1)	0.166	0.308
B1534+12	2.678463(4)	1.3330(2)	1.3455(2)	1.165870(2)	0.9907(3)	0.421	2.734
B2127+11C	2.71279(13)	1.358(10)	1.354(10)	1.18043(8)	(0.975, 1)	0.335	0.217
J1757-1854	2.73295(9)	1.3384(9)	1.3946(9)	1.18930(4)	0.960(1)	0.184	0.076
J0509+3801	2.805(3)	1.34(8)	1.46(8)	1.215(5)	(0.793, 1)	0.380	0.574
B1913+16	2.828378(7)	1.4398(2)	1.3886(2)	1.230891(5)	0.9644(3)	0.323	0.301
J1913+1102	2.875(14)	1.64(4)	1.24(4)	1.239(8)	0.76(4)	0.206	0.475
Systems will not merge within a Hubble time							
J1807-2500B	2.57190(73)	1.3655(21)	1.2064(21)	1.1169(3)	0.883(3)	9.957	1044
J1518+4904	2.7183(7)	1.41(8)	1.31(8)	1.181(5)	(0.794, 1)	8.634	8832
J0453+1559	2.733(4)	1.559(5)	1.174(4)	1.175(2)	0.753(5)	4.072	1453
J1411+2551	2.538(22)	< 1.64	> 0.92	(1.05, 1.11)	(0.57, 0.95)	2.616	466
J1811-1736	2.57(10)	< 1.75	> 0.91	(1.02, 1.17)	(0.58, 0.95)	18.78	1794
J1829+2456	2.59(2)	< 1.36	> 1.25	(1.08, 1.14)	(0.65, 1)	1.176	55
J1930-1852	2.59(4)	< 1.32	> 1.30	(1.07, 1.15)	(0.58, 0.96)	45.06	$\sim 10^5$

表 3-2 表格引用自 [41]，列出了银河系内双中子星系统的质量信息，包括了双星总质量 ( $M_T$ )，再生中子星质量 ( $m_r$ ) 和满转中子星 ( $m_s$ )，双星啁啾质量 ( $\mathcal{M}_c$ )，质量比 ( $q$ )，双星轨道周期 ( $P_b$ ) 和并合时标  $T_c$ 。括号中的数字表示了最后一位数字的  $1-\sigma$  不确定度。对五个没有子星质量测量的双星系统，列出了  $m_r$  和  $m_s$  的 99% 概率上下限。

星并合率同样低于由 LIGO/Virgo 给出的并合率估计。当然，随着观测时间的累积，未来有望对这一事件率给出更精确的估计。

LIGO/Virgo 后续又探测到 GW190425，但并未观测到电磁对应体 [7]。基于其伴星质量，极有可能是双中子星并合，但无法完全排除其较重天体为黑洞的可能性。根据这一事件，双中子星的并合率更新为  $1090^{+1720}_{-800} \text{ Gpc}^{-3} \text{ yr}^{-1}$ 。经过 O3b 更长时间的观测后，虽然积累数据更多，但是由于 LIGO 在折算双中子星并合率时将更多模型纳入考虑中，因此其并合率的置信区间反而进一步拉大（跨越更大的量级范围），目前的估计值为  $10 - 1700 \text{ Gpc}^{-3} \text{ yr}^{-1}$ 。

除了引力波测量以外，通过对脉冲星的电磁观测限定中子星性质，因此可以编撰出银河系中的双中子星源表，如表3-2所示。虽然其一般离并合时间极远，无法在较

近的未来被探测到，但是这些中子星双星的质量、并合时标等信息可以帮助我们更好地理解这类恒星级致密双星的性质。有意思的是，GW190425 的双星总质量与电磁波段发现的双中子星总质量有明显偏离，可能预示着不同的起源。

### 3.3.2.2 中子星-黑洞双星

关于中子星-黑洞双星系统的观测，不确定度极大，截止目前，并没有公认的电磁观测结果。但一些研究认为在中子星-黑洞双星并合时能产生电磁信号，从而被记录到，如 [61] 认为部分伴有红外超的伽玛暴，其物理起源可以用中子星-黑洞双星并合解释，并且可以用这些电磁观测限制中子星物态方程。

经过一年多的仪器升级和灵敏度提升，LIGO/Virgo 开始了第三次观测运行（O3）。O3 分为两部分，2019 年 4 月 1 日-10 月 1 日为 O3a，2019 年 11 月 1 日-2020 年 3 月 27 日为 O3b，中间为期一个月的修整进一步提升了灵敏度。2020 年 10 月 29 日，O3 的第一部分观测结果以引力波暂现源表 2 (GWTC-2) 的形式公布 [68]。2021 年 11 月 5 日，LIGO-Virgo 科学合作组发表了引力波暂现源表 3 (GWTC-3) [92]，并更新了族群性质的估计 [12]。这其中包括了 GW191219\_163120 ( $1.2 - 31 M_{\odot}$ )，GW200115\_042309 ( $1.4 - 6 M_{\odot}$ )，GW200210\_092254 ( $2.8 - 24 M_{\odot}$ )，上述事件未观测到电磁对应体。其中最后一个事件的质量较为特殊，有可能是双黑洞并合。另外，注意到，曾经有一例事件 GW200105\_162426 被单独发表、报道 [11]，但后续分析认为其物理性质与族群性质符合度不高，因此其来自天文学起源的可能性被降低，最终被移出源表。LIGO-Virgo 给出的中子星-黑洞并合率为  $7.8 - 140 \text{ Gpc}^{-3} \text{ yr}^{-1}$ 。

### 3.3.2.3 引力波宇宙学

如果将广义相对论应用于宇宙学研究，就可以得到宇宙不同时期宇宙的演化历史。Hubble 第一次用观测证明，几乎所有的星系都在离我们远去，并且离我们越远的星系（即光度距离  $D_L$  越大），远去的速度越快（即红移  $z$  越大），换句话说，宇宙处于膨胀状态。这就是著名的 Hubble's Law

$$v = H_0 D_L \quad (3.58)$$

这其中  $H_0$  是 Hubble 常数。注意其中  $v$  是星系退行速度，和红移  $z$  的关系可以由下式给出：

$$1 + z = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (3.59)$$

1998 年，Supernova Cosmology Project 和 High-Z Supernova Search Team 分别搜寻并测量了 Ia 型超新星的光度距离和红移，发现宇宙的膨胀在加速，由此，两个团队的领

导者 Saul Perlmutter, Brian P. Schmidt 和 Adam G. Riess 分获了 2011 年的 Nobel 奖, “for the discovery of the accelerating expansion of the Universe through observations of distant supernovae”。

宇宙学研究, 可以通过分别测量天体的光度距离和红移来研究。传统天文学研究中, 光度距离通常无法直接测量, 需要通过逐级定标, 包含了较大的系统误差。目前, 两种不同的方法精确测得的 Hubble 常数有着明显的不一致。而引力波测量过程中, 可以直接得到光度距离  $D_L$  的测量, 无需层层定标, 因此可以提供独立的宇宙学参数测量, 具有重大意义。

如图 3-11 通过 LIGO/Virgo 对 GW170817 的电磁波观测, 结合对其对应体及宿主星系的红移观测, 可以对 Hubble 常数进行限定 [1]。虽然精度上比不上电磁波手段获得的宇宙学参数限制能力, 但这表明引力波宇宙学的方法切实可行, 随着观测数量的增加将有望得到更精确的测量结果。

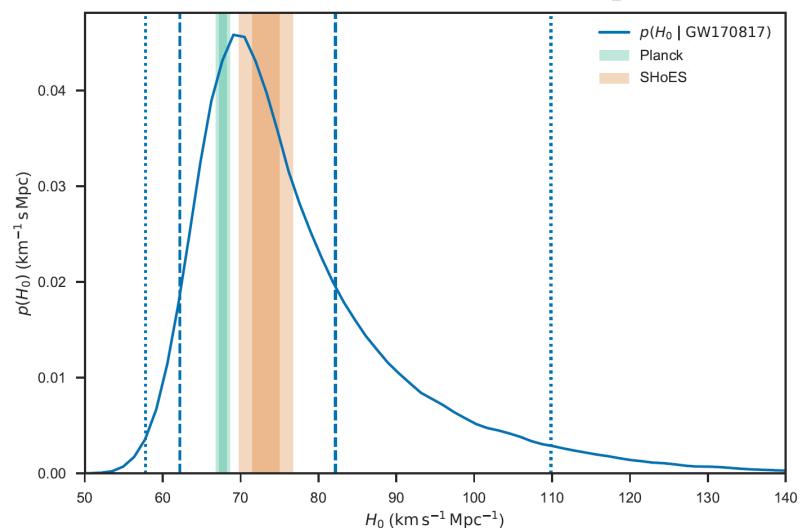


图 3-11 不同测量方法下得到的 Hubble 常数概率分布。

Figure 3-11 Inference on the Hubble constant using three methods. Credit: [1]

除了利用多信使手段同时测量光度距离和红移以研究宇宙学, 还可以利用星系在宇宙中分布不均匀的特性, 无需确定宿主星系即可得到 Hubble 常数的估计值。这样的方法已经运用于双黑洞并合事件 GW170814, 并得到了对 Hubble 常数的限制结果。虽然精度上更比不上 GW170817 对宇宙学参数的测量, 但是这一方法独立于电磁对应体的发现, 具有更好的普适性和独立性 [90]。

针对天琴等空间引力波探测器, 朱良贵等人利用恒星级双黑洞和大质量双黑洞, 开展了引力波宇宙学能力评估研究。研究表明, 利用星系中心大质量黑洞的质量和核球光

度的线性相关关系对星系进行加权，可以有效提升引力波宇宙学的参数评估能力。天琴与 LISA 的联合可以将哈勃参数定到 4%-7%，在有电磁对应体的前提下能准确至 1.7% [97]。针对恒星级质量双黑洞引力波源，利用多个巡天波段测光光度信息给候选宿主星系加权，也可以获得类似的宇宙学参数评估能力提升。结合天琴与 ET 的多波段引力波探测可以用暗汽笛方法将哈勃参数定到 1% [98]。

## 3.4 EMRI

### 3.4.1 天文学背景

在星系中心的核区，除了中心的大质量黑洞之外，还有大量的小质量天体，聚集在大约秒差距的尺度内。一般预期，核区的质量密度高达  $\sim 10^6 pc^{-3}$ ，并以  $100 - 1,000 km/s$  的高速运动着。实际上，大质量黑洞与其宿主星系之间的相互作用，很大程度上是藉由中心核区实现的。在核区的复杂动力学环境下，也会上演着精彩的天文大戏：如果恒星靠近大质量黑洞太过接近，大质量黑洞的引潮力可能会将恒星变形、撕裂，从而形成潮汐瓦解事件，tidal disruption event (TDE)。

**例 3.5** 假设一颗类太阳的恒星发生 TDE，考虑其多方指数为 3，则其潮汐瓦解半径为

$$r_t \approx 1.4 \times 10^{11} \left( \frac{M_\bullet}{M_\odot} \right)^{1/3} \text{ cm} \quad (3.60)$$

如果接近大质量黑洞的天体密度增加，那么其瓦解半径将会减小。更进一步地，小天体的碰撞参数也足够大，能避免小天体直接掉落进大质量黑洞的视界中，那么这就构成了一种极为有趣的研究对象：这种一大一小、质量比及其悬殊的双星系统，将成为一种极佳的引力波源，我们通常称之为极端质量比旋近，extreme mass-ratio inspiral (EMRI)。我们对于如白矮星、中子星或恒星级质量黑洞作为小天体的EMRI系统特别感兴趣，是因为小天体可以非常靠近大质量黑洞而不被潮汐瓦解。



**注 3.6.** 在相关领域的研究中，通常定义 EMRI 为质量比小于  $10^{-4}$  的系统，并将质量比处于  $10^{-4} - 10^{-2}$  范围内的系统称为中等质量比旋近，intermediate mass-ratio inspiral (IMRI)。在这里，我们不特别区分 EMRI 和 IMRI。

在EMRI系统中，其动力学演化末期阶段将由引力波主导，随着引力波辐射能量，小天体绕大质量黑洞的轨道半长径不断缩减，总体上有频率增强的趋势。

通过研究小天体围绕大质量黑洞绕转的过程，我们可以得到大质量黑洞周围时空的信息，也由此可以获得大质量黑洞的相关物理参数。而这一信息可以通过引力波传播，正因此，对EMRI系统的引力波探测将实现对广义相对论的高精度检验。与此同时，对EMRI的研究也可以用来研究星系核区的动力学、大质量黑洞成长过程等等。

由于大质量黑洞尺度极小，除非对TDE之类的灾变现象进行观测，或者，花费极大观测资源进行观测，否则对于大质量黑洞的各类观测技术只能间接地展示中心黑洞及其周围环境。因此，到目前为止，我们对于EMRI系统的理解也相当间接，在引力波探测器运作之前，无法通过直接测量的方式研究。因此，当前对EMRI的研究目前主要集中于理论模拟。而这类研究的重点集中于小质量天体如何在各种机制的作用下，靠近大质量黑洞，并且由引力波辐射机制接管之后的轨道演化的。研究的手段主要是利用多体动力学模拟或经验公式给出。

在将小天体带入大质量黑洞附近的过程中，主要有三种作用可能起作用：小天体的两体弛豫、碰撞和久期演化。

大质量黑洞周围的引力势，除了光滑的引力背景，也存在小尺度的起伏。而实际上，当两个小天体极为靠近时，互相之间的引力作用将可能导致能量和角动量的互换，导致小天体之间的引力作用将盖过大质量黑洞和背景引力势的影响。

通常，在天文学领域讨论的“碰撞”并非物理意义上的撞击，而是指引力相互作用导致作用结束后新的轨道和旧的轨道之间不存在相关性。但由于在大质量黑洞附近恒星的密度极端地高  $\sim 10^* M_\odot / pc^3$ ，因此真实发生的两星碰撞过程也不可避免，但这一过程对星系中心核区星团的影响可以忽略不计。

最后，小天体轨道的久期演化不会改变能量，但是可以在不同天体之间实现角动量的交换。

通常定义，当通过作用后，小天体的径向速度分量的改变  $\delta v_\perp$  与原径向速度  $v_\perp$  处于同一量级时，认为系统已经充分弛豫。早在 1942 年，Chandrasekhar 就已研究过两体弛豫过程的典型时标。可以估算出，星系中心核区的两体弛豫时标为

$$t_{rlx}(r) \approx 1.8 \times 10^8 \text{ yr} \left( \frac{\sigma(r)}{100 \text{ km s}^{-1}} \right)^3 \left( \frac{10 M_\odot}{m_{CO}} \right) \left( \frac{10^6 M_\odot \text{ pc}^{-3}}{\langle m \rangle n(r)} \right) \quad (3.61)$$

这其中  $\sigma(r)$  是局部的速度弥散； $n(r)$  是局部恒星数密度； $\langle m \rangle$  是平均恒星质量；而  $m_{CO}$  是小天体的质量，通常主要考虑约  $10 M_\odot$  的恒星级质量黑洞。另一个常用于计算的是动力学时标  $t_{dyn}$ ，通常有  $t_{dyn} \ll t_{rlx}$ 。

可以发现，在  $R_{infl}$  内，当  $r$  减小时，弛豫时标  $t_{rlx}$  也会拉长。通过两体弛豫作用，不同小天体之间的能量会大致均分，相同的动能对于较重的天体而言意味着较小的速度；正因此，平均而言较重的天体会处于离大质量黑洞更近，引力势更深的区域。这一过程

通常被称为质量分隔 (mass segregation) 或动力学摩擦 (dynamical friction)，最终导致在大质量黑洞附近的恒星数密度按照幂律分布，并形成所谓的数密度尖峰 (number density cusp)

$$n(r) \propto r^{-\gamma} \quad (3.62)$$

其中，幂指数取  $\gamma \approx 1.75 - 2.1$ 。对类似银河系的星系而言，其中心的大质量黑洞附近的弛豫时标是短于 Hubble 时标的。值得一提的是，由于动力学摩擦让大质量天体沉入引力势中，因此在大质量黑洞周围恒星级质量黑洞的数密度极高，远高于星系的其他区域。

对于小天体而言，要形成我们感兴趣的EMRI系统，就必须尽可能靠近大质量黑洞。通常，可以用损失锥 (loss cone) 来描述这一条件，即小天体的速度矢量必须处于一个给定的范围内，才能足够近地靠近大质量黑洞  $r < r_{lc}$ 。可以借此将符合条件的临界角度定义为损失锥角  $\theta_{lc}$ ，如图3-12所示

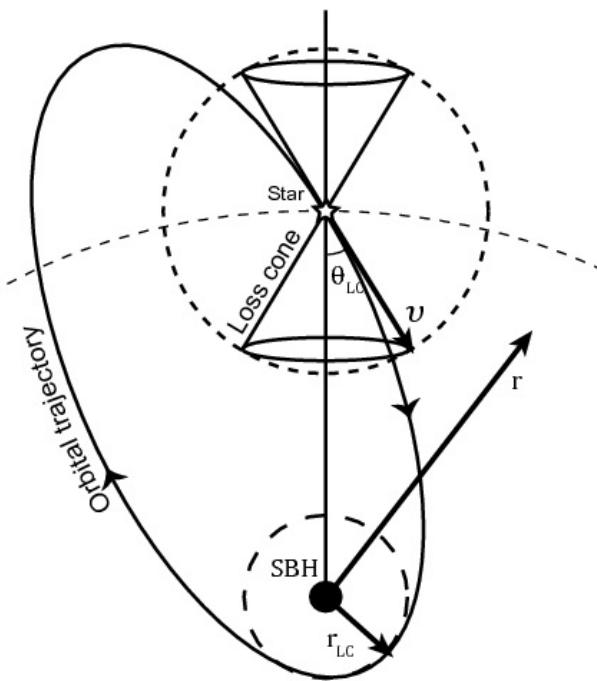


图 3-12 损失锥角  $\theta_{lc}$  的示意图

Figure 3-12 Illustration of loss cone angle  $\theta_{lc}$  Credit:[72]

通过上述描述的星系中心大质量黑洞及附近恒星级小天体动力学过程的理解，可以通过数值的多体动力学模拟、或半解析的蒙特卡洛模拟以开展分析。基于上述研究，可以对大质量黑洞俘获致密小天体的事件率进行估算，进而计算EMRI事件率。

如前所述，EMRI的事件率依赖于多种不同因素，包括：不同红移处的大质量黑洞质量、红移分布；星系中心包含大质量黑洞的比例；不同质量、不同小天体性质的系统对应的EMRI事件率。研究过程中，可以自然产生模拟的引力波源表。

对于大质量黑洞的族群性质，有多种可能模型。如 Barausse12 模型，在大质量黑洞由轻种子黑洞成长的模型下 [54]，大质量黑洞质量函数为

$$\frac{dn}{d \log M} = 0.005 \left( \frac{M}{3 \times 10^6 M_\odot} \right)^{-0.3} \text{Mpc}^{-3} \quad (3.63)$$

或者 Gair10 模型，以保守起见，选择唯象的质量函数估计 [45]，其中大质量黑洞的数量更偏向相对大质量端

$$\frac{dn}{d \log M} = 0.002 \left( \frac{M}{3 \times 10^6 M_\odot} \right)^{0.3} \text{Mpc}^{-3} \quad (3.64)$$

对于大质量黑洞的自旋，也有多种可能性。如轻种子黑洞模型下的 a98 模型，大质量黑洞通过不断吸积，其自旋也接近极端旋转黑洞的极限，上限为  $a = 0.998$ ，而中位数为  $a = 0.98$ [87]。对应比较的是自旋均匀分布的 aflat 模型和无自旋的 a0 模型。

对于黑洞周围的恒星数密度尖峰 (cusp)，在星系并合的情形下，由于来自两个星系中心的大质量黑洞会通过动力学摩擦往引力势中心下沉，由于能量守恒，等效于“加热”周围的小天体，特别是耗尽处于损失锥中的天体。这会导致恒星数密度尖峰的损耗，而在系统稳定后又会由于两体弛豫等作用，使得数密度尖峰再次增长和损失锥的再填充。

通常，将再次形成数密度尖峰所需要的时长定义为  $t_{\text{cusp}}$ ，其具体的取值与  $M-\sigma$  关系的标定有很大联系。在根据 [49] 所定义的 Gultekin09 模型下， $t_{\text{cusp}} \approx 6 \text{Gyr}$ ；最悲观的 KormendyHo13 模型 [55] 和最乐观的 GrahamScott13 模型 [46] 则分别给出  $t_{\text{cusp}} \approx 10 \text{Gyr}$  和  $t_{\text{cusp}} \approx 2 \text{Gyr}$  的估计。

对于单个大质量黑洞周围，不同小天体性质所对应的EMRI事件率。根据 [18] 考虑强质量分层效应的多体数值动力学模拟，可以得到EMRI 事件率依赖大质量黑洞质量的函数

$$R_0 = 300 \left( \frac{M}{10^6 M_\odot} \right)^{-0.19} \text{Gyr}^{-1} \quad (3.65)$$

然而，相关研究并没有考虑小天体直接撞向大质量黑洞的过程 (plunge)，这一效应会导致小天体无法产生足够长时间的旋近过程，进而无法被引力波探测器有效地探测到，需要在计算EMRI事件率的过程中略去，并且大质量黑洞的质量增长需要考虑直接碰撞的效应，具体影响由参数  $N_p$  决定

$$\dot{M} = 3000(1 + N_p) \left( \frac{m}{10 M_\odot} \right) \left( \frac{M}{10^6 M_\odot} \right)^{-0.19} M_\odot \text{Gyr}^{-1} \quad (3.66)$$

Model	Mass	MBH	Cusp	$M-\sigma$	$N_p$	CO	EMRI rate [ $\text{yr}^{-1}$ ]
	function	spin	erosion	relation		mass [ $M_\odot$ ]	Total
M1	Barausse12	a98	yes	Gultekin09	10	10	1600
M2	Barausse12	a98	yes	KormendyHo13	10	10	1400
M3	Barausse12	a98	yes	GrahamScott13	10	10	2770
M4	Barausse12	a98	yes	Gultekin09	10	30	520 (620)
M5	Gair10	a98	no	Gultekin09	10	10	140
M6	Barausse12	a98	no	Gultekin09	10	10	2080
M7	Barausse12	a98	yes	Gultekin09	0	10	15800
M8	Barausse12	a98	yes	Gultekin09	100	10	180
M9	Barausse12	aflat	yes	Gultekin09	10	10	1530
M10	Barausse12	a0	yes	Gultekin09	10	10	1520
M11	Gair10	a0	no	Gultekin09	100	10	13
M12	Barausse12	a98	no	Gultekin09	0	10	20000

表 3-3 根据大质量黑洞的质量分布、自旋分布、数密度尖峰是否存在消耗、 $M-\sigma$  关系、直接碰撞和EMRI的比例、以及小天体的典型质量等模型的不同，预言出较大差异的EMRI事件率。注意到，在最乐观和最悲观的模型下，事件率有着三个量级的差距。注意所有的模型都将事件率考虑的红移上限设定为  $z = 4.5$ ，但对 M4 模型也同时考虑了红移上限设定为  $z = 6.5$  的情形，列入括号中。

Table 3-3 EMRI event rate under various of models. Credit: [24]

这其中  $m$  是小质量天体的典型质量。

在表3-3中，罗列了考虑不同模型组合下，所预言的EMRI 事件率。可以发现，不同的模型下，EMRI的事件率有 3 个量级的不确定性。

### 3.4.2 波形

通常，可以将EMRI波形的计算按照方法分为：微扰法、Teukolsky 法、kludge 法、半相对论性法、后牛顿, post-Newtonian (PN)法和 EOB 法。

对EMRI波形的计算，最为简单的是通过黑洞微扰的方式进行展开。即，将时空度规  $g_{\mu\nu}$  分解为两个不同的部分。具体来说，包含了稳态的，由大质量黑洞提供的背景时空度规，以及运动的，由小天体引起的扰动成分。通过对质量比进行展开，原则上就可以得到微扰法下EMRI系统的波形。由于EMRI讨论的系统中，大小黑洞之比远小于 1，因此上述的分解是成立的。

虽然这一思路较为简单，但在实际操作中非常困难。特别是，EMRI的系统演化实

际上还依赖于小天体的自引力 (self-force)，否则的话，整个系统不会辐射引力波，由于能量守恒，小天体将无法旋近至大质量黑洞的视界中。

另外一种解析方式获得EMRI波形的方式是 Teukolsky 法。这种方法首先选定一组轨道根数以作为初始条件，从而得到一根测地线，作为小天体轨迹的基准。在此基础上，利用上述测地线，求解线性化的 Einstein 场方程，从而获得轨道根数的演化。接下来，用新的轨道根数确定新的测地线方程，并重复上述过程。通过求解非齐次的 Teukolsky 方程，可以获知微扰下的 Weyl 张量  $\Psi_4$ ，并可以基于此计算出系统相应的引力波波形。通过这种方法，既可以计算频率域的EMRI波形，也可以计算时间域的EMRI波形。Teukolsky 法得到的波形虽然相对精确，但是由于需要对 Teukolsky 方程进行数值积分，并且对大量高阶模进行求和，其计算时耗时长久，也由此因此难以实际运用于引力波数据处理中。

值得注意的是，通过 Teukolsky 法推算，EMRI 系统有时会经历一种所谓“zoom-whirl”的效应。小天体通过椭圆轨道进入大质量黑洞附近，接着可能会在大质量黑洞附近沿近圆轨道运动，接着偏心率又增加，小天体又一次远离大质量黑洞。这种 zoom-whirl 效应随大质量黑洞自旋的增强而增强。

另一类方法被归类为“kludge”法。*kludge* 在英语中的原意是七拼八凑的组装品，意指强行拼凑、勉强够用的产品。顾名思义，*kludge* 方法并不追求求解EMRI波形的物理严谨性，而仅为了追求实用。而利用 *kludge* 方法得到波形的过程，正是综合从不同角度出发得到的对轨道演化的描述，拼凑而成的EMRI系统波形。由此，*kludge* 波形自身在物理上并非是自洽的，它的主要目的是为了在进行科学目标分析和初步数据处理能力研究的过程中，可以实现对EMRI特性的体现。

*Kludge* 方法首次由 Barack 等人提出 [26]。在所谓的analytical *kludge* (AK)波形中，EMRI系统每时每刻都遵循 Newtonian 轨道，并且按照 Peters & Matthews [77] 的公式，考虑椭圆轨道对应的质量四极矩的加速变化而产生引力波。在此基础上，以PN方程对轨道根数进行久期演化。特别是，AK波形中还考虑了辐射反作用、近星点进动和轨道面 Lense-Thirring 进动。这两种进动的产生主要是因为径向、轴向和方位角方向的频率不一致而导致。通过这样的处理，AK波形可以考虑除了小天体自旋以外的所有物理参数，在参数空间上和EMRI真实参数空间一致，包含了 14 个参数  $[t_0, \ln \mu, \ln M, S/M^2, e_0, \tilde{\gamma}_0, \Phi_0, \alpha_0, \mu_s \equiv \cos \theta_S, \phi_S, \cos \lambda, \mu_K \equiv \cos \theta_K, \phi_K, \ln(\mu/D)]$ 。这其中， $t_0$  是参考时刻； $\mu$  是小天体质量； $M$  是大质量黑洞质量； $S$  是大质量黑洞的自旋强度 (有  $0 \leq S/M^2 \leq 1$ )； $e_0$  是轨道偏心率， $\tilde{\gamma}_0$  描述了轨道近星点在轨道面上的方位， $\Phi_0$  则是平近点角， $\alpha_0$  是小天体轨道角动量  $\mathbf{L}$  相对大质量黑洞自旋  $\mathbf{S}$  方向的角度。上述几个参数都是随时间演化的，因此在选定初始条件时都是选择  $t = t_0$  时刻的具体取值； $\theta_S$  和  $\phi_S$  描述了EMRI系统相对观测者的方位角； $\lambda$  是  $\mathbf{L}$  与  $\mathbf{S}$  的夹角； $\theta_K$  和  $\phi_K$  描述了  $\mathbf{S}$  的方向； $D$  是源距离。

需要注意的是，虽然采用AK方法得到的EMRI波形可以定性地描述EMRI系统，但是其波形的准确性并不可靠，甚至会在小时量级上与真实波形发生偏离。此外，一些特殊效应（如 zoom-whirl 效应）无法在AK方法中描述。同时，AK波形中，轨道角动量  $\mathbf{L}$  与大质量黑洞自旋角动量  $\mathbf{S}$  的夹角  $\lambda$  保持常数，但是真实情况下，其值会发生一定变化。

在AK的基础上，Babak 等人扩展了numerical kludge (NK)波形 [23]。NK的逻辑是，将小天体轨道的精确解与引力波辐射的近似解相结合，由此，虽然NK依然对真实引力波波形的近似，但依然能够较为准确地表达EMRI系统的动力学性质。首先，在给定轨道能量  $E$ 、轴向角动量  $L_z$  和 Carter 常数  $Q$  等初始条件下，获得相空间上的旋近轨道；其次，在旋近轨道上，对 Kerr 测地线进行数值积分，由此得到小天体的 Boyer-Lindquist 坐标随时间的演化；最后就是从上述轨道中提取引力波波形。与AK方法类似，NK对轨道根数也只计算轨道平均后的久期演化，从而简化计算；并且NK同样从多极矩中提取引力波信号，只不过除了最低阶的四极矩外，NK还会考虑次领头阶。

在 NK 的处理中，会导致所谓“beads-on-a-wire”问题。为了简化计算，在弯曲时空背景下获得小天体的测地线方程后，具体轨迹被置于平直时空中用于计算引力波信号。然而，这一轨道与物理条件不一致的情形，会使得整个系统的能量、角动量不守恒——小天体并不是按照背景时空而运动，而像是穿在线上的珠子一般受迫运动。这种物理图景上的不一致是采用 kludge 方法下可以直接预期的，也会带来代价：NK的波形与真实波形间，可能会显著地影响参数估计时的结果，这一点在开展物理参数测量时需要特别留意。

NK虽然相对精确的 Teukolsky 法而言速度更快，但相对AK波形而言仍然较为耗时。在AK 波形中，轨道频率  $f_{\text{orb}}$  由 Kepler 第三定律决定；而两个进动的频率：近星点进动频率  $f_{\text{peri}}$  和 Lense-Thirring 进动频率  $f_{\text{LT}}$  均由PN展开引入的修正获得。然而，上述三种频率与真实的频率  $(\omega_r, \omega_\phi - \omega_r, \omega_\phi - \omega_\theta)$  并不对应。这导致了AK波形很快就与真实物理系统失去了对应，甚至在旋近早期阶段就可能与NK波形产生分歧。

近年来，蔡靖康与 Gair 等人发展了一套augmented analytical kludge (AAK)波形的计算方法，在与NK相近精度计算EMRI的前提下，与AK耗时接近，同时结合了两种方法的所长 [35]。在AAK波形的计算过程中，通过对数值的波形进行分析，可以将EMRI系统中的物理参数，如大质量黑洞质量  $M$ ，自旋  $a$  和轨道半通径  $p$  并映射到一组拟合得出的、非物理的新参数  $\tilde{M}, \tilde{a}$  和  $\tilde{p}$ 。虽然新的参数并不对应物理，但通过校准，反而可以准确的反应出不同成分的频率  $f_{\text{orb}}, f_{\text{peri}}$  和  $f_{\text{LT}}$ 。这样，在采用AK方法计算波形的快速波形生成基础上，可以得到更为真实的波形，就结果而言与NK有着极高的一致性。

半相对论性法由 Ruffini 和 Sasaki 于 1981 年首先提出 [83]。简单来说，这一方法的

思路就是：在弯曲时空下计算小天体的运动，以及引力波的产生；又在平直时空下计算引力波的传播。这一方法的主要问题是，整个系统中不会考虑辐射反作用，能量并不守恒，导致小天体永远不会往大质量黑洞旋近。可以将半相对论性法视作不考虑辐射反作用的NK法。半相对论性法可以被视作质量比趋于零时的绝热近似的极限情形。基于半相对论性法开展的研究主要是为了理解在 Kerr 时空背景下的轨道特性，其研究出发点更多偏向于理论分析；由于其引力波波形仅在相当段的时长内正确，其对数据处理帮助甚微。

对许多系统而言，在 Newtonian 力学基础上展开的PN波形可以很好地描述双星并合之前的引力波信号。对双星系统的PN波形展开的研究最早可以追溯至上世纪六十年代。然而，对于质量比  $q \ll 1$  的双星系统而言，在PN展开下，高阶项系数收敛速度极慢。同时，PN展开的推导随着阶数的增加而迅速变得复杂，这就导致了用PN展开得到的EMRI无法准确地描述EMRI系统的演化特性。

类似于第3.2.4节中，即使质量比悬殊，也可以利用 EOB 方法进行研究。Nagar 等人开始针对质量比较为悬殊的情形开展的 EOB 研究，并在之后逐渐发展完善 [50]。

如前所述，准确地描述引力自力对EMRI系统演化的影响对于自洽地产生EMRI波形而言至关重要。在极端质量比情形下，Barack 和 Sago 研究了 Schwarzschild 背景下小天体的引力自力 [27]。随后，van de Meent 等人对 Kerr 背景下的引力自力计算进行了推广，目前可以对封闭轨道关于质量比展开到一阶 [94]。这些工作为考虑引力自力的EMRI波形提供了基础，并已取得一定突破 [19]。

### 3.4.3 天琴探测能力

范会敏等人利用 AK 波形，分析了天琴对 EMRI 的探测能力和参数估计精度 [40]。对于质量质量为  $10 M_{\odot}$  的小黑洞所形成的EMRI系统，天琴可以观测的最远红移为 *sim2*。在最乐观的情景中，天琴每年的探测量可能达到数十甚至数百次。在参数估计层面，天琴对EMRI的内禀参数估计精度可以达到  $\sim 10^{-6}$  的相对精度，而对光度距离和天空定位这类外禀误差，典型误差为 10% 和 10 平方度。

## 3.5 Continuous Wave

一些引力波源本身功率并不高，在较长的一段观测时间内系统构型并不会发生明显的变化，这类源统称为连续引力波（Continuous Wave）。对这类信号的描述往往可以用正弦函数或其基础上的修正得到。这类源主要包括高频（ $10 - 1,000$  Hz）的旋转中子星、中低频（ $10^{-4} - 1$  Hz）的双白矮星、低频（ $10^{-9} - 10^{-6}$  Hz）的大质量双黑洞。大质量双

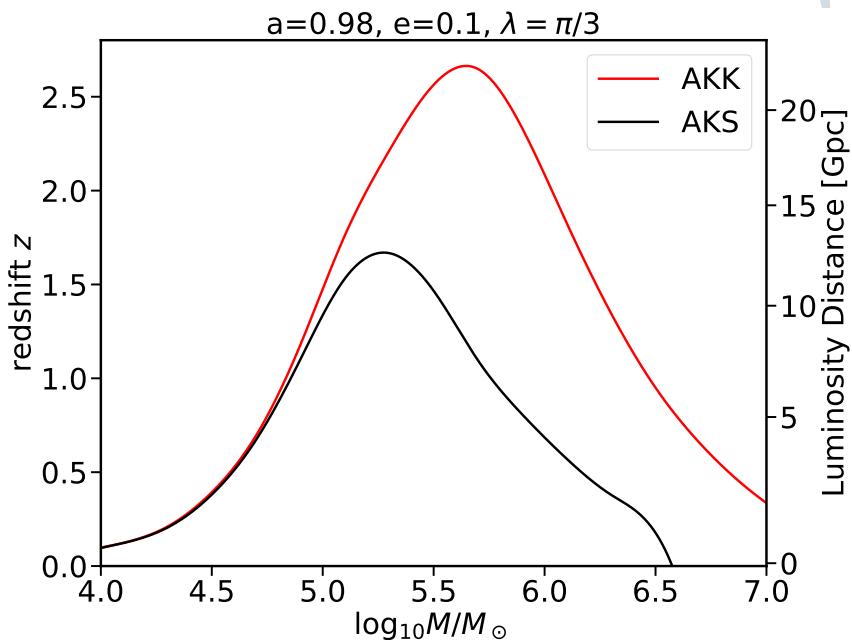


图 3-13 考虑大质量黑洞自旋为 0.98，偏心率为 0.1，倾角为  $\pi/3$  时，天琴对 EMRI 系统的探测极限距离。黑线代表 AKS 波形，红色曲线代表 AKK 波形

Figure 3-13 TianQin's horizon distance for EMRI systems as a function of MBH mass, assuming a MBH spin of 0.98, eccentricity of 0.1 and inclination angle of  $\pi/3$ . The black line adopts the AKS waveform, while the red curve adopts the AKK waveform. Credit: [40]

黑洞的物理图景已经在第3.2.5.2节作过介绍，这里不再详述。对于旋转中子星和双白矮星系统，由于信号强度本身较弱，只能观测到较近的信号，通常仅将目光关注于银河系内。

### 3.5.1 变形的中子星

为简单起见，中子星可以视作三轴椭球（triaxial ellipsoid）刚体。其对应的质量四极矩  $\mathbf{I}$ ，

$$\mathbf{I} = \begin{pmatrix} I^{xx} & 0 & 0 \\ 0 & I^{yy} & 0 \\ 0 & 0 & I^{zz} \end{pmatrix} \quad (3.67)$$

若中子星自转轴为  $z$  轴方向, 其角速度为  $\omega$ , 则其在  $t$  时刻的四极矩张量为

$$\begin{aligned} \mathbf{I}' &= \mathbf{R}_3(\omega t) \mathbf{I} \mathbf{R}_3^{-1}(\omega t) \\ &= \begin{pmatrix} \cos \omega t & \sin \omega t & 0 \\ -\sin \omega t & \cos \omega t & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I^{xx} & 0 & 0 \\ 0 & I^{yy} & 0 \\ 0 & 0 & I^{zz} \end{pmatrix} \begin{pmatrix} \cos \omega t & -\sin \omega t & 0 \\ \sin \omega t & \cos \omega t & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \quad (3.68)$$

$$= \begin{pmatrix} I^{xx} \cos^2 \omega t + I^{yy} \sin^2 \omega t & (I^{yy} - I^{xx}) \cos \omega t \sin \omega t & 0 \\ (I^{yy} - I^{xx}) \cos \omega t \sin \omega t & I^{xx} \sin^2 \omega t + I^{yy} \cos^2 \omega t & 0 \\ 0 & 0 & I^{zz} \end{pmatrix} \quad (3.69)$$

通过对其求时间二阶导, 并且定义椭率为

$$\varepsilon \equiv \frac{I^{xx} - I^{yy}}{I^{zz}} \quad (3.70)$$

其中默认  $I^{xx} > I^{yy}$ 。代入公式3.6, 可以得到

$$h_{ij}^{TT} = \frac{4G\mu\varepsilon I^{zz}\omega^2}{c^4 D} \begin{pmatrix} -\cos 2\omega t & \sin 2\omega t & 0 \\ \sin 2\omega t & \cos 2\omega t & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.71)$$

再引入轨道倾角  $\iota$ , 并用  $f_0 = \omega/\pi$  来表示引力波频率<sup>1</sup>, 可以得到,

$$h_+ = -h_0 \frac{1 + \cos^2 \iota}{2} \cos 2\pi f_0 t \quad (3.72)$$

$$h_\times = h_0 \cos \iota \sin 2\pi f_0 t \quad (3.73)$$

其中, 振幅  $h_0$  定义为

$$h_0 \equiv \frac{4\pi^2 G \varepsilon I^{zz} f_0^2}{c^4 D} \quad (3.74)$$

 **注 3.7.** 若考虑更一般的情况, 即, 连续引力波并非单频波, 可以设中子星自转的相位为  $\phi$ 。将四极矩对时间求二阶导时, 会出现项

$$\dot{\phi}^2 \cos 2\phi + \frac{1}{2} \ddot{\phi} \sin 2\phi$$

可以看到, 当忽略高阶项时, 回归公式3.72

<sup>1</sup>我们遵从惯例, 默认  $f$  指  $f_{\text{GW}}$ , 而  $\omega$  指  $\omega_{\text{rotation}}$

由于中子星形态极为简单，因此一般会预期中子星具有球对称性质。当然，在现有主流理论下，中子星表面是固态，因此可能可以支撑中子星表面的形变，这样的形变会导致中子星具有非零椭率  $\epsilon$ 。其对应的引力波信号波形与公式3.72是一致的。

除了由转动引起的引力波以外，对中子星的微扰会激发其正则模，这些正则模的震荡也会激发引力波。按照模式物理机制的不同，分为 f-modes (fundamental modes), p-modes (pressure modes), g-modes(gravity modes) 等。这些模式的存在可以通过理论推导得出，也通过对中子星的电磁观测，特别是 X 射线的 giant flare 现象中的 Quasiperiodic oscillations (QPOs)，可以间接证明不同模式的存在。

上述这些基础模式都是建立在近稳态、“冷”的中子星模型基础上的。当中子星刚刚形成的时候，其温度很高，且偏离稳态较大，会有一些新的模式产生，如 r-modes (rotation modes) 和 w-modes (gravitational-wave modes)。有关不同模式的推导和性质，详见 [70] 第 11.2.1 节的讨论。

按照对源信息掌握多少的不同，可以将中子星引起的连续引力波源分为 targeted search, directed search 和 blind search 三种。对如蟹状星云等的射电观测，可以精确测定出一部分中子星的方位、频率、甚至距离等。这种有的放矢的 targeted search 可以很好地避免由过多地匹配尝试引入的误警，从而实现对探测灵敏度的提升。此外，还有一部分中子星，如 Scorpius X-1 (Sco X-1)，由于是在 X 射线实现的探测，无法获得频率信息；同时，此类系统中，中子星吸积伴星物质从而有频率提升，因此其频率具体演化更是未知。这类中子星的搜寻适用 directed search，即知道方位，但其他参数未知。参数空间的减小可以带来灵敏度的提升，但终究要比 targeted search 灵敏度更差。blind search 即全盲搜索，无具体电磁观测信息作为输入，相对来说探测灵敏度最差。

LIGO/Virgo 利用现有的数据，尚未探测到来自旋转中子星的连续引力波信号，但已经给出有意义的上限。特别的，对部分中子星而言，如第3.3.1节所述，可以观测到整体自转减慢的效应，可以由磁偶极矩辐射解释，也可以由引力波辐射解释。假设完全由引力波引起时，可以得到对应的理论最强引力波强度，即所谓 spin-down limit。对 11 个已知参数的中子星的 targeted search 表明，其自转减慢不能完全由引力波辐射导致，具体比较如图3-14所示。另外，对 Cassiopeia A, Vela Jr. and G347.3 等中子星的 directed search 表明，其椭率应当小于  $7.2 \times 10^{-9}$  [8, 73]。

### 3.5.2 双白矮星

白矮星是密度极高的天体，仅次于中子星和黑洞，一个太阳质量的白矮星体积只有地球那么大。白矮星的内部已不再有热核反应，故没有能量继续产生，自身的光度和温度会随着能量的释放而降低，此时与引力抗衡的只有电子的简并压。白矮星体积虽小，

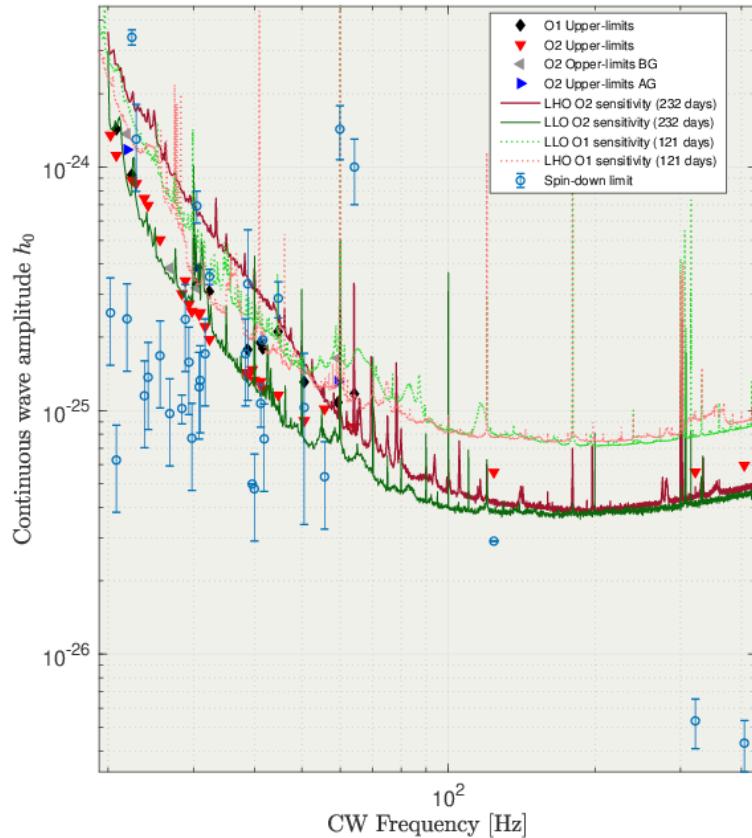


图 3-14 LIGO/Virgo 给出的中子星连续引力波强度上限，与部分已知中子星的 spin-down limit 的比较

Figure 3-14 Upper limit of the gravitational wave from neutron star continuous wave with LIGO/Virgo, comparing with some neutron star's spin-down limit. Credit: [93]

但数量众多，是天体中最常见的致密天体。

致密双星系统是由两颗致密天体相互绕转形成，是星系中极其常见的天体系统，广泛的分布在星系各处。在银河系中，双白矮星的数量最多，比其他致密双星高两个数量级。根据目前的模型预言其数量甚至高达 $\sim 10^8$ ，数量远多于其他致密双星，其中轨道周期小于几个小时能够辐射出 mHz 的引力波的数量约数千万。按照组成双星的致密天体的类型，致密双星可以分为：恒星质量双黑洞、双中子星、双白矮星以及包括黑洞-中子星和黑洞-白矮星等的混合致密双星系统。

在空间引力波探测器灵敏的频段，有大量的致密双星旋近的连续引力波信号（通常是指周期（秒~小时）的双星），主要来源银河系内的双白矮星，无论是足够强的能够被单独识别出来的信号，还是由很多无法单独识别的信号非相干叠加形成的前景噪声，都希望在未来被空间引力波探测器所探测到。

双白矮星辐射的引力波，作为电磁波观测的补充，能够帮助我们丰富双白矮星以及白矮星本身的认识：包括双白矮星的形成、演化、归宿以及白矮星的内部结构，具体如下。第一，致密的双白矮星系统在演化过程至少经历了两个质量转移阶段，因此给测试双星演化模型提供了良好的对象，特别是对于我们理解质量转移和公共包层阶段。第二，双白矮星是许多有趣系统的前身星：Ia 型超新星，它们可用作宇宙距离指示器；AM CVn 系统，即半接型的双白矮星，其谱线通常是由 He 线主导；.Ia 型超新星，光度比 Ia 型超新星弱很多，但数量更多。此外，一般认为两个白矮星合并可以产生稀有的星体，如大质量的白矮星（甚至孤立的中子星）、亚矮星-O 和 R Corona Borealis 星。第三，轨道周期从一小时到几分钟的分离型的双白矮星特别适合于研究潮汐的物理，这种现象与白矮星的内部性质直接相关。研究恒星内部结构对潮汐力的反应可以给我们提供重要的信息，例如，关于白矮星黏度和它的起源，这将完善我们关于白矮星内部物质的理解。

另一方面，研究双白矮星的引力波不仅有助于研究双白矮星相关的物理，还可以揭示其他天体物理学或者基础物理的本质。第一，可以通过引力波探测反推银河系结构。一些来自双白矮星的引力波信号将能够由引力波探测器单独分辨（即测量其个体的属性）。来自银河系双星中不可分辨的重叠信号将形成随机背景信号。可分辨和不可分辨的信号都将提供关于银河系恒星总体数量的信息，从而可用于研究银河系的重子物质的含量和形状。第二，围绕双白矮星旋转的行星或第三颗恒星，会给双白矮星的引力波信号带来额外的频率调制，因此我们可以从双白矮星的引力波信号提取关于围绕双星的系外行星或第三颗恒星的信息。第三，双星的引力波辐射由广义相对论所预言和描述。因此，双白矮星的引力波可以用来检验引力理论并对修改引力理论作出限制。

对于致密双星辐射的连续引力波信号，其振幅可由下式估计

$$\begin{aligned} h(n, e) &= \left[ \frac{16\pi G}{c^3 \omega_g^2} \frac{L(n, e)}{4\pi d^2} \right]^{1/2} \\ &= 1.0 \times 10^{-21} \frac{\sqrt{g(n, e)}}{n} \left( \frac{\mathcal{M}}{\text{M}_\odot} \right)^{5/3} \left( \frac{P_{\text{orb}}}{1 \text{ hr}} \right)^{-2/3} \left( \frac{d}{1 \text{ kpc}} \right)^{-1}, \end{aligned} \quad (3.75)$$

其中  $\mathcal{M} = (M m)^{3/5}/(M + m)^{1/5}$  和  $P$  分别是双星的啁啾质量、轨道周期。高次谐波的相对功率系数  $g(n, e)$ <sup>[78]</sup> 形式较为复杂，但有

$$\sum_{n=1}^{\infty} g(n, e) = \frac{1 + (73/24)e^2 + (37/96)e^4}{(1 - e^2)^{7/2}}. \quad (3.76)$$

特别地，对于圆轨道的双星只有二次谐波辐射能量，即  $e = 0$ ,  $n = 2$ ，此时振幅

$$h = 5.0 \times 10^{-22} \left( \frac{\mathcal{M}}{\text{M}_\odot} \right)^{5/3} \left( \frac{P_{\text{orb}}}{1 \text{ hr}} \right)^{-2/3} \left( \frac{d}{1 \text{ kpc}} \right)^{-1}. \quad (3.77)$$

当双星演化到两颗星靠的足够近时，像星风、吸积和物质交流等对演化的影响非常显著。像这样相互影响的双星系统称为密近双星。根据密近双星整体的演化情况，我们可以把密近双星分为 3 种类型：a、分离型，每一子星都比各自的 Roche 瓣部分小很多；b、半接型，一颗子星已膨胀并充满自己的 Roche 瓣；c、相接型，两颗子星都充满 Roche 瓣。对于质量都为  $M$ ，并合前  $\tau$  时刻的致密双星，可以应用第3.2.1.1节中的公式，对其辐射的引力波频率和振幅进行估算

$$f = 1.9 \text{ Hz} \left( \frac{1.4 \text{ M}_\odot}{M} \right)^{5/8} \left( \frac{1 \text{ day}}{\tau} \right)^{3/8}, \quad (3.78)$$

$$h = 1.7 \times 10^{-23} \left( \frac{15 \text{ Mpc}}{d} \right) \left( \frac{1 \text{ day}}{\tau} \right)^{1/4} \left( \frac{M}{1.4 \text{ M}_\odot} \right)^{5/4}. \quad (3.79)$$

当前，已有不少研究对空间引力波探测器探测双白矮星系统的能力进行研究。如，Nelemans 等人<sup>[76]</sup> 使用双星的星族合成模型，结合双白矮星的银河系分布模型，连同对中子星和黑洞等致密双星的结果，来模拟用 LISA 能够观测到的银河系致密双星的信号，他们的工作给出了 LISA 预期能够探测到的不同类型的双白矮星的数目。Korol 等人<sup>[56]</sup> 给出了最新的 (250 万公里臂长构型的)LISA 对银河系双白矮星可分辨源的数目估计，显示 LISA 运行 4 年预期探测到约 25000 对银河系双白矮星。此外，由大量的白矮星或中子星组成的银河系致密双星，在低频处会形成一个不可分辨的随机引力波前景。

黄顺佳等人<sup>[52]</sup> 研究了天琴对双白矮星的探测能力。天琴预期在五年观测时间内可以探测到约 10000 对银河系双白矮星。同时预计天琴将分别以  $\sim 10^{-7}$  和  $\sim 0.2$  的精度

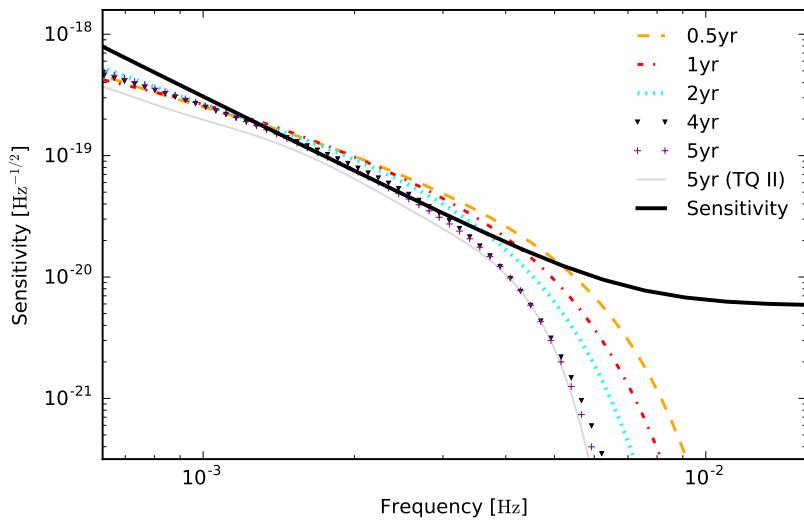


图 3-15 银河系致密双星形成的前景

Figure 3-15 GW foreground due to Galactic binaries. Credit: [52].

测量其轨道周期和振幅，很高比例的双星（39%）都可以被精确定位到 1 平方度以内。在图3-15中展示了由双白矮星引起的前景信号对天琴灵敏度曲线的影响。

目前有数十对双白矮星已被电磁波望远镜观测到，其中引力波信号足够强的能够被天琴所探测到，从而实现电磁波和引力波的联合观测，这些源我们称之为验证双星（verification binaries）。在表3-4中，给出了多个电磁波段观测到的双白矮星系统。验证双星可以用来对天琴的仪器噪声和探测能力进行校准和评估。对于已知的银河系双白矮星，Thomas Kupfer 等人根据电磁波望远镜的 Gaia2018 年最新观测数据，计算了 LISA 预期探测数以及参数估计 [57]，结果表明 LISA 运行 4 年预期能探测到 13 个银河系双白矮星，这些已被电磁波望远镜观测的并且有望被未来引力波探测器探测的源被称之为验证双星。黄顺佳等人的研究结果表明，天琴可以探测到其中的 12 个。Korol 等人 [56] 结合 LISA 和电磁波段的 Gaia 和 LSST，计算了电磁波和引力波联合探测的银河系双白矮星数目，预期 LISA 和 Gaia 能够联合探测到 24 个银河系双白矮星、LISA 和 LSST 能够联合探测到 73 个银河系双白矮星，该结果依赖于双白矮星模型。

### 3.6 SGWB

如第3.5.2节所示，当许多不同的、强度相近的引力波源交叠在一起，单个信号无法分辨。前景信号指的是信号综合而言强于噪声，如图3-15所示；而若信号弱于噪声，则对应随机引力波背景，在某些场景下，并不区分前景信号和背景噪声，统一归于随机引

源	$\lambda$ [deg]	$\beta$ [deg]	$f = 2/P$ [mHz]	$d$ [kpc]	$M$ [ $M_{\odot}$ ]	$m$ [ $M_{\odot}$ ]	$\iota$ [deg]
<b>半接型双白矮星</b>							
J0806	120.4442	-4.7052	6.22	[5]	0.55	0.27	38
V407 Vul	294.9832	46.7858	3.51	1.786	[0.8]	[0.177]	[60]
ES Cet	24.6099	-20.3365	3.22	1.584	[0.8]	[0.161]	[60]
SDSS J1351	208.3820	4.4724	2.12	1.317	[0.8]	[0.100]	[60]
AM CVn	170.3831	37.4390	1.94	0.299	0.68	0.125	43
SDSS J1908	298.2032	61.4571	1.84	1.044	[0.8]	[0.085]	15
HP Lib	235.0803	4.9608	1.81	0.276	0.645	0.048	30
PTF1 J1919	308.9868	69.0329	1.48	1.338	[0.8]	[0.066]	[60]
CXOGBS J1751	268.0519	-6.2510	1.45	0.971	[0.8]	[0.064]	[60]
CR Boo	202.2673	17.8960	1.36	0.337 <sup>a</sup>	0.885 <sup>a</sup>	0.066	30
V803 Cen	216.1614	-30.3132	1.25	0.347 <sup>a</sup>	0.975 <sup>a</sup>	0.084	13.5
<b>分离型双白矮星</b>							
SDSS J0651	101.3396	5.8048	2.61	0.933	0.247	0.49	86.95
SDSS J0935	130.9795	28.0912	1.68	0.645 <sup>b</sup>	0.312 <sup>b</sup>	0.75	[60]
SDSS J1630	231.7612	63.0501	0.84	1.019	0.298	0.76	[60]
SDSS J0923	133.7151	14.4268	0.51	0.229	0.275	0.76	[60]
<b>双热亚矮星</b>							
CD-30°11223	221.1718	-16.6125	0.47	0.337	0.54	0.79	82.9

表 3-4 验证双星候选者的源参数，包括：源的黄道坐标 ( $\lambda, \beta$ )、引力波频率  $f$ 、源到地球的距离  $d$ 、源的倾角  $\iota$  以及双星的质量  $M$  和  $m$ 。带方括号 [] 的质量和倾角的测量是基于演化阶段和质量比估计的结果。J0806 的距离测量误差较大，计算中假设为 5 kpc。表中所有的数据都可以在文献及其参考文献中找到。<sup>a</sup>[82], <sup>b</sup>[31]。

Table 3-4 Parameters for the verification binaries, values with bracket indicates indirect estimates. Credit:

[57] and reference within.

力波背景。

中心极限定理保证了，随机引力波背景在不同频率上符合 Gaussian 分布，由此，其统计学特性类似于噪声。随机引力波背景与噪声的主要差别在于：不同探测器上的噪声具有独立性；但其记录的随机引力波背景则具有相关性。除了天文学起源的大量引力波叠加，随机引力波背景还可以源自宇宙学过程，如暴胀、再加热、弱电相变、宇宙弦，等等。当前，并没有任何实验探测到随机引力波背景，但是由宇宙微波背景辐射等手段的探测已经可以给出一定上限。根据当前对恒星级双黑洞的事件率估算，可以预期 LIGO/Virgo 可以积累约 40 个月的数据，探测到对应的随机引力波背景。LISA 等空间引力波探测可以通过约一个月的数据积累，探测到对应的随机引力波背景。

可以定义一个无量纲的密度参数  $\Omega_{\text{GW}}$  描述引力波强度：

$$\Omega_{\text{GW}}(f) = \frac{1}{\rho_c} \frac{d\rho_{\text{GW}}}{d(\ln f)} \quad (3.80)$$

其中，临界密度  $\rho_c = 3H_0^2c^2/8\pi G$ ，对应着让宇宙空间曲率保持平直时对应的宇宙平均密度。引力波能量密度  $\rho_{\text{GW}}$  可由下式得到：

$$\rho_{\text{GW}} = \frac{c^2}{32\pi G} \langle \dot{h}_{ab}(t, \vec{x}) \dot{h}^{ab}(t, \vec{x}) \rangle \quad (3.81)$$

其中， $h_{ab}(t, \vec{x})$  是位矢为  $\vec{x}$  的  $t$  时刻度规微扰， $\langle \dots \rangle$  表示统计学期望。

对于来自某一个族群  $k$  的信号，假设单个信号的物理参数由  $\theta_k$  决定，其随机引力波背景可以由下式计算得到

$$\Omega_{\text{gw}}(f) = \frac{f}{\rho_c} \int d\theta_k \int_0^{z_{\max}} dz \frac{R_m(z|\theta_k) \frac{dE_{\text{gw}}}{df_s}(f_s, \theta_k)}{(1+z)H(z)} \quad (3.82)$$

这其中， $R_m(z|\theta_k)$  是并合率， $\frac{dE_{\text{gw}}}{df_s(f_s, \theta_k)}$  是在源参考系  $f_s = f(1+z)$  下辐射的能谱分布。 $H(z)$  是 Hubble 参数，描述了宇宙膨胀的历史。

针对天文学起源和宇宙学起源的随机引力波背景，梁正程等人研究了天琴、天琴联合 LISA 等不同情形下对各类随机引力波背景的探测能力。研究表明，对于天文学起源的随机引力波背景，天琴可以以不到一百的信噪比实现探测；对于宇宙学起源的随机引力波背景，其模型不确定较大，最极端的模型下，信噪比可以达到一万。

脉冲星计时阵列可以探测来自大质量黑洞旋近的随机引力波背景。在 2020 年，NANOGrav 团队报告了他们分析 12.5 年间积累数据后的发现：仅凭高斯噪声无法完全解释数据??。然而，数据和广义相对论下所预言的随机引力波背景也并不一致。之后，在 2021 年，PPTA 团队通过分析他们的数据，也确认了这一结果??。业内以“共同谱过程”(common-spectrum process) 来描述这一非噪声、非随机引力波背景的有趣发现。黄

庆国等人认为，如果考虑引力波可以存在标量横波偏振这一非广义相对论预言的偏振，则可以很好地解释数据 [34]。脉冲星计时阵列团队的后续分析中也确认了这一发现，但指出如果移除某一颗特殊的脉冲星的话，则这一结论会被大大削弱。相关问题仍然是随机引力波背景探测领域的有趣问题。

## 3.7 其他波源

### 3.7.1 Supernova

大质量恒星在临近死亡时，由于中心核聚变停止，无法有效提供热能和光压抵抗引力塌缩，加速向内的物质会装上致密内核，从而向外反弹，最终可能会经历超新星爆发。引力塌缩和超新星爆发都是可以产生引力波信号的过程。超新星从观测上分为 I 型超新星和 II 型超新星，其中 Ia 型超新星主要由白矮星的核爆发产生，并无有效地引力波辐射；而其余的过程，包括 Ib, Ic 和 II 型等，一般均认为源于大质量恒星的塌缩，因此一般认为有可观测的引力波效应，特别是若反弹过程缺乏球对称性，更是会有较大的  $\tilde{h}$ 。

对超新星爆发过程的精确数值模拟在计算上依然一个困难的问题，超新星爆发过程是不多的几个宇宙中四个基本作用力都需要仔细考虑的场景。因此，对应的精确引力波波形尚不可得，不同的小组只能通过设定不同的初始条件，定性地理解超新星爆发过程及对应的引力波信号。当前研究表明，引力波能将整个系统中约  $10^{-7} - 10^{-5}$  的质能提取，而对应的引力波能量集中于  $\sim 200 - 1000\text{Hz}$  频段。解析来看，大质量天体的塌缩过程可以由自由落体 (free fall) 时标表征，对应的动力学时标为  $\tau_{ff} = \sqrt{\frac{3\pi}{32} \frac{1}{G\rho}}$ ，其中  $\rho \sim 10^{18}\text{kg m}^{-3}$  是核心物质密度，对应的时标约为  $0.1\text{ms}$ ，与前述频段大致吻合。其典型强度为

$$h \sim 6 \times 10^{-21} \left( \frac{E}{10^{-7} M_\odot} \right)^{1/2} \left( \frac{1\text{ms}}{T} \right)^{1/2} \left( \frac{1\text{kHz}}{f} \right) \left( \frac{10\text{kpc}}{r} \right) \quad (3.83)$$

### 3.7.2 引力波暴

除了如双星绕转般物理图景比较简单的引力波源系统，还存在如超新星爆发般无法精确理论建模的引力波源。如果能量爆发较为集中，则可以将其视为引力波暴，通过对具体波形依赖程度较小的数据处理方法进行探测。第3.7.1节中描述超新星爆发一定程度上也可以归类为引力波暴。首先发现第一起引力波事件 GW150914 的数据处理流水线 cWB 全称 coherent WaveBurst，就是设计用于处理难以准确建模的引力波波形的。

由于引力波探测手段与电磁波手段完全不同，我们有理由相信，引力波探测可能会发现一些在电磁波观测中前所未见的物理过程，而这种未知过程触发的信号，将极大地

加深我们对于宇宙和基础物理的认知。

如，宇宙弦的断折，可能会产生引力波信号。宇宙弦是类似于弦一般的一维拓扑缺陷，某些理论预言在早期宇宙相变过程中会产生这类宇宙弦。宇宙弦由其单位长度上的质量  $\mu = m/l$  表征，更一般地由无量纲量  $G\mu/c^2$  表征。 $G\mu/c^2$  的具体取值取决于相变发生的能标，如，大一统理论的对称破缺对应着  $G\mu/c^2 \sim 10^{-6}$ 。宇宙弦在偶然情况下会卷曲并与自身发生碰撞，由于宇宙弦本身巨大的张力，交叠处尖端会获得巨大的 Lorentz boost (Lorentz 增速)，在这一固定方向上会释放大量引力波辐射。

### Exercise 3.1

计算一个均质球绕对称轴旋转时的引力波强度。

### Exercise 3.2

根据《西游记》的设定，金箍棒重一万三千五百斤，合 6.75 吨。孙悟空在要金箍棒时，通常的长度设为 1 米，旋转频率设为 1Hz。估算在一个波长外测量这一系统的引力波强度。

### Exercise 3.3

请根据稳相近似，推导出两个准圆轨道黑洞旋近的频率域波形（公式3.39与公式3.40）。

### Exercise 3.4

考虑质量为  $10M_\odot$ - $10M_\odot$  的双黑洞系统，假设其距离我们 100Mpc，倾角  $\iota = 0$ 。请编写程序作图，

- ▶  $t \in [t_c - 1\text{yr}, t_c]$  时间内频率随时间变化的图。
- ▶  $t \in [t_c - 1\text{yr}, t_c]$  的时间内相位随时间变化的图。
- ▶  $t \in [t_c - 1\text{s}, t_c]$  的时间域波形（假设  $Q(\text{angles}) = 1$ ）。
- ▶  $t \in [t_c - 1\text{yr}, t_c]$  对应的频率域波形，考虑 OPN（以 Bode 图形式给出）。
- ▶ (加分题 1) 功率谱密度 (PSD) 是描述探测器不同频率处噪声强度的图形。振幅谱密度 (ASD) 是 PSD 开根号后的结果。请找出 LIGO 的 ASD，(任意一种均可，请注明来源)，并将其与上述双黑洞系统的波形振幅一同给出（以 log-log 形式画出）。
- ▶ (加分题 2) 请自行找出天琴的 ASD，并将其与上述双黑洞系统的波形振幅一同给出。设从双星并合前 5 年开始观测(以 log-log 形式画出)。

### Exercise 3.5

设中子星表面磁场强度  $B$ , 有  $\mu_{\perp} = BR^3$ , 其中  $\mu_{\perp}$  是磁偶极矩,  $R = 10\text{km}$  是中子星半径,  $I = 10^{45}\text{g cm}^2$  是转动惯量。

- ▶ 请推导出  $B$  的表达式 (请用  $P$  和  $\dot{P}$  表示)
- ▶ 若该中子星有观测量  $P = 1\text{ms}$ ,  $\dot{P} = 10^{-20}$ , 假定中子星辐射完全由磁偶极矩辐射主导, 请计算出其年龄
- ▶ 若该中子星有观测量  $P = 10\text{s}$ ,  $\dot{P} = 10^{-10}$ , 假定中子星辐射完全由磁偶极矩辐射主导, 请计算出其磁场强度
- ▶ 若该中子星有观测量  $P = 1\text{s}$ ,  $\dot{P} = 10^{-14}$ , 假定中子星辐射完全由磁偶极矩辐射主导, 请计算出其辐射功率
- ▶ 假定中子星辐射完全由引力波辐射主导, 请根据公式3.53, 计算出上述三种情况下的  $\epsilon$

### Exercise 3.6

请在  $P - \dot{P}$  平面上作图,  $P$  的取值范围为  $0.001 - 10\text{s}$ ,  $\dot{P}$  的取值范围为  $10^{-20} - 10^{-8}$ , 以对数展示, 画出

- ▶  $\tau = 10^3, 10^5, 10^7, 10^9\text{yr}$  的等年龄线 (假定中子星辐射完全由磁偶极矩辐射主导)
- ▶  $B = 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13}, 10^{14}\text{G}$  的等磁场强度线
- ▶  $L = 10^{30}, 10^{33}, 10^{36}\text{erg/s}$  的等辐射强度线

### Exercise 3.7

考虑  $\dot{\omega} = \dot{\omega}_{\text{EM}} + \dot{\omega}_{\text{GW}}$ , 假设磁场强度  $B$ 、椭率  $\epsilon$ 、转动惯量  $I$  等均为常数。记  $\dot{\omega}_{\text{GW}}/\dot{\omega}_{\text{EM}} = \epsilon$ , 试证明  $\epsilon = \frac{n-3}{5-n}$ 。已知 PSR J1640-4631 的制动指数  $n = 3.15$ , 计算相应的  $\epsilon$ 。

2023-03-13

# 4

## 引力波探测手段

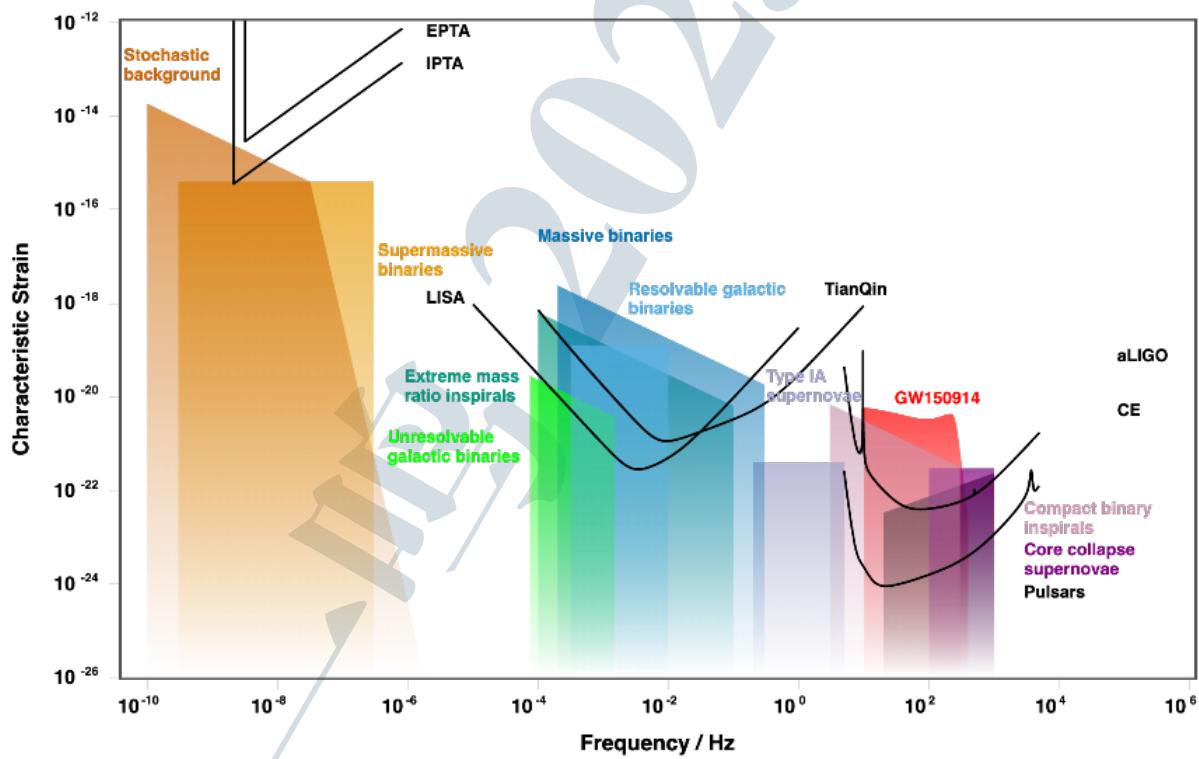


图 4-1 不同频段下的主要引力波源和引力波探测手段。

Figure 4-1 Gravitational wave sources and detection methods in different wavelengths. Credit: [74]

在第三章中，我们了解到，在不同的机制下，可以产生不同类型、不同特性的引力波信号。正如同在电磁波段，人类可以用波长短至  $10^{-12}$ m 的伽玛射线进行观测，也可

以利用长至  $10^3\text{m}$  的射电波段进行观测；引力波的波长/频率涵盖范围也很广；而针对不同频段的波源，也有相对更为合适的探测手段。

在图4-1中，展示了第三章中所讨论的几类不同的引力波波源。并且，根据频率的不同，展示了不同探测手段和对应的灵敏度曲线。在本章中，我们主要讨论不同探测手段的探测原理、噪声源、灵敏度曲线的估计。

## 4.1 棒状探测器

### 4.1.1 原理



图 4-2 Joseph Weber 和他的棒状探测器。

Figure 4-2 Joseph Weber and bar detector. Credit: AIP Emilio Segré Visual Archives

最早开展引力波实验的装置是棒状探测器（bar detector），有时候又被称为共振物质探测器（resonant mass detector）。如图4-2所示，一个典型的棒状探测器为长约 1 米的铝圆柱体，在中间放置压电传感器，能将轴向方向的引力波效应记录下来。为了分析棒

状探测器受引力波的效应，不妨考虑如下简化的模型：两个质点，质量分别为  $m_1$  和  $m_2$ ，以弹性系数为  $k$  的弹簧相接，阻尼常数  $\beta$ ，在静止时长度为  $L$ 。假设系统沿横向无迹坐标系下的  $x$  轴放置，有一引力波  $h_{11} = -h_{22} = h_+ = h_0 \cos \omega t$ ，两个质点的距离变化以  $x$  表示。可以计算出，两个质点间距离符合振动方程：

$$\frac{d^2x}{dt^2} + 2\beta \frac{dx}{dt} + \omega_0^2 x = -\frac{2L}{\pi^2} \ddot{h}_{11} = -\frac{2Lh_0\omega^2}{\pi^2} \cos \omega t \quad (4.1)$$

这其中， $\omega_0 = \sqrt{k/\mu}$  是系统的特征频率， $\mu \equiv m_1m_2/(m_1 + m_2)$  是系统的约化质量。

当系统稳定后，可以得到

$$x(t) = x_{\max} \cos(\omega t + \delta) \quad (4.2)$$

这其中

$$x_{\max} = \frac{2h_0L}{\pi^2} \frac{\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}} \quad (4.3)$$

以及

$$\tan \delta = \frac{2\omega\beta}{\omega^2 - \omega_0^2} \quad (4.4)$$

可以发现，当引力波频率  $\omega$  非常接近探测器特征频率  $\omega_0$  时，探测器的共振会被激发。当  $\omega = \bar{\omega} = \sqrt{\omega_0^2 - 2\beta^2}$  时，共振棒可以达到最大振幅。而当  $\omega = \omega_0$  时，系统的振幅为

$$x_{\max, \text{res}} = \frac{2h_0L}{\pi^2} Q \quad (4.5)$$

其中

$$Q \equiv \frac{\omega_0}{2\beta} \quad (4.6)$$

是系统的品质因子 (quality factor)。

 **注 4.1.** 品质因子可以用来表示振子阻尼性质，品质因子  $Q$  高，对应着振子损失能量慢，振动持续的时间长（余音绕梁，三日不绝）。高的品质因子一般对应较小的阻尼。高的品质因子也意味着较窄的共振频率。

### 4.1.2 天线响应函数

对于一般方位的引力波源而言，探测器的响应不同，具体而言，探测器记录到的信号是探测器张量  $D^{ij}$  和引力波张量  $h_{ij}$  的双缩并

$$h(t) = D^{ij} h_{ij}(t). \quad (4.7)$$

以上述沿  $x$  轴放置的棒状探测器为例，可以发现其中  $D^{11} = 1$ ，而其他项皆为 0。对于更一般的情形，假设  $\hat{\mathbf{l}}$  是探测器轴线方向的单位向量，则有  $h(t) = \hat{l}^i \hat{l}^j h_{ij}(t)$ ，换言之，对于棒状探测器而言，其探测器张量为

$$D^{ij} = \hat{l}^i \hat{l}^j \quad (4.8)$$

而引力波张量则可以写成

$$h_{ij}(t) = \sum_{A=+, \times} e_{ij}^A(\hat{\mathbf{n}}) h_A(t) \quad (4.9)$$

这其中  $h_A$  是对应偏振下的引力波强度，而  $e_{ij}^A(\hat{\mathbf{n}})$  则是来自  $\hat{\mathbf{n}}$  方向的单位偏振张量，其定义为

$$e_{ij}^+(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j - \hat{\mathbf{v}}_i \hat{\mathbf{v}}_j \quad (4.10)$$

$$e_{ij}^\times(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{v}}_j + \hat{\mathbf{v}}_i \hat{\mathbf{u}}_j \quad (4.11)$$

其中  $\hat{\mathbf{u}}$ 、 $\hat{\mathbf{v}}$  是垂直于源方向  $\hat{\mathbf{n}}$  且互相垂直的单位向量，并且  $\hat{\mathbf{u}}$ 、 $\hat{\mathbf{v}}$  与  $-\hat{\mathbf{n}}$  张成一个右手系的正交基底。不难发现，当  $\hat{\mathbf{n}} = \hat{\mathbf{z}}$  时，系统回归到第2.2.2节描述的横向无迹规范下的偏振态

$$e_{ij}^+(\hat{\mathbf{z}}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.12)$$

及

$$e_{ij}^\times(\hat{\mathbf{z}}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.13)$$

通常，为简化计算，对于来自方位  $\hat{\mathbf{n}} = (\theta, \phi)$  的引力波，可以定义探测器天线响应函数

$$F_A(\hat{\mathbf{n}}) = D^{ij} e_{ij}^A(\hat{\mathbf{n}}) \quad (4.14)$$

这样，公式4.7便可以简化为

$$h(t) = h_+(t) F_+(\theta, \phi) + h_\times(t) F_\times(\theta, \phi) \quad (4.15)$$

为了计算  $F_A(\hat{\mathbf{n}})$ ，我们继续假设探测器沿  $x$  轴放置。对位于  $\hat{\mathbf{n}}$  方向的源而言，其传播到观测者的引力波波矢方向的单位向量为  $\hat{\mathbf{k}} = -\hat{\mathbf{n}}$ 。至此，我们只确定了  $(x, y, z)$  坐标系

中  $x$  轴的选取,  $y$  轴和  $z$  轴的方向选取具有任意性。为了计算方便, 不妨设源处于  $x - z$  平面内 (即  $\phi = 0$ ), 由此也可以唯一确定  $y$  轴方向。注意到, 上述假设意味着,  $\mathbf{k} = -\hat{\mathbf{n}}$  也处于  $x - z$  平面中。考虑源坐标系  $(u, v, k)$ , 不妨设  $v$  轴平行于  $y$  轴, 则通过绕  $v$  轴旋转  $\alpha = \frac{\pi}{2} - \theta$  即可以将  $(u, v, k)$  坐标系转至  $(x, y, z)$  坐标系, 相应的旋转矩阵为

$$\mathcal{R}^V(\alpha) = \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix} \quad (4.16)$$

并且有  $\sin \alpha = \cos \theta$  和  $\cos \alpha = \sin \theta$ 。

在源坐标系  $(u, v, k)$  下, 引力波的空间分量表示为

$$h'_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (4.17)$$

在旋转以后,  $(x, y, z)$  坐标系下的分量变为  $h_{ij} = \mathcal{R}_{ik}^V(\alpha)\mathcal{R}_{jl}^V(\alpha)h'_{kl}$ 。对于棒状探测器而言, 只有  $h_{11}$  分量有意义。可以得到

$$\begin{aligned} h_{11} &= \mathcal{R}^V(\alpha)_{1k}\mathcal{R}^V(\alpha)_{1l}h'_{kl} \\ &= [\mathcal{R}_{11}^V(\alpha)^2 - \mathcal{R}_{12}^V(\alpha)^2]h_+ + 2\mathcal{R}_{11}^V(\alpha)\mathcal{R}_{12}^V(\alpha)h_\times \\ &= h_+ \sin^2 \theta \end{aligned} \quad (4.18)$$

换言之, 有

$$\begin{aligned} F_+(\theta, \phi) &= \sin^2 \theta \\ F_\times(\theta, \phi) &= 0 \end{aligned} \quad (4.19)$$

考虑在  $\hat{\mathbf{u}} - \hat{\mathbf{v}}$  矢量张成的平面内, 旋转  $\psi$  角度对应的旋转矩阵为

$$\mathcal{R}(\psi) = \begin{pmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{pmatrix} \quad (4.20)$$

旋转后得到的新基底  $\hat{\mathbf{u}}'$  和  $\hat{\mathbf{v}}'$

$$\begin{aligned} \hat{\mathbf{u}}' &= \begin{pmatrix} \cos \psi \\ \sin \psi \end{pmatrix} \\ \hat{\mathbf{v}}' &= \begin{pmatrix} -\sin \psi \\ \cos \psi \end{pmatrix} \end{aligned} \quad (4.21)$$

这时，由公式4.10和公式4.11，可以得到在新的坐标基底下的引力波偏振单位张量：

$$\begin{aligned}(e_{ij}^+)'(\hat{\mathbf{n}}) &= \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_j' - \hat{\mathbf{v}}_i' \hat{\mathbf{v}}_j' \\(e_{ij}^\times)'(\hat{\mathbf{n}}) &= \hat{\mathbf{u}}_i' \hat{\mathbf{v}}_j' + \hat{\mathbf{v}}_i' \hat{\mathbf{u}}_j'\end{aligned}\quad (4.22)$$

正如同线偏光经过检偏器后会产生一定的投影，引力波在经历坐标变换后同样可以得到引力波的投影转换。不难计算，其对应的投影公式为：

$$\begin{aligned}h'_+ &= h_+ \cos 2\psi - h_x \sin 2\psi \\h'_x &= h_+ \sin 2\psi + h_x \cos 2\psi\end{aligned}\quad (4.23)$$

将棒状探测器的探测器张量  $D^{ij}$  代入天线响应函数式4.14中，可以得到

$$\begin{aligned}F_+(\theta, \phi, \psi) &= \sin^2 \theta \cos 2\psi \\F_x(\theta, \phi, \psi) &= \sin^2 \theta \sin 2\psi.\end{aligned}\quad (4.24)$$

### 4.1.3 噪声

棒状探测器的噪声来源，主要有热噪声、读取噪声和量子噪声。

考虑一个振幅  $h_0 \sim 10^{-21}$  的引力波入射至  $L \sim 1\text{m}$ 、 $\mu \sim 1000\text{kg}$ 、 $Q \sim 10^6$ 、并且  $f = \omega_0/2\pi \sim 1\text{kHz}$  的棒状探测器，将会激起  $x_{\max, \text{res}} \sim 10^{-15}\text{m}$  的振动。注意到系统振动的平均能量为  $\langle E_{\text{tot}} \rangle = \frac{1}{4}\mu x_{\max}^2 (\omega^2 + \omega_0^2) \sim 10^{-21}\text{J}$ 。而这时，室温  $T = 300\text{K}$  的共振棒将具有  $k_B T \sim 10^{-21}\text{J}$  的热能，因此，即使此时的引力波信号处于共振频率处，也无法将其从噪声中分辨出来。

室温环境下的棒状探测器由于仪器中的热能，会影响探测器对引力波的响应，对此，后续建造的棒状探测器都会在极低温中运行，约  $100mK$ ，这时，由热噪声引起的位移测量误差为

$$\langle \delta l^2 \rangle_{th}^{1/2} = \left( \frac{k_B T}{4\pi^2 M f^2} \right)^{1/2} \sim 6 \times 10^{-18}\text{m}. \quad (4.25)$$

这一振幅相对于预期的引力波信号而言仍然偏大。对于高 Q 值的探测器来说，在共振频率处，由热噪声引起的振动会有很强的相关性。对于一个  $Q \sim 10^6$  的探测器而言，在共振频率附近的热噪声，其持续时标在  $Q/f \sim 10^3\text{s}$ ；而引力波信号在  $\text{ms}$  量级上就会发生显著变化。在  $\text{ms}$  量级上，热噪声引起的随机性相比公式4.25降低  $(10^3\text{s}/1\text{ms})^{1/2} = Q^{1/2}$  倍。因而，热噪声在  $6 \times 10^{-18}\text{m}$  水平上的棒状探测器，可以探测弱得多的引力波信号

$$\langle \delta l^2 \rangle_{th: \text{ms}}^{1/2} = \left( \frac{k_B T}{4\pi^2 M f^2 Q} \right)^{1/2} \sim 6 \times 10^{-21}\text{m}. \quad (4.26)$$

我们发现，在极低温条件下共振棒探测器的热噪声可以压制到  $10^{-20}$  以下的水平。

在 kHz 频段处，棒状探测器的零点振动能约为

$$\langle \delta l^2 \rangle_{\text{quant}}^{1/2} = \left( \frac{\hbar}{2\pi^2 M f} \right)^{1/2} \sim 4 \times 10^{-21} \text{ m.} \quad (4.27)$$

注意到，在 ms 的时间尺度上，由量子力学引起的噪声已经和热噪声处于同一量级，进一步地压低探测器整体噪声必须依赖于对量子噪声的压制。这一标准量子极限并非不可逾越的：量子力学给出的不确定性原理只是限制了一对共轭量的综合测量不确定度。在光学系统中，通过引入压缩光（squeezing light）技术，可以一定程度上改变量子力学不确定性原理在不同分量上的简并态，使得感兴趣的可观测量上误差减小。可以在棒状探测器中引入此类压缩态测量技术实现对标准量子极限的突破。

要记录引力波信号，需要通过压电传感器，将引力波产生的探测器形变转换成电信号并放大。这一信号读取过程也会引入读取噪声。结合压电传感器，分析受迫振动，结果表明，棒状探测器会在本征频率产生两个共振峰。其噪声曲线如4-3所示。

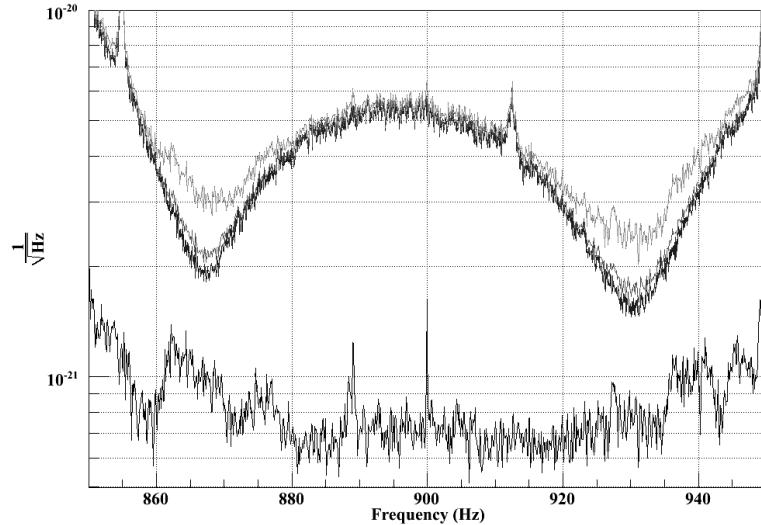


图 4-3 2005 年 9 月 15 日棒状探测器 AURIGA 探测器和 Virgo 探测器的噪声功率谱密度。

Figure 4-3 Strain noise power spectral densities of AURIGA and Virgo detectors on September the 15<sup>th</sup>, 2005. Credit: [28]

## 4.2 地面激光干涉探测器

在常见的引力波相关科普材料中，通常描述的说法是：引力波会引起时空的变化，通过类似于激光干涉，可以测出微小的位移，从而探测出引力波的踪迹。但这里似乎存在一个矛盾之处，常有初学者针对这一物理图像提出问题。聪明的你不妨放下手中的讲义，想一想，这个图像具体有什么问题。

### 4.2.1 原理

有前面提到的困惑一点也不奇怪。在高中、本科的大学物理实验中，与引力波干涉仪相近的实验是 Michelson 干涉实验，如图 4-4 所示。实际上，地基激光干涉引力波探测器，如激光干涉引力波观测台，Laser Interferometer Gravitational-wave Observatory (LIGO) 等，的确类似于一个极大的 Michelson 干涉仪。对于标准的 Michelson 实验而言，可以把激光看做一把标准不动的标尺，而反射镜位置的移动会影响干涉臂臂长，用一把标准的尺测量变化长度的物体，自然可以读出读数的不同。

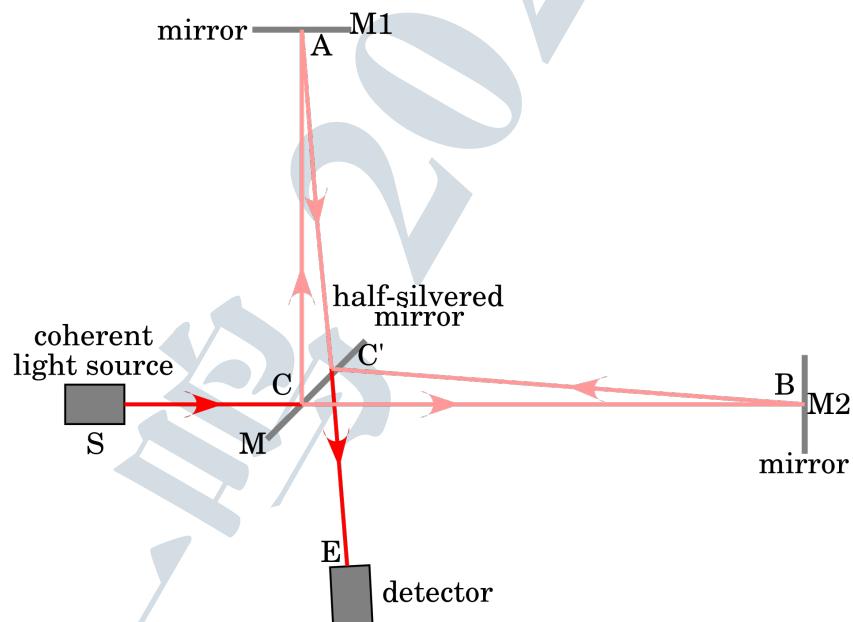


图 4-4 Michelson 干涉仪光路示意图。

Figure 4-4 Path of light in Michelson interferometer. Credit:

[https://en.wikipedia.org/wiki/Michelson\\_interferometer](https://en.wikipedia.org/wiki/Michelson_interferometer)

然而，在第 2.3.1 节中，我们提到引力波会对两个相隔一定距离的物体间距产生影响，这一影响不依赖于物体的性质，只依赖于物体的尺度。换言之，引力波的效应对于待测

的臂长和用于测量的激光波长而言是完全相同的。就好像用热膨胀系数相同的木尺去测量木桌的长度，由于温度对待测物和测量标准的影响相同，不同温度下量出的读数不会发生变化。这样的逻辑似乎没有问题，那么LIGO究竟是如何探测到引力波的呢？我们在第2.3.2节中简单分析了电磁波传播过程中受引力波影响的效应。在这一节中，我们讨论激光干涉型引力波探测器受引力波的影响。

考虑一个 $+$ 模的引力波信号

$$h_+(t) = h_0 \cos(\omega t) \quad (4.28)$$

则有，横向无迹规范下，时空间隔可以表示为

$$ds^2 = -c^2 dt^2 + [1 + h_+(t)] dx^2 + [1 - h_+(t)] dy^2 + dz^2 \quad (4.29)$$

对于走零测地线的光子而言， $ds^2 = 0$ 。我们考虑在 $t_0$ 时刻从分束器出发的沿 $x$ 方向的光子，在一阶近似下有

$$dx = \pm c dt \left[ 1 - \frac{1}{2} h_+(t) \right] \quad (4.30)$$

设 $t_1$ 时刻下，到达尽头，坐标 $x = L_x$ ，又在 $t_2$ 时刻下，光子回到分束器，其坐标 $x = 0$ 。可以得到

$$t_2 - t_0 = \frac{2L_x}{c} + \frac{1}{2} \int_{t_0}^{t_2} dt' h_+(t') \quad (4.31)$$

这就意味着，光子经过终端反射镜反射后，其所耗时间为 $2L_x/c$ ，在此基础上还有 $h_0$ 的一阶修正。代入引力波信号（公式4.28），有

$$\begin{aligned} t_2 - t_0 &= \frac{2L_x}{c} + \frac{h_0}{2\omega} \left[ \sin\left(\omega(t_0 + 2L_x/c)\right) - \sin(\omega t_0) \right] \\ &= \frac{2L_x}{c} + \frac{h_0 L_x}{c} \frac{\sin(\omega L_x/c)}{\omega L_x/c} \cos\left(\omega(t_0 + L_x/c)\right) \\ &\approx \frac{2L_x}{c} + \frac{L_x}{c} h_+(t_0 + L_x/c) \end{aligned} \quad (4.32)$$

其中，第二行运用了和差化积公式，第三行在低频近似（也称长波近似）下成立，即引力波频率 $\omega \ll 2\pi c/L_x$ 。

类似地，在 $y$ 方向上，可以得到

$$t'_2 - t_0 \approx \frac{2L_y}{c} - \frac{L_y}{c} h_+(t_0 + L_y/c) \quad (4.33)$$

不难计算出，当假设等臂长 $L_x = L_y = L$ 时，两束光干涉会产生相位差

$$\begin{aligned} \Delta\phi_{\text{Mich}}(t) &= \Delta\phi_x(t) - \Delta\phi_y(t) = 2\Delta\phi_x(t) \\ &\approx 2h_+(t_0 + L/c)k_L L \end{aligned} \quad (4.34)$$

其中  $k_L$  是激光波数  $k_L = 2\pi/\lambda$ 。由于引力波探测器记录的光强信息对应于激光在分束器干涉时的强度，通常会将  $t_2$  选为时间变量，简化为  $t$ ，有  $t_1 = t_0 + L/c = t - L/c$ ，因此有

$$\Delta\phi_{\text{Mich}}(t) = 2\Delta\phi_x(t) = 2h_+(t - L/c)k_L L \quad (4.35)$$

注意到，从数学形式上看，上述相位差等价于前述的物理图景：将激光作为长度不变的标尺，测量时刻变化的臂长  $\Delta L = \Delta\phi_x/2k_L$ 。由此我们也可以将臂长变化与引力波强度建立联系：当引力波在  $x$  方向和  $y$  方向分别引起臂长变化  $\Delta L_x$  和  $\Delta L_y$  时，可以得到

$$\frac{\Delta(L_x - L_y)}{L} \simeq h_+(t - L/c). \quad (4.36)$$

另一个不是那么显然的问题是：通常预期的引力波强度弱于  $10^{-20}$ ，即使考虑 km 级别的基线长度，对应的臂长变化量也很容易算出为  $\Delta L = L \times h \sim 10^3 m \times 10^{-20} = 10^{-17} m$ 。与此同时，通常用于激光干涉的激光波长都在  $10^{-6} m$  量级，也就是说，直观上来看，要探测到引力波信号，必须对  $10^{-17} m / 10^{-6} m = 10^{-11} rad$  的相位差实现探测，这一要求似乎太过夸张。Kip Thorne 早年不相信可以通过激光干涉方法测量引力波信号，正是因为这一量级估算的结果。事实真的如此么？

为了能使LIGO这样的引力波探测器可以运作，探测信号，需要通过几个不同的操作提高灵敏度。首先，通过采用 Fabry-Pérot cavity 的形式，可以使光在臂长中往返穿梭多次，人为提高等效臂长。考虑若激光在单个干涉臂中往返  $\sim 10^3$  次，则等效臂长变为  $\sim 10^6 m$ ，而对于激光干涉相位测量精度的要求下降为  $10^6 m \times 10^{-20} / 10^{-6} m = 10^{-8} rad$ 。对于激光干涉的引力波探测器而言，对波长约等于等效干涉臂长的引力波具有最强的响应，因此对应的探测频段约为  $f \sim 300 Hz$ 。

$10^{-8} rad$  的相位精度要求是否可实现？对激光干涉相位测量的主要噪声贡献是散粒噪声 (shot noise)，这是由于恒定光源射出的激光在单位时间内射出光子数符合 Poisson 分布。在预期有  $N$  个光子数时，具体光子数的标准差  $\sigma = \sqrt{N}$ 。当由臂长变化引起的相位差导致光源流量变化大于 Poisson 分布对应的随机起伏时，即认为对应的臂长变化幅度是引力波探测器的灵敏度。

$$\Delta\phi = \Delta l / \lambda \sim \frac{\sqrt{N}}{N} \quad (4.37)$$

其中  $\lambda$  是干涉仪采用的激光波长。考虑到单个光子的能量为  $hc/\lambda$ ，而采样频率由最敏感频段  $f \sim 300 Hz$  代入，对应的单次积分时长约为  $\tau \sim 1/f$ ，可以计算出光子数

$$N = \frac{P}{hc/\lambda} \tau \sim \frac{P}{hc/\lambda} \frac{1}{f} \quad (4.38)$$

其中  $P$  是激光功率。代入典型数值激光功率  $P = 1W$ , 波长  $\lambda = 1\mu m$ , 引力波频率  $f \sim 300Hz$ , 则有光子数  $N \sim 10^{16}$ , 对应的有, 相位测量精度约为  $\Delta\phi \sim \frac{1}{\sqrt{N}} \sim 10^{-8} rad$ 。这一精度刚好满足我们对于测量  $h \sim 10^{-20}$  的引力波的要求。

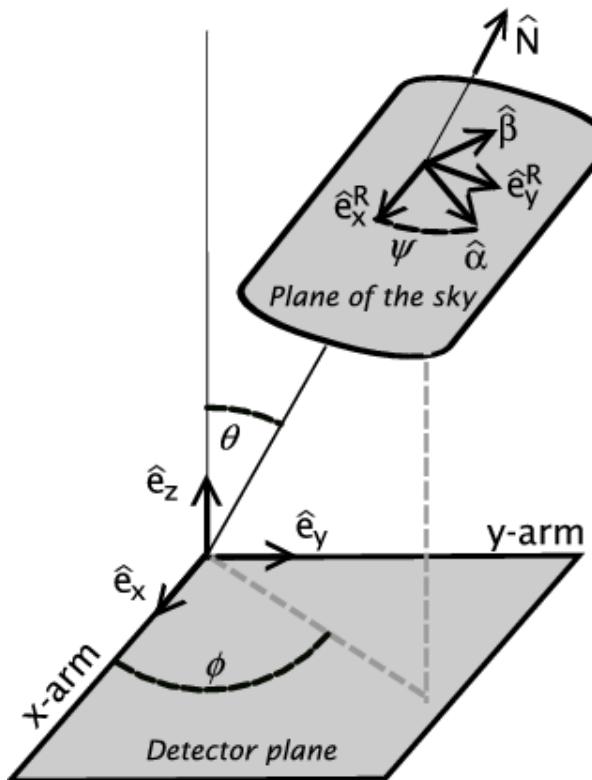


图 4-5 类LIGO的激光干涉引力波探测器下源坐标系与探测器坐标系的相互关系。

Figure 4-5 Relative orientation between source frame and detector frame for a LIGO-like interferometer.

Credit: [84]

可以注意到, 高新 LIGO 的精度比起  $10^{-20}$  更高, 而计划中的第三代引力波探测器, 如 Einstein telescope 或 Cosmic Explorer 等, 都将获得更高的探测灵敏度。这一高精度又是如何获得的呢? 首先, 可以提高激光功率。在探测过程中, 光电探测器放置位置为暗端, 激光经过两条干涉臂终端的镜子反射后在此处形成相消干涉。除了在光学界面中的耗散外, 绝大部分激光功率均由另一面(激光入射方向)逃逸。可以通过放置功率回收镜 (power recycling mirror), 将外溢的激光重新注入 FP 腔, 从而提高等效激光功率。当然, 激光功率的提高会增大辐射压噪声, 综合考虑下, 功率回收装置可以将灵敏度提高约一个量级。

在探测端放置一个半透镜, 并小心调节位置, 可以对某些频段有针对性地提高灵敏度。这样的镜子被称为信号回收镜 (signal recycling mirror), 这一灵敏度的提升是反直

觉的，但是在光学平台的实验证明了这一机制的可靠性。类似收音机中的调频操作，信号回收镜放置位置的不同会改变整体敏感频率的取值。

另外，标准量子极限是激光干涉引力波探测器的重要噪声来源。通过使用压缩光技术，可以在感兴趣的位置方向上提高精度，将量子不确定度压缩至不感兴趣的共轭物理量上。

如图4-5所示，在固有探测器系下，考虑长波近似， $x$  轴方向干涉臂两端的两面镜子之间的距离  $\xi_x$  与 + 模引力波  $h_+$  的关系为  $\ddot{\xi}_x = \frac{1}{2}\ddot{h}_+\xi_x$ 。类似地，在  $y$  轴方向上有  $\ddot{\xi}_y = -\frac{1}{2}\ddot{h}_+\xi_y$ 。联系  $h(t) = D^{ij}h_{ij}(t)$  (公式4.7)，可以得到

$$D^{ij} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (4.39)$$

在更一般的情况下，有

$$D^{ij} = \frac{1}{2}(\mathbf{e}_x^i \mathbf{e}_x^j - \mathbf{e}_y^i \mathbf{e}_y^j) \quad (4.40)$$

其中， $\mathbf{e}^x$  与  $\mathbf{e}^y$  分别为两条干涉臂方向的单位矢量。

类似于第4.1.1节，按照两个偏振模式进行分解，就有

$$\begin{aligned} F_+(\theta, \phi, \psi) &= D_{+}^{ij} e_{ij}^+ = \frac{1}{2}(1 + \cos^2 \theta) \cos 2\phi \cos 2\psi - \cos \theta \sin 2\phi \sin 2\psi \\ F_\times(\theta, \phi, \psi) &= D_{\times}^{ij} e_{ij}^\times = \frac{1}{2}(1 + \cos^2 \theta) \cos 2\phi \sin 2\psi + \cos \theta \sin 2\phi \cos 2\psi \end{aligned} \quad (4.41)$$

## 4.2.2 噪声

如前所述，由于相位测量实际是通过光子数计数测量转换而来的，因此激光测量过程不可避免地会引入散粒噪声。注意到，在公式4.37中，当给定采样频率  $f$  后，时间域上的散粒噪声与引力波频率  $f_{\text{GW}}$  无关。但由于地面探测器引入了 FP 腔以提高灵敏度，但 FP 腔对高频信号的响应会减弱，对应的在高频处，由 FP 腔导致的噪声会占据主导。

除了散粒噪声外，由激光引起的噪声还包括辐射压 (radiation pressure) 噪声。考虑以功率  $I$  入射到镜面的激光，由辐射压引起的力为

$$F_{\text{rad}} = \frac{2I}{c} \quad (4.42)$$

由于在时间域上有  $d^2x/dt^2 = F_{\text{rad}}/M$ ，因此有频率域上  $-4\pi^2 f^2 \tilde{x} = \tilde{F}_{\text{rad}}/M$ ，其中  $M$  是镜子的质量，而  $\tilde{g}$  是时间域函数  $g(t)$  的 Fourier 变换。描述镜子位置不确定度的功率谱密度  $S_x(f)$  可以用激光强度不确定度的功率谱密度  $S_I(f)$  描述

$$S_x(f) = \frac{1}{16\pi^4 f^4 M^2} \frac{4}{c^2} S_I(f) \quad (4.43)$$

功率谱密度是不同频率处的噪声的方差，其具体定义将在第6.2.4章中给出。而引力波测量噪声与镜子位置不确定度的关系为

$$S_h(f) = \frac{8}{L^2} S_x(f) \quad (4.44)$$

由于在 FP 腔中，起点和终点的镜子耦合在一起，两者的噪声强相关，FP 腔的整体噪声就是单面镜子噪声的两倍，在功率谱密度下则是四倍。又因为 x 轴方向和 y 轴方向的 FP 腔不相干，两者的功率谱的叠加符合线性叠加，因此总体干涉仪的功率谱是单个镜子功率谱密度的 8 倍。注意，由于辐射压噪声与频率成反比，所以在低频时辐射压噪声会占据主导。

当功率  $P$  增加时，散粒噪声降低，但是辐射压噪声会抬升；而功率减小时相反，辐射压噪声降低，但散粒噪声太高，综合而言量子噪声在低频处有一个所谓标准量子极限，standard quantum limit (SQL)。只有通过压缩量子态的压缩光 (squeezed light) 才可以突破这一标准量子极限。

与激光相关的噪声，在不同的文献中有不同的表述，有些文献将其统称为读取噪声，有些文献则统称为量子噪声。与之对应的是环境噪声，这其中主要包含**地震噪声**、**热噪声**和**引力梯度噪声**。

**地震噪声**是指由于地震和人类活动等过程产生的地面震动而引起的镜子位置随机变化。地震噪声在主要体现在低频， $S_{h,seis}^{1/2} \sim 10^{-12} Hz^{-1/2} \left( \frac{10Hz}{f} \right)$  而如果不对地震噪声做任何限制，则其影响将远大于量子噪声。可以采用在地面多级悬挂减震的方法，将地震噪声的影响降低。其原理是，不同频率的震动通过单摆系统后，在频率域上的响应在高频是会快速衰减的。若采用多级悬挂，可以加快衰减速率，当然，其代价就是在低频部分噪声会累加放大。地震噪声在  $10Hz$  以下占主导。

**热噪声**在约  $40Hz$  处开始占据主导。这一噪声主要来源于镜子表面和光学器件悬挂器件的分子 Brownian 运动。因此，计划中的下一代地面引力波探测器计划在低温环境中运行，以降低由热噪声带来的影响。

**引力梯度噪声**主要是指由于探测器周围质量分布发生变化，导致引力场变化，从而干扰测量的一种噪声。由于这一噪声来源可以归结为 Newtonian 引力，因此有时也称为 Newtonian 噪声。即使是探测器周围的空气密度发生起伏，也会产生引力梯度噪声。与地震噪声相同，引力梯度噪声也是在低频主导，对于初代引力波探测器而言，引力梯度噪声并非主导因素，但高新 LIGO 等下一代探测器则需要考虑引力梯度噪声的影响。然而，由于引力作用不需要借助地面等媒介进行传导，所以无法通过悬挂减震等技术对其进行压制。下一代地面引力波探测器计划在地下进行试验，通过山体等大量物质使得噪声源远离探测器，从而实现对引力梯度噪声的降低。

日本的 KAGRA 由于采用了低温测量，并在山洞中进行探测，视为地下探测，因此被称为“2.5”代探测器。

### 4.2.3 灵敏度

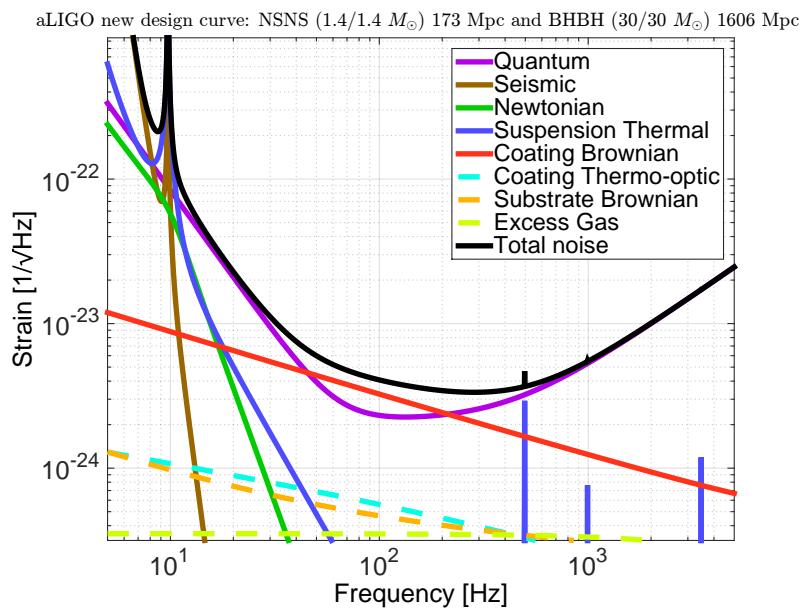


图 4-6 高新 LIGO 的预计振幅功率谱及不同噪声源贡献。

Figure 4-6 Expected amplitude spectrum density and contribution from different sources for advanced LIGO. Credit: [63]

综合考虑各种不同噪声来源后，可以得到高新LIGO等地面引力波探测器的灵敏度曲线。通常由功率谱密度 (power spectrum density, PSD)，或者其平方根振幅谱密度 (amplitude spectrum density, ASD) 表示。在不同时期，由于技术升级，灵敏度不断提升。较为常用的、包含不同时期灵敏度的 ASD 可以参考 [6] 中的图 1，其数据可以从响应的 DCC 页面中获得 [62]。

关于高新 LIGO 升级以后的最终灵敏度，可以参考 LIGO 相关技术文档 [63]，具体的噪声构成如图4-6。

感兴趣的同学还可以下载免费的游戏“Space-Time Quest”，直观感受不同的实验设计对最终噪声曲线的改变（以及对应的经费成本）[58]。该游戏可以在苹果 App Store 或安卓的 Google Play 上下载，也可以下载 Mac、Windows 或 Linux 版。

## 4.3 空间引力波探测

### 4.3.1 原理

在地面，通过激光干涉，结合 F-P 腔，可以探测相对高频的引力波信号。LIGO/Virgo 的敏感频段在约 10-1000Hz 范围内，讨论中的下一代升级可能向更低频或更高频延拓，但一般不会有量级上的差别。根据前述，只有引力波波长约等于等效干涉臂长的引力波才可以最有效的被探测器探测。如果要测量赫兹频段以下的引力波信号，可以计算其等效臂长必须  $\gtrsim$  光秒，长于 10 万公里的量级。这一尺度已经超出地球的尺度，由此，中低频的引力波信号探测，通常均考虑在空间中实施。如图 4-7 所示，天琴计划就计划通过发射三颗卫星，在空间形成编队，实现在空间通过激光干涉测量引力波信号的目标，我们将这种测量手段称为空间引力波探测。

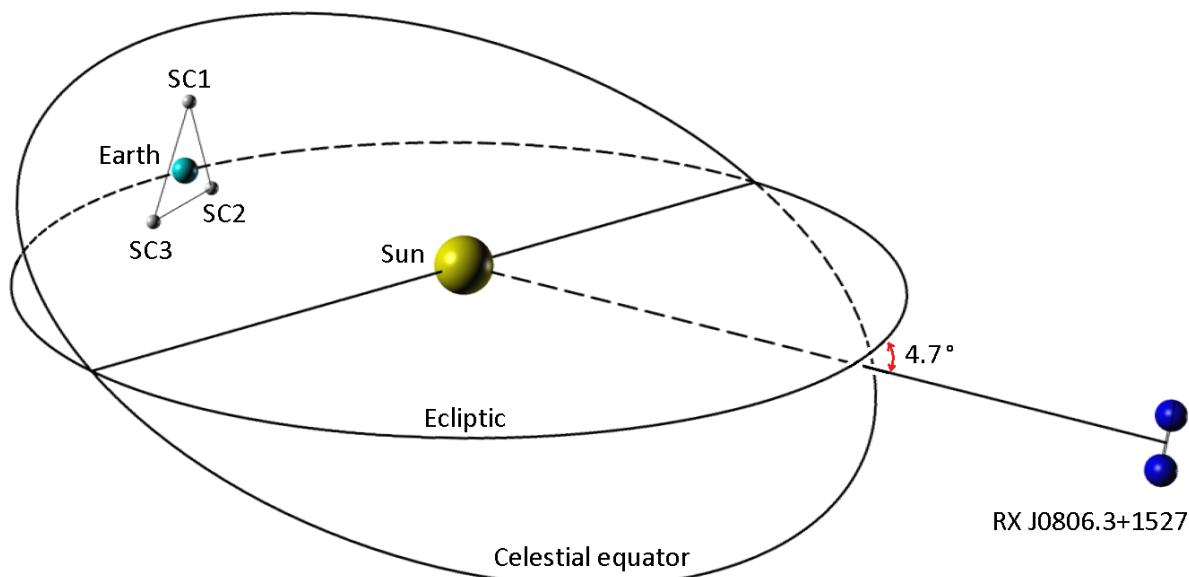


图 4-7 天琴示意图。

Figure 4-7 Illustration of TianQin. Credit: [67]

在空间测量引力波，首先需要一个只受引力作用的基准，再对这样的基准进行精密的测量，从而反推出引力波的作用。为此，以天琴为例的空间引力波探测器，将在卫星中放置完全悬浮的参考物体，也称为检验质量。为了让这些检验质量只受引力作用，而不受磁场、太阳风、太阳光压等效应的干扰，需要对检验质量进行保护，将其与外界环境隔绝，并通过无拖曳控制（drag-free control）方法，尽量使检验质量处于自由落体状态。具体来说，就是在敏感方向不做任何控制，让卫星平台完全跟随检验质量的运动；

另一方面，在非敏感方向，对检验质量做控制，使其保持在一定范围内。在天琴团队中，由**惯性基准组**和**无拖曳控制组**一起负责攻克相关的技术难关。

有了一组在敏感方向处于自由落体状态的检验质量，还需要对他们之间的距离进行精确测量的手段才可以。引力波测量需要测量臂长  $L$  受引力波影响而引起的微小臂长变化  $\Delta L$ ，且有  $h \sim \Delta L/L$ 。如果将目标的  $10^{-20}$  的引力波强度和  $L \sim 10^8\text{m}$  的臂长代入，那么需要对距离测量精度高至  $10^{-12}\text{m}$ ，或皮米(pm)的水准。这里需要注意的是，与地面引力波探测类似，引力波探测过程中，对于绝对距离的测量精度要求并不过分，上述距离测量精度的要求是针对相对测距精度提出的。考虑接收端激光光强  $100\text{nW}$ ，采样频率  $1\text{Hz}$ ，可以发现，由 Poisson 涨落导致的相位测量精度约为  $10^{-6}\text{rad}$ ，再叠加上光波波长  $1\mu\text{m}$ ，可以最终实现  $1\text{pm}$  的相对测距精度。在天琴团队中，由**地月测距组**和**空间光学组**负责攻克相关的技术难关。

最终，**系统技术组**还需要对整体任务实施保障，确保温度、磁场、结构等各方面因素不会影响最终引力波测量目的。尤其是，在传统的卫星项目中（例如天琴一号），科学载荷与卫星平台相互分离，两者有明确的界面。然而对于天琴这样的空间引力波探测项目而言，由于其对测量精度提出了非常严苛的指标要求，卫星平台的各方面因素都有可能会耦合进来，引入误差，因此，载荷与平台的界面相对变得模糊，甚至可以将整颗卫星都视作科学载荷。与卫星平台相关的光机热相关的技术挑战由**光机热组**研究、实现。

我们在第4.2节里提到，在计算公式4.32时，为简化计算，可以采用长波近似， $\omega \ll 2\pi c/L_x$ 。这对于地面探测器而言是可行的。而空间引力波探测器无法利用 F-P 腔延长等效臂长，因此在高频部分，长波近似失效。

我们回顾，在第2.3.2节中，我们讨论了在引力波背景下传播的电磁波，其 Doppler 频移量为

$$z = \frac{\nu_{\text{em}} - \nu_{\text{rec}}}{\nu_{\text{rec}}} \approx \frac{1}{2} \frac{p^j p^k}{1 + n_i p^i} [h_{jk}^{\text{TT}}(t_{\text{rec}}, x_{\text{rec}}^l) - h_{jk}^{\text{TT}}(t_{\text{em}}, x_{\text{em}}^l)] \quad (2.47)$$

在空间引力波探测过程中，需要将激光从一颗卫星中发出，另一颗卫星接收到信号后通过弱光锁相放大，并发射回来。沿用在第4.2.1节中的描述，我们将出射时刻用 0 标记，抵达远端时刻用 1 标记，返回原卫星时刻用 2 标记。记录到的引力波信号为

$$h(t) = \frac{1}{2} \hat{p}^i \hat{p}^j h_{ij}(t) \quad (4.45)$$

则有

$$\frac{\nu_0 - \nu_1}{\nu_1} = \frac{1}{2} \frac{p^i p^j}{1 - \mathbf{n} \cdot \mathbf{p}} [h_{ij}^1 - h_{ij}^0] \quad (4.46)$$

注意，在公式2.47中，我们假定接收者到发射者为  $\mathbf{p}$ 。而在下文中我们都将  $\mathbf{p}$  记作作分

束器至终端的单位矢量。相似地，有

$$\frac{\nu_1 - \nu_2}{\nu_2} = \frac{1}{2} \frac{p^i p^j}{1 + \mathbf{n} \cdot \mathbf{p}} [h_{ij}^2 - h_{ij}^1] \quad (4.47)$$

由于我们考虑引力波扰动很小，因此可以得到

$$\frac{\nu_0 - \nu_2}{\nu_0} = \frac{1}{2} p^i p^j \left( \frac{h_{ij}^2 - h_{ij}^1}{1 + \mathbf{n} \cdot \mathbf{p}} + \frac{h_{ij}^1 - h_{ij}^0}{1 - \mathbf{n} \cdot \mathbf{p}} \right) \quad (4.48)$$

注意，其中，

$$h_{ij}^0 = h_{ij}(t) \quad (4.49)$$

$$h_{ij}^1 = h_{ij} \left[ t + \frac{L}{c} (1 - \mathbf{n} \cdot \mathbf{p}) \right] \quad (4.50)$$

$$h_{ij}^2 = h_{ij} \left( t + \frac{2L}{c} \right) \quad (4.51)$$

对应地，转换到频率域后，可以得到

$$\tilde{h}_{ij}^0 = \tilde{h}_{ij}(f) \quad (4.52)$$

$$\tilde{h}_{ij}^1 = e^{i2\pi f L(1-\mathbf{n}\cdot\mathbf{p})/c} \tilde{h}_{ij}(f) \quad (4.53)$$

$$\tilde{h}_{ij}^2 = e^{i4\pi f L/c} \tilde{h}_{ij}(f) \quad (4.54)$$

可以得到

$$\widetilde{\frac{\nu_0 - \nu_2}{\nu_0}}(f) = \frac{1}{2} p^i p^j \tilde{h}_{ij}(f) \left( \frac{e^{i4\pi f L/c} - e^{i2\pi f L(1-\mathbf{n}\cdot\mathbf{p})/c}}{1 + \mathbf{n} \cdot \mathbf{p}} + \frac{e^{i2\pi f L(1-\mathbf{n}\cdot\mathbf{p})/c} - 1}{1 - \mathbf{n} \cdot \mathbf{p}} \right) \quad (4.55)$$

这一 Doppler 频移会导致相位偏移

$$\phi = 2\pi \int dt [\nu_0 - \nu_2(t)] \quad (4.56)$$

到了频率域，这一相移为

$$\tilde{\phi}(f) = \frac{(\widetilde{\frac{\nu_0 - \nu_2}{\nu_0}})(f)}{if} \quad (4.57)$$

整理后可得

$$\tilde{\phi}(f) = k L p^i p^j \tilde{h}_{ij}(f) \mathcal{T}(\mathbf{n} \cdot \mathbf{p}, fL/c) \quad (4.58)$$

其中，

$$\begin{aligned} \mathcal{T}(\mathbf{n} \cdot \mathbf{p}, fL/c) &= \frac{1}{2} e^{i2\pi f L/c} \left\{ e^{i\pi f L(1+\mathbf{n}\cdot\mathbf{p})/c} \text{sinc}[\pi f L(1+\mathbf{n}\cdot\mathbf{p})/c] \right. \\ &\quad \left. + e^{-i\pi f L(1+\mathbf{n}\cdot\mathbf{p})/c} \text{sinc}[\pi f L(1-\mathbf{n}\cdot\mathbf{p})/c] \right\} \end{aligned} \quad (4.59)$$

称为传递函数 (**transfer function**)。注意到当  $2\pi f L/c \ll 1$  时, 或者  $f \ll f_* = c/2\pi L$  时, 传递函数近似为 1。

不管是天琴还是 LISA, 空间引力波探测器通常采用的构型都是三角形型的, 如果每两颗卫星之间都有激光相连, 可以构成 3 个干涉仪。对于每一个干涉仪, 都可以相应地对写出 Michelson 构型下的探测器张量

$$\mathbf{D}^{ij}(\hat{\mathbf{n}}, f) = \frac{1}{2} \left( \mathbf{e}_1^i \mathbf{e}_1^j \mathcal{T}(\mathbf{e}_1 \cdot \hat{\mathbf{n}}, f) - \mathbf{e}_2^i \mathbf{e}_2^j \mathcal{T}(\mathbf{e}_2 \cdot \hat{\mathbf{n}}, f) \right) \quad (4.60)$$

像天琴或 LISA 这样的空间引力波探测器, 由三个探测臂 (来回六条激光链路) 构成等边三角形。这种构型可以组合出两个线性独立的光程差测量值  $2(\delta L_1(t) - \delta L_2(t))$  和  $2(\delta L_2(t) - \delta L_3(t))$ 。对应两个测量信号  $s_I(t) = 2(\delta L_1(t) - \delta L_2(t))/L$  和  $s'_{II}(t) = 2(\delta L_2(t) - \delta L_3(t))/L$ 。

第一个探测器接收到的信号

$$\begin{aligned} h_I(t) = s_I(t) &= 2[\delta L_1(t) - \delta L_2(t)]/L \\ &= 2 \frac{1}{2} h_{ab}(t)(l_1^a l_1^b - l_2^a l_2^b) \end{aligned} \quad (4.61)$$

$$= \frac{\sqrt{3}}{2} (h_{xx} - h_{yy}). \quad (4.62)$$

$$= \frac{\sqrt{3}}{2} h_+(t). \quad (4.63)$$

从公式4.61到公式4.63用到了

$$l_i^a = \cos \gamma_i x^a + \sin \gamma_i y^a, \quad \gamma_i = \pi/12 + (i-1)\pi/3. \quad (4.64)$$

$l_i^a$  是沿探测臂的单位矢量,  $x^a$  和  $y^a$  分别是探测器坐标下沿 x 轴和 y 轴的单位矢量, 见图 4-8。

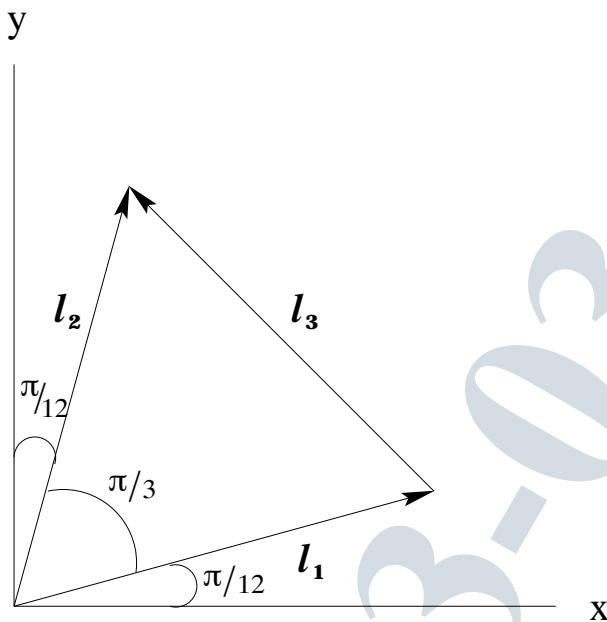
$s'_{II}(t)$  与第一个干涉仪虽然线性独立, 但两者并不正交。定义第  $i$  条臂中的噪声  $n_i = \delta L_i(t)/L$ , 则有

$$\langle n_i(f) n_j(f') \rangle = C_{ij}(f) \delta(f - f'). \quad (4.65)$$

其中  $\langle \cdots \rangle$  是统计学期望。通过计算, 可以发现,  $\langle s_I(f) s_I(f) \rangle = C_{11}(f) - 2C_{12}(f) + C_{22}(f)$ 。此外,  $\langle s_I(f) s'_{II}(f) \rangle = C_{12}(f) - C_{13}(f) - C_{22}(f) + C_{23}(f)$ 。

因此, 在  $s'_{II}(f)$  中扣除与  $s_I(f)$  相关的项, 即定义

$$s_{II}(f) \propto s'_{II}(f) - s_I(f) \frac{C_{12}(f) - C_{13}(f) - C_{22}(f) + C_{23}(f)}{C_{11}(f) - 2C_{12}(f) + C_{22}(f)}. \quad (4.66)$$

图 4-8  $x - y$  平面上探测器 3 条臂的指向Figure 4-8 Direction of 3 arms in the  $x - y$  plane. Credit:[38]

即可得到与  $s_I(f)$  正交的信号通道。在没有进一步实验数据的前提下，可以采用完全对称的假设，并要求归一化，有

$$s_{II}(t) = \frac{2}{\sqrt{3}} (\delta L_1(t) + \delta L_2(t) - 2\delta L_3(t)) \quad (4.67)$$

在这种构型下，其接收到的信号

$$\begin{aligned} h_{II}(t) &= \frac{\sqrt{3}}{2} (h_{xy} + h_{yx}) \\ &= \frac{\sqrt{3}}{2} h_x \end{aligned} \quad (4.68)$$

由公式4.63和公式4.68可知  $60^\circ$  角的干涉仪接受到的振幅是  $90^\circ$  角的干涉仪接受到的振幅的  $\sin(60^\circ) = \sqrt{3}/2$  倍。

对于天琴而言，传递频率 (transfer frequency)  $f_* = c/2\pi L = 0.28\text{Hz}$ 。在低频极限下，探测器响应后的引力波信号由下式给出

$$h_\alpha(t) = \frac{\sqrt{3}}{2} h_+(t) F^+(t) + \frac{\sqrt{3}}{2} h_x(t) F^x(t), \alpha = I, II \quad (4.69)$$

其天线响应函数与地面探测器的一致 (公式4.41)。而对于构建出的第二个干涉仪，其天

线响应函数可以由第一个探测器旋转  $\pi/4$  后得到。换言之，两者正交。

$$F_{\text{II}}^+(t, \theta_S, \phi_S, \psi_S) = F_{\text{I}}^+(t, \theta_S, \phi_S - \frac{\pi}{4}, \psi_S), \quad (4.70)$$

$$F_{\text{II}}^\times(t, \theta_S, \phi_S, \psi_S) = F_{\text{I}}^\times(t, \theta_S, \phi_S - \frac{\pi}{4}, \psi_S), \quad (4.71)$$

也可以把  $60^\circ$  角的干涉仪到  $90^\circ$  角的干涉仪响应差异的因子  $\sqrt{3}/2$  与天线响应函数结合在一起，由此上式改写为

$$h_\alpha(t) = h_+(t)F^+(t) + h_\times(t)F^\times(t), \quad (4.72)$$

此时其中的天线响应函数可改写成

$$\begin{aligned} F_{\text{I}}^+(t, \theta_S, \phi_S, \psi_S) &= \frac{\sqrt{3}}{2} \left( \frac{1}{2}(1 + \cos^2 \theta_S) \cos 2\phi_S(t) \cos 2\psi_S \right. \\ &\quad \left. - \cos \theta_S \sin 2\phi_S(t) \sin 2\psi_S \right), \end{aligned} \quad (4.73)$$

$$\begin{aligned} F_{\text{I}}^\times(t, \theta_S, \phi_S, \psi_S) &= \frac{\sqrt{3}}{2} \left( \frac{1}{2}(1 + \cos^2 \theta_S) \cos 2\phi_S(t) \sin 2\psi_S \right. \\ &\quad \left. + \cos \theta_S \sin 2\phi_S(t) \cos 2\psi_S \right), \end{aligned} \quad (4.74)$$

### 4.3.2 噪声

如表4-1所示，天琴对诸多关键参数提出了明确指标。天琴初步方案要求，在天琴的敏感频段内，即毫赫兹频段内，检验质量上的残余扰动加速度不超过  $10^{-15}\text{m/s}^2/\text{Hz}^{1/2}$  量级，约 17 万公里基线上的激光位移测量误差不超过  $10^{-12}\text{m}/\text{Hz}^{1/2}$  量级。在卫星平台方面，天琴初步方案要求，在天琴的敏感频段内，即毫赫兹频段内，卫星内部核心载荷区域的温度稳定性达到  $10^{-4}\text{K}$  量级，部分核心区域达到  $10^{-6}\text{K}$  量级，用于卫星无拖曳控制的微推进器的推力分辨率应达到  $0.1\text{N}/\text{Hz}^{1/2}$  量级、推力范围达到  $1\text{-}100\mu\text{N}$ ，寿命应长达若干年 [67]。

对空间引力波探测器而言，总的噪声功率谱密度为位置噪声  $S_n^x(f)$  和加速度噪声  $S_n^a(f)$  的叠加：

$$P_n = \left[ S_n^a \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_n^x \right]. \quad (4.75)$$

同时将考虑传递函数以后的噪声曲线定义为灵敏度曲线，符号是  $S_n$ 。在上式中， $\left( 1 + \frac{10^{-4}\text{Hz}}{f} \right)$  是由惯性传感器引起的  $1/f$  噪声。对应的  $\sqrt{P_n}$  是振幅谱密度。 $S_n^a$  是加速度噪声的 PSD， $S_n^x$  是位置噪声的 PSD。以上噪声可以转化为相位噪声，用相位计测量。

参数	取值
卫星数目	$N = 3$
星座构型	等边三角形
轨道类型	地心轨道
臂长	$L \sim 10^5 \text{ km}$
位置测量精度	$1 \text{ pm}/\text{Hz}^{1/2} @ 6 \text{ mHz}$
残余加速度精度	$10^{-15} \text{ m/s}^2/\text{Hz}^{1/2} @ 6 \text{ mHz}$
观测窗口	每年 $2 \times (3 \text{ 个月})$
激光波长	$\lambda = 1064 \text{ nm}$
光功率	$P_{\text{out}} = 4 \text{ W}$
望远镜口径	$D = 20 \text{ cm}$
光学效率	$\eta_{\text{opt}} = 70\%$

表 4-1 天琴基础技术参数。

Table 4-1 Key parameters for TianQin

考虑位置噪声，每个相位计测量的位置噪声都是不相关的，可以直接把光路中所有的相位计各自测得位置噪声相加以得到总的位置噪声。

$$S_n^x = \frac{4S_x}{(2L)^2} = \frac{S_x}{(L)^2}. \quad (4.76)$$

$S_x$  是单个相位计测量的位置噪声， $L$  是探测器臂长，分母的 2 是由于一个干涉仪的光路由两条臂组成，分子的 4 是由于每条臂都有两个相位计测量的噪声。

对于加速度噪声，除了每个相位计测量的加速度噪声可以直接相加，加速度噪声对入射和出射激光的相干作用也需要考虑。

$$S_n^a = \frac{16S_a}{(2\pi f)^4(2L)^2} = \frac{4S_a}{(2\pi f)^4(L)^2}. \quad (4.77)$$

$S_a$  是单个相位计测量的加速度噪声，干涉仪的加速度噪声  $P_n^a$  推导中与  $P_n^x$  相比分子多乘的 4 是由于加速度噪声对入射和出射激光都有 2 倍的效应

把公式 4.76、4.77) 代入公式 4.75，由此我们得到一个干涉仪总的噪声 PSD

$$\begin{aligned} P_n &= \left[ S_n^a \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_n^x \right] \\ &= \frac{1}{L^2} \left[ \frac{4S_a}{(2\pi f)^4} \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_x \right]. \end{aligned} \quad (4.78)$$

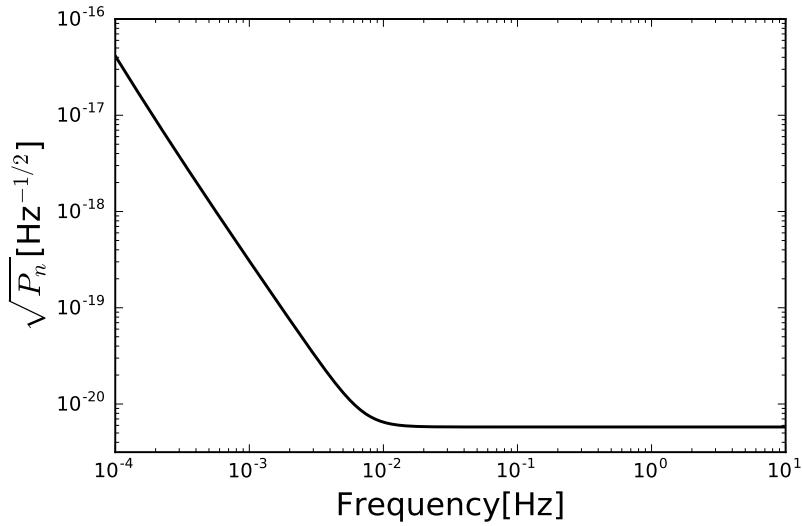


图 4-9 天琴的噪声曲线

Figure 4-9 Noise curve for TianQin

用这个式子代入具体  $S_a$ 、 $S_x$  和  $L$  的值就可以画出相应的噪声曲线，见图4-9。天琴的探测器臂长  $L = \sqrt{3} \times 10^5 \text{ km}$ ，加速度噪声和位置噪声指标分别是  $S_a = 1 \times 10^{-30} \text{ m}^2 \text{s}^{-4} \text{ Hz}^{-1}$  和  $S_x = 1 \times 10^{-24} \text{ m}^2 \text{ Hz}^{-1}$ ，天琴的参数见表4-1。

下面我们考虑对天线响应函数的全天平均。引力波引起的激光信号变化，也可以理解为激光的 Doppler 红移  $z(t, \theta, \psi)$

$$z(t, \theta, \psi) = \frac{\sqrt{T}}{4\pi\tau} \int_{-\infty}^{+\infty} d\omega \cos(2\psi) \tilde{h}(\omega) [(1 - \mu) + 2\mu e^{-i\omega\tau(1+\mu)} - (1 + \mu)e^{-i2\omega\tau}] \frac{1}{i\omega} e^{-i\omega t}, \quad (4.79)$$

其中  $\mu = \cos \theta$ ， $\theta$  是引力波波源位置矢量与探测臂的夹角， $\psi$  是引力波偏振角， $\tau$  是光在探测臂传播的时间。

对于一般的空间引力波探测器，给定引力波源，在探测器坐标下对应的方位和极化角会随着时间变化。定义从引力波振幅到探测器信号的关系为引力波响应函数  $R(\omega)$ ，即

$$S_{\bar{\Delta}}(\omega) = S_h(\omega)R(\omega), \quad (4.80)$$

其中  $\Delta(t) = z_1(t) - z_2(t)$ ， $S_{\bar{\Delta}}(\omega)$  是对源的偏振角和方位角平均的信号的功率谱，由

$$\langle \bar{\Delta}^2 \rangle = \frac{1}{2\pi} \int_0^\infty S_{\bar{\Delta}}(\omega) d\omega = \frac{1}{T} \int_0^\infty |\bar{\Delta}|^2 dt \quad (4.81)$$

定义,  $S_h(\omega)$  是引力波振幅的功率谱  $\langle X \rangle$  表示对时间平均,  $\bar{X}$  表示对源的方位和偏振角平均。利用公式4.79计算得

$$\langle \Delta^2 \rangle = \frac{1}{2\pi} \int_0^\infty d\omega \tilde{h}^2(\omega) \frac{1}{(\omega\tau)^2} [\mathcal{T}_1(\omega) + \mathcal{T}_2(\omega) - 2\mathcal{T}_3(\omega)], \quad (4.82)$$

其中  $\mathcal{T}_1(\omega)$ ,  $\mathcal{T}_2(\omega)$  和  $\mathcal{T}_3(\omega)$  具体表达式为

$$\begin{aligned} \mathcal{T}_1(\omega) &= \cos^2(2\psi_1) [\mu_1^2 (1 + \cos^2(\omega\tau)) + \sin^2(\omega\tau) \\ &\quad - 2\mu_1^2 \cos(\omega\tau) \cos(\omega\tau\mu_1) - 2\mu_1 \sin(\omega\tau) \sin(\omega\tau\mu_1)], \end{aligned} \quad (4.83)$$

$$\begin{aligned} \mathcal{T}_2(\omega) &= \cos^2(2\psi_2) [\mu_2^2 (1 + \cos^2(\omega\tau)) + \sin^2(\omega\tau) \\ &\quad - 2\mu_2^2 \cos(\omega\tau) \cos(\omega\tau\mu_2) - 2\mu_2 \sin(\omega\tau) \sin(\omega\tau\mu_2)], \end{aligned} \quad (4.84)$$

$$\mathcal{T}_3(\omega) = \cos(2\psi_1) \cos(2\psi_2) \eta(\omega), \quad (4.85)$$

其中

$$\begin{aligned} \eta(\omega, \theta_1, \theta_2) &= [\cos(\omega\tau) - \cos(\omega\tau\mu_1)][\cos(\omega\tau) - \cos(\omega\tau\mu_2)]\mu_1\mu_2 \\ &\quad + [\sin(\omega\tau) - \mu_1 \sin(\omega\tau\mu_1)][\sin(\omega\tau) - \mu_2 \sin(\omega\tau\mu_2)], \end{aligned} \quad (4.86)$$

对源的偏振角和方位角做平均用到

$$\overline{\langle \Delta^2 \rangle} = \frac{1}{8\pi^2} \int_0^{2\pi} d\psi \int_0^{2\pi} d\epsilon \int_0^\pi \sin \theta d\theta \langle \Delta^2 \rangle. \quad (4.87)$$

$\epsilon$  是波源位置矢量与探测臂所在平面与探测器平面的倾角。

结合 4.80- 4.87 式, 定义  $\alpha = \psi_2 - \psi_1$  最终可以得到引力波响应函数为

$$\begin{aligned} R(\omega) &= 2 \frac{1}{(\omega\tau)^2} \left[ (1 + \cos^2(\omega\tau)) \left( \frac{1}{3} - \frac{2}{(\omega\tau)^2} \right) + \sin^2(\omega\tau) + \frac{4}{(\omega\tau)^3} \sin(\omega\tau) \cos(\omega\tau) \right. \\ &\quad \left. - \frac{1}{4\pi} \int d\epsilon d\theta_1 \sin \theta_1 (1 - 2 \sin^2 \alpha) \eta(\omega) \right], \end{aligned} \quad (4.88)$$

用数值积分给出的响应  $R(\omega = 2\pi f)$ , 见图4-10。在低频时令  $\omega\tau \rightarrow 0$ , 由公式4.88 可以得到  $R(\omega)$  的低频极限, 低频极限下平均后信号的功率谱

$$S_{\bar{\Delta}} = R(\omega) S_h(\omega) = \frac{4}{5} \sin^2 \gamma S_h, \quad (4.89)$$

其中  $\gamma$  是探测器两臂的张角, 当  $\gamma = \pi/3$  时,  $R_{\omega\tau \rightarrow 0} = 4/5 \times 3/4 = 3/5$ 。可以如此理解这些数字:4 来自单次测量噪声与干涉仪测量噪声的比值, 如公式4.76、4.77等, 即  $S_n^x$  与  $\frac{S_x}{4L^2}$  之间的比值关系; 1/5 来自全天平均,

$$\mathcal{R} = \langle F^+ \rangle^2 = \langle F^\times \rangle^2 = \frac{1}{32} \int (1 + 6x^2 + x^4) dx = \frac{1}{5}$$

$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
1	$\frac{1}{10^4}$	$\frac{2639}{10^4}$	$\frac{231}{5 \times 10^4}$	$-\frac{2093}{1.25 \times 10^4}$	$\frac{2173}{10^5}$
$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$
$\frac{2101}{10^6}$	$\frac{3027}{2 \times 10^5}$	$-\frac{42373}{5 \times 10^6}$	$\frac{176087}{10^8}$	$-\frac{8023}{5 \times 10^7}$	$\frac{5169}{10^9}$

表 4-2 拟合响应函数的公式 4.94 中的多项式系数

Table 4-2 Polynomial parameters from fitting equation 4.94.

$\frac{3}{4} = \sin^2(\pi/3)$  来自三角形构型引力波干涉仪与垂直构型干涉仪在灵敏度上的差别。

根据公式 4.80, 再令  $S_{\bar{\Delta}}$  等于探测器噪声的功率谱密度  $S_N$ , 全天平均的灵敏度曲线 (ASD)

$$h_f = \sqrt{S_h} = \sqrt{\frac{S_{\bar{\Delta}}}{R}} = \sqrt{\frac{S_N}{R}} = \sqrt{4 \frac{P_n}{R}}, \quad (4.90)$$

这其中  $P_n$  为单程光路测量的仪器噪声功率谱密度。考虑到一个干涉仪包含来去双程、两条干涉臂, 因此有  $S_N = 4P_n$  成立。可以认为是消去公式 4.89 中的一个因子 4。

根据公式 4.90 取  $R_{\omega\tau \rightarrow 0} = 3/5$ , 可得低频极限下全天平均的灵敏度曲线

$$S_h^{SA} = \frac{S_N}{R} = \frac{20}{3} P_n. \quad (4.91)$$

低频极限下非全天平均的灵敏度曲线为

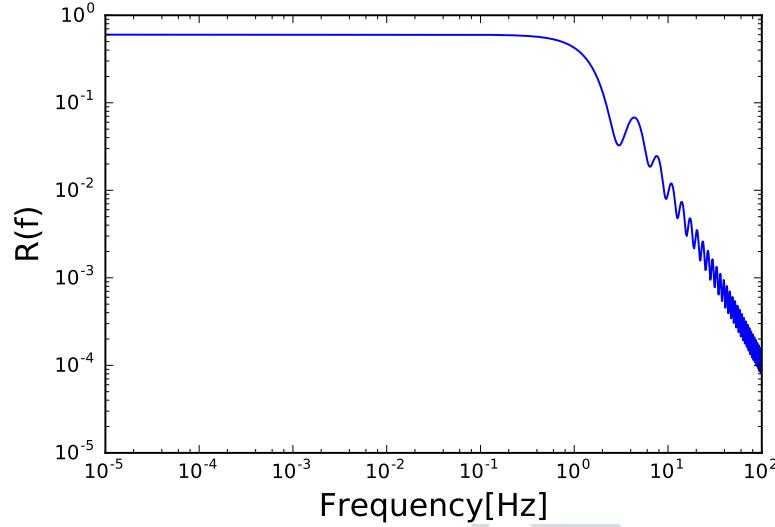
$$S_h^{NSA} = \frac{3}{20} S_h^{SA}. \quad (4.92)$$

我们拟合数值积分后的响应给出了  $R(f)$  的解析表达式

$$R(f) = \frac{3}{5} \times \frac{g(f/f_*)}{1 + 0.6(f/f_*)^2}, \quad (4.93)$$

其中函数  $g(x)$  为

$$g(x) = \begin{cases} \sum_{i=0}^{11} a_i x^i & : x < 4.1, \\ \exp [-0.322 \sin(2x - 4.712) + 0.078] & : 4.1 \leq x < \frac{20\pi}{\sqrt{3}}, \end{cases} \quad (4.94)$$

图 4-10 响应函数  $R(f)$ Figure 4-10 Response curve  $R(f)$ 

其中拟合的二项式系数  $a_i$  的取值见表 4-2。通常在计算时会取响应函数  $R(f)$  的高频部分的领头阶，即

$$R_0 = \frac{3}{5} \times \frac{1}{1 + 0.6(f/f_*)^2}. \quad (4.95)$$

根据公式 4.78、4.90 和 4.95，全天平均的灵敏度曲线为

$$S_h^{SA} = \frac{20}{3} \frac{1}{L^2} \left[ \frac{4S_a}{(2\pi f)^4} \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_x \right] \times \left[ 1 + 0.6 \left( \frac{f}{f_*} \right)^2 \right] \quad (4.96)$$

根据公式 4.96 可以画出天琴全天平均的灵敏度曲线，见图 4-11。

这里得到的全天平均的灵敏度曲线与 [67] 所给的有所不同，文章中公式 2-3 的天琴灵敏度曲线的表达式如下

$$S_h'^{SA} = \frac{4}{(\frac{8}{15})L^2} \left[ \frac{S_a}{(2\pi f)^4} \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_x \right] \times \left[ 1 + \left( \frac{f}{0.41\frac{c}{2L}} \right)^2 \right] \quad (4.97)$$

两者的差别主要在全天平均的因子  $R$  不同。公式 4.97 的系数相当于将  $\frac{20}{3} = 6.67$  写作  $\frac{15}{2} = 7.5$ ，在数值上相差  $1/9$ ，约为 11%。在 ASD 上的差别则为 5.7%。可以发现，公式 4.88 的积分项在低频极限下不为零，可以对  $\eta$  泰勒展开后取线性项求得低频极限就是式 4.89 的  $R = 4 \sin^2 \gamma / 5$ 。但是若把积分项的极限错当成零，则低频极限为  $8/15$ ，而非  $9/15$ 。根据公式 4.97 可以画出 [67] 中天琴全天平均的灵敏度曲线，见图 4-11。

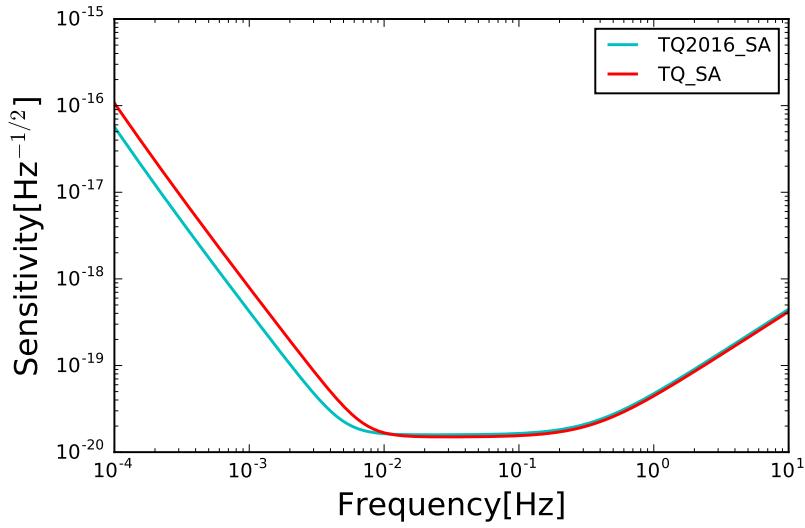


图 4-11 天琴的全天平均的灵敏度曲线。红色是公式 4.96 给出的，浅蓝色来自公式 4.97 给出的灵敏度曲线。

Figure 4-11 Sky-averaged sensitivity curve. Red curve from equation 4.96, blue curve from equation 4.97.

Credit:[67]

对于非全天平均的灵敏度曲线，可以表达为

$$S_h^{NSA} = \frac{3}{20} S_h^{SA} = \frac{1}{L^2} \left[ \frac{4S_a}{(2\pi f)^4} \left( 1 + \frac{10^{-4}\text{Hz}}{f} \right) + S_x \right] \times \left[ 1 + 0.6 \left( \frac{f}{f_*} \right)^2 \right]. \quad (4.98)$$

根据公式 4.98 可以画出天琴非全天平均的灵敏度曲线，见图 4-12 中红色的线，黑色的线用到了 4.93 给出的完整  $R(f)$ 。

值得注意的是，理论上预期，大量的双白矮星引力波信号会形成前景噪声。在上述描述中，我们并没有考虑前景噪声。

## 4.4 脉冲星计时阵

### 4.4.1 原理

脉冲星计时阵列探测引力波是指通过对地球周边一系列脉冲星信号进行高精度测时记录，以分析是否有引力波信号的信息。当前国际上主要的脉冲星计时阵列引力波探测实验主要有：澳大利亚的 PPTA[71]，欧洲的 EPTA[39]，北美的 NANOGrav[21]。我国已经建成的 FAST 也有望在未来的 PTA 引力波探测中发挥作用 [75]。脉冲星计时阵列主

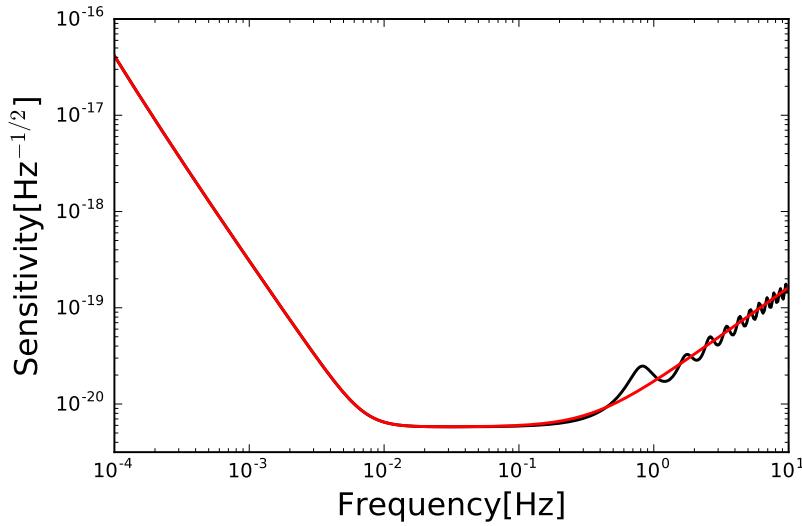


图 4-12 天琴的非全天平均灵敏度曲线

Figure 4-12 Non sky-averaged sensitivity curve for TianQin

要针对纳赫兹 ( $10^{-9}$ Hz) 附近频段的引力波。经过十多年的科学运行，这些现有的实验已经对纳赫兹频段附近的引力波给出了有意义的限制。

脉冲星在长时标上有很高的测时精度，可以将其视为极为稳定的时钟。从原理上来说，PTA 测量引力波的方法类似于 Doppler Tracking，引力波的经过会导致脉冲星辐射脉冲信号的频率产生微小变化

$$\frac{\nu_{\text{em}} - \nu_{\text{rec}}}{\nu_{\text{rec}}} \approx \frac{1}{2} \frac{p^j p^k}{1 + n_i p^i} [h_{jk}^{\text{TT}}(t_{\text{rec}}, x_{\text{rec}}^l) - h_{jk}^{\text{TT}}(t_{\text{em}}, x_{\text{em}}^l)] \quad (2.47)$$

这一频率变化反映到脉冲到达地球的时刻上 (times-of-arrival, ToA)，会产生时间残余

$$R(t) = - \int_0^t \frac{\nu_{\text{em}} - \nu_{\text{rec}}}{\nu_{\text{rec}}} dt \quad (4.99)$$

同时测量多个不同的脉冲星信号，就可以利用不同脉冲星的时间残余直接的相关关系反推出随机引力波背景的相关信息。通常预期 PTA 方法可以测量来自宇宙暴胀、宇宙弦网络、大质量双黑洞并合等过程的随机引力波背景。

2020 年，NANOGrav 团队宣布在 12.5 年的数据中观测到了符合幂率谱分布的随机信号，相比起独立无关的噪声模型，具有共同谱分布的引力波信号可以更好地解释数据。考虑形如  $f^{-2/3}$  的引力波信号，在  $f = 1\text{yr}^{-1}$  频率处，其强度的中值为  $1.92 \times 10^{-15}$  然而，实际数据和理论预期所给出的时间参与相关关系并不吻合，故此暂时还无法确定其天文学起源。[22]

## 4.5 宇宙微波背景辐射

### 4.5.1 原理

宇宙微波背景辐射探测引力波是指通过对宇宙微波背景辐射中光子的 B 模极化的高精度探测，来获取有关原初引力波的信息。美国于 2001 年发射的 WMAP 卫星及欧洲于 2009 年发射的 Planck 卫星已经对原初引力波给出了很好的限制 [91, 14]。位于南极的 BICEP 望远镜于 2006 年开始运行，并于 2016 年升级为 BICEP3[13]。我国正在进行的阿里计划也将通过微波背景辐射观测来探测原初引力波 [59]。宇宙微波背景辐射探测主要针对阿赫兹 ( $10^{-18}\text{Hz}$ ) 到飞赫兹 ( $10^{-15}\text{Hz}$ ) 频段的原初引力波。

## 4.6 其他探测方案

目前，还有一些研究组提出建造小尺度的引力波探测器，探测更高频的引力波信号；或者利用原子干涉技术，实现空间引力波探测。然而，目前这类讨论均处于相对早期阶段。

### Exercise 4.1

请根据公式4.21和公式4.22，推导出旋转  $\psi$  角后  $\times$  模的表达式

### Exercise 4.2

参照图4-5，证明对于激光干涉引力波探测器，在长波极限下的天线响应函数为公式4.41。（提示：先得到  $(\alpha, \beta, N)$  坐标系到  $(x, y, z)$  坐标系的旋转矩阵  $\mathcal{R}$ ，再根据  $h_{ij} = \mathcal{R}_{ik}\mathcal{R}_{jl}h'_{kl}$  分别求出投影到探测器坐标系下两个偏振基底的分量，即为  $F_+$  和  $F_\times$ ，最后再考虑偏振角  $\psi$  的旋转带来的改变，即从  $(x^R, y^R, N)$  坐标系到  $(\alpha, \beta, N)$  坐标系的转变）。

### Exercise 4.3

请根据公式4.41，编写程序作图，画出 LIGO 等地面引力波探测器的天线响应函数。限定极化角  $\psi = 0$ 。

- ▶ 画出曲面，使得在  $(\theta, \phi)$  方向上，与原点的距离为  $|F_+(\theta, \phi)|$ 。
- ▶ 画出曲面，使得在  $(\theta, \phi)$  方向上，与原点的距离为  $|F_\times(\theta, \phi)|$ 。
- ▶ 画出曲面，使得在  $(\theta, \phi)$  方向上，与原点的距离为  $\sqrt{F_+^2(\theta, \phi) + F_\times^2(\theta, \phi)}$ 。

# 5

## “穷人版” 引力波数据处理示例

2015 年 9 月 14 号，LIGO 团队第一次探测到了引力波信号 GW150914，并证实它是来自于组成质量分别为  $36M_{\odot}$  和  $29M_{\odot}$  的恒星级双黑洞并合事件，这也再一次检验了广义相对论的正确性。1915 年，Einstein 创立了广义相对论，并在 1916 年在自己的场方程里面发现了一个波动解，且发现能量以光速传播，以引力波的形式辐射出去。同年，德国物理学家 Karl Schwarzschild 找到了广义相对论引力场方程的解，并在之后被认为是描述黑洞的解。它描述的是无自旋的大质量物体外的一定位置内逃逸速度超过了光速，这也就意味着在处在这个半径内部的任何物质和辐射都将不能溢出，而这个临界距离就叫做 Schwarzschild 半径：

$$R_s(m) = 2Gm/c^2 = 2.95 \left( \frac{m}{M_{\odot}} \right) \text{ km} \quad (5.1)$$

其中  $M_{\odot} = 1.99 \times 10^{30} \text{ kg}$ ,  $G = 6.67 \times 10^{-11} \text{ kg m}^3 \text{ s}^{-2}$ ,  $c = 2.998 \times 10^8 \text{ m/s}$ 。任何物体如果压缩到其 Schwarzschild 半径之内，那它就一定会形成一个黑洞，这个结论用于后面论证 GW150914 信号是来源于恒星级双黑洞并合。对于一个双星系统，若其组成质量分别为  $m_1$  和  $m_2$ ，总质量可记为  $M = m_1 + m_2$ ，质量比  $q = m_1/m_2$ ，定义啁啾质量 chirp mass，

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \quad (5.2)$$

接下来我们将基于探测器采集的数据，通过几个方面来论证 GW150914 的天文学起源是双黑洞并合，并讨论所做假设的合理性以及假设对于结论的影响。

## 5.1 从观测数据内提取信息

引力波观测信号本质上是大质量物体相互旋转所辐射出的引力波，反映到探测器激光回路里就是产生光程差，被记录下来的就是仪器观测到的应变数据  $h(t)$ 。为了处理数据，首先我们从 Gravitational Wave Open Science Center<sup>1</sup> 下载了 GW150914 的 strain data。

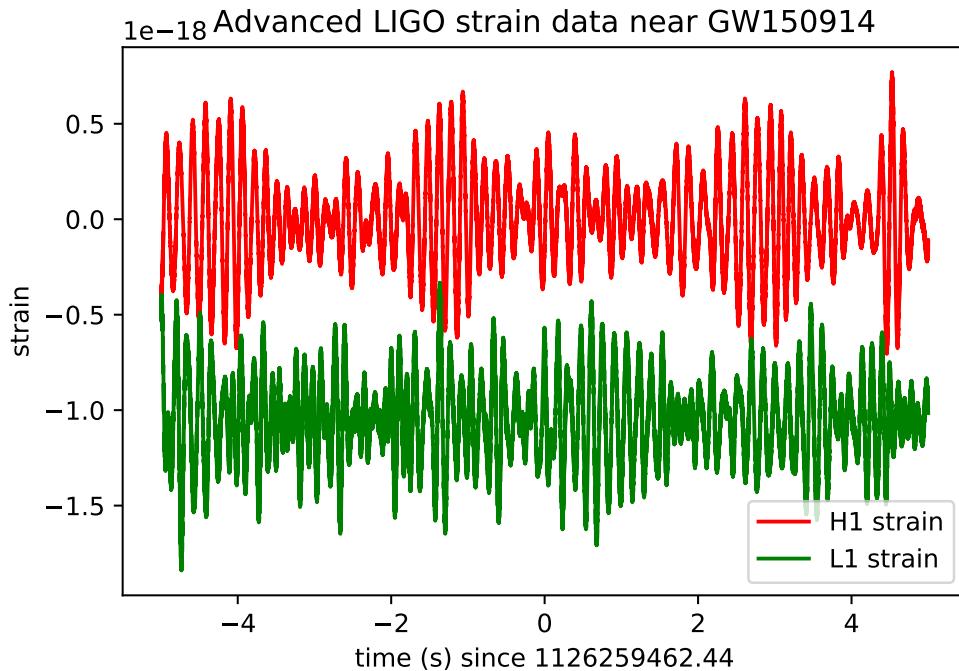


图 5-1 GW150914 事件点左右 5 秒内的应变数据，红色为 Hanford 的数据，绿色为 Livingston 的数据。

Figure 5-1 Strain data around GW150914.

绘制出了在事件发生前后 5s 的信号，如图 5-1 所示。图中的信号主要是由低频噪声占据主导，需要一些数据处理方法才能找到引力波信号。H1 的数据的中间值相比于 L1 数据有大约  $-2.0 \times 10^{-18}$  的偏离，但是经过数据处理（例如 bandpass）之后，该偏移可以消除。

### 5.1.1 白化 (Whitening)

数据处理的第一步通常是白化操作，因而对于图 5-1 里面的信号，我们首先白化整个数据。白化仅需要数据，不需要有关信号的信息，它所做的就是通过把观测信号数据除以数据的振幅谱密度 (ASD)，来消除噪声间的相关性。而要获得数据中信号频域的

<sup>1</sup><https://www.gw-openscience.org/about/>

ASD，我们把时域上的数据进行快速 Fourier 变换到频域，然后采用 Welch 平均周期图法计算功率谱密度，如图5–2所示。图5–2中展示的频率是从 10~2000Hz，因为在频率低

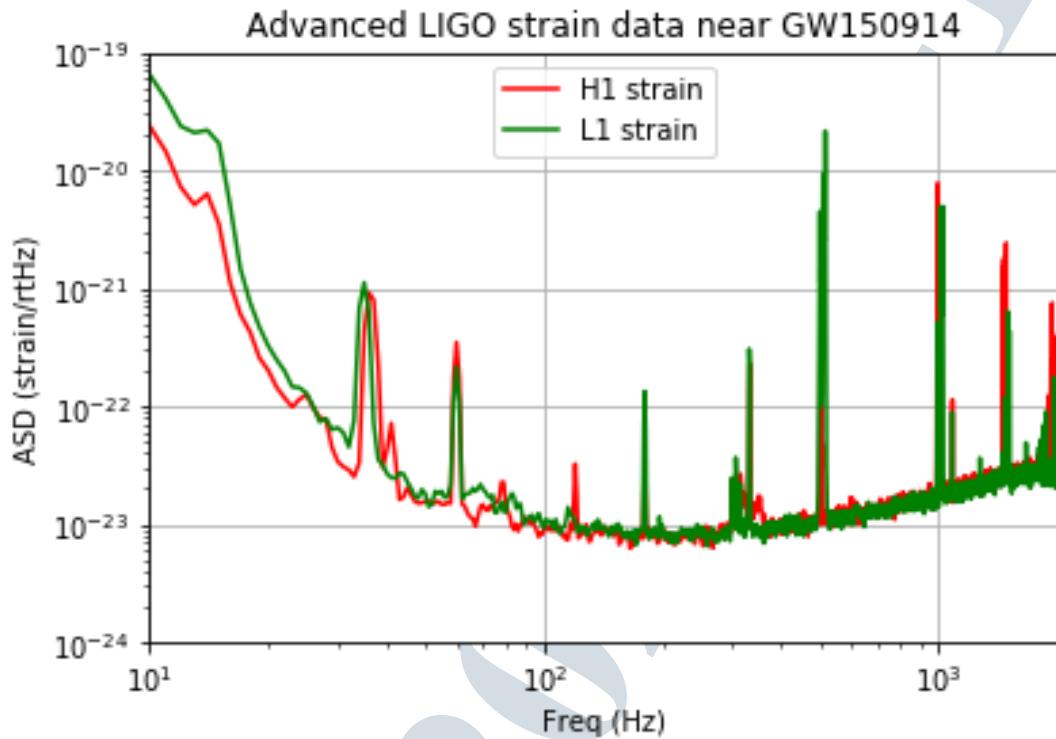


图 5–2 绿线和红线分别是 Hanford 与 Livingston 的地面引力波探测器装置的探测信号的振幅谱密度，其中一些很强的竖线来自于仪器的噪声 (~500Hz 镜子悬挂绳索共振)。

Figure 5–2 Amplitude spectral densities of LIGO detectors.

于 10Hz 的区域，aLIGO 的噪声过大导致很难从噪声中寻找到引力波信号。在图5–2中所示的部分是利用 32s 的数据所获得的 ASD。可以利用上述 ASD 进行白化操作。

### 5.1.2 带通滤波（Bandpass）

由于 GW150914 的信号主要贡献在中间频段，因此为了去除其他频率噪声的影响，我们可以采用带通滤波来保留 [20Hz,300Hz] 频率范围内的数据。在这样的操作下就可以看到例如图5–3展示的数据。

我们利用时域上的波形制作时频图，来显示数据中的 GW150914 信号，如图5–4所示。

在图5–3当中，我们可以明显发现数据有震荡的行为出现，并且在波形在一定时间内呈现频率与振幅均大幅增加，并紧接着迅速衰减。这个数据很难由仪器受微扰后的系

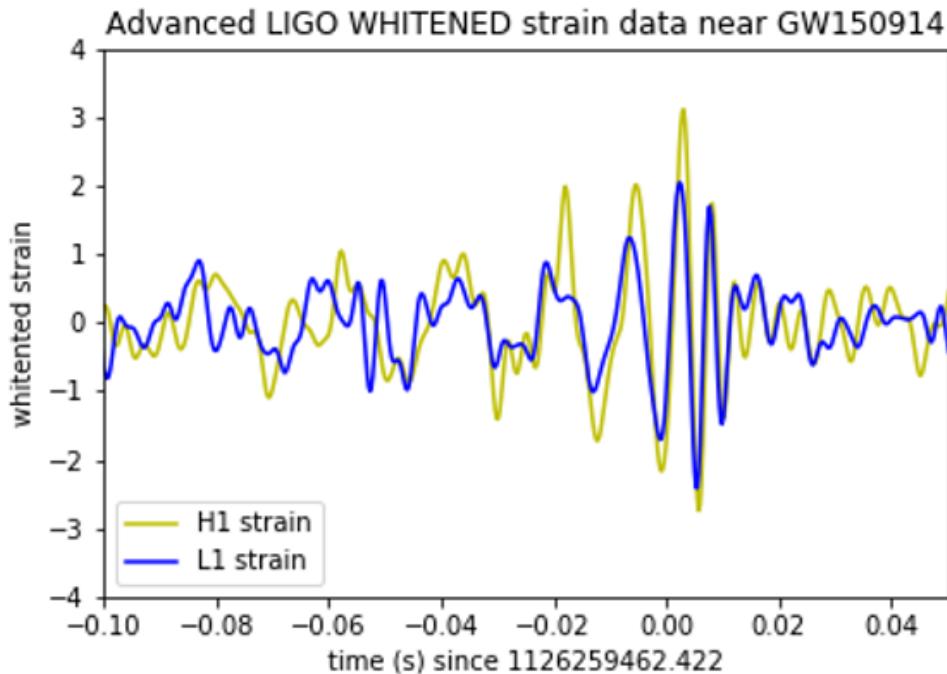


图 5-3 黄线和蓝线是经过白化与带通滤波之后的来自于 LIGO 的 Hanford 与 Livingston 地面探测器所采集的数据中所包含的引力波信号。

Figure 5-3 Data after bandpass

统恢复平衡过程中所产生，因为那将是一个围绕稳定状态来回震荡的过程，因而预期会有一个保持不变的频率与逐步衰减的振幅，而这显然与 GW150914 的性质不符。而一个两体系统的轨道旋绕运动可以是一个很好的解释，通过向外辐射引力波，两体之间的轨道半径在不断衰减，从而角速度变大，两体越来越近，进一步导致辐射的引力波振幅与频率上升，最后在两体并合之后轨道运动结束变成了一个稳定的系统，这样也对应了振幅的迅速衰减的现象。接下来我们基于此结论，要继续论证，做两体旋绕运动的是两个黑洞。

利用图 5-1 当中的数据，我们可以简单地估计频率随时间演化的关系，所采用的方法就是计算波形当中经过数据为零 (zero-crossing) 的点，并计算它们的对应频率。

由于实际探测数据是离散的并不连续，因而我们寻找所有相邻的对应的 strain data 在横坐标上下的相邻时间点 (e.g.  $T_1, T_2$ )，再并求其中间值  $T = (T_1 + T_2)/2$ ，经过筛选所有在时间间隔  $0.35s < t < 0.43s$  内的点，得到一系列近似的 zero-crossing 点，每个点之间的距离记为  $\delta t$ ，而每个 zero-crossing 点对应的  $f_{\text{GW}}$  是由以下公式给出

$$f_{\text{GW}} = 1/(2\delta t) \quad (5.3)$$

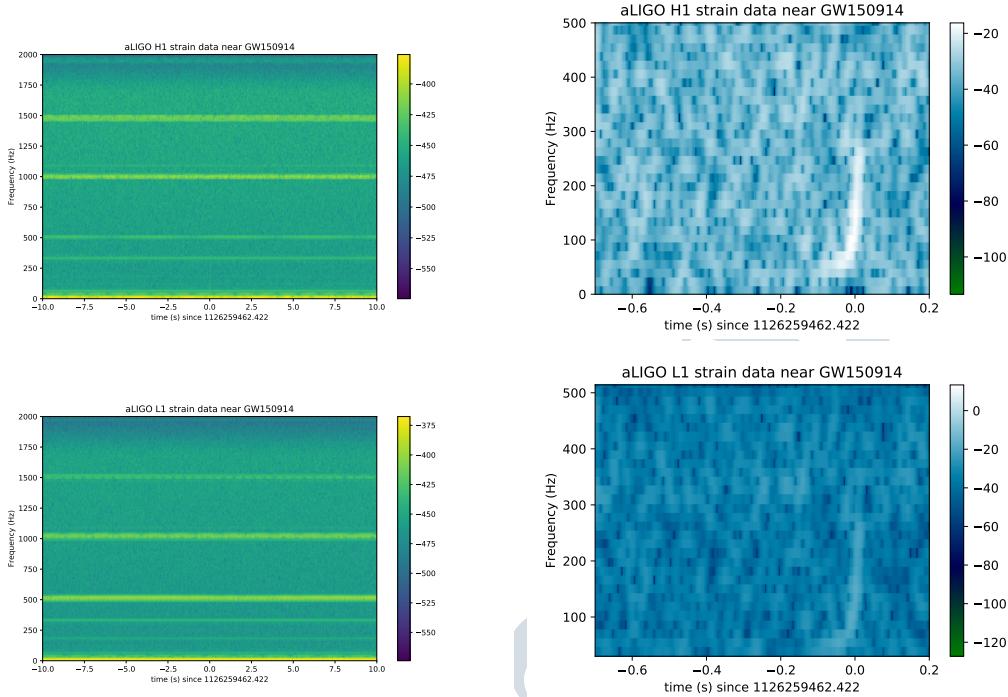


图 5-4 左上角与左下角的图分别为 H1 和 L1 探测数据的时频图，中间靠下的位置就是信号 GW150914，放大后就如右边两图所示。

Figure 5-4 Time-frequency plot of LIGO data around GW150914.

最终频率随时间演化的示意图如图 5-5。

图中绿色的点代表经过上述计算，得到的频率-时间数据，而绿线则是我们对于这一些数据所做的拟合，红线和蓝线则是代表了由 5.4 式所计算出来的，chirp mass 分别为  $30M_{\odot}$ ， $40M_{\odot}$  的时候，频率随时间演化的关系。

$$f_{GW}^{-8/3}(t) = \frac{(8\pi)^{8/3}}{5} \left( \frac{GM}{c^3} \right)^{5/3} (t_c - t) \quad (5.4)$$

## 5.2 论证引力波源为双黑洞系统

由图 5-3 和图 5-4 可见，振幅最大的时刻也对应于频率最高的时刻，这个现象与我们之前所做的两体轨道运动的假设是相一致的，记这个最高振幅对应的频率为  $f_{GW}|_{\max}$ ，图 5-4 中的预估值为  $\sim 150$  Hz。进而我们可以把观测数据解释为两体在以一定轨道频率上限，做两体旋绕轨道运动。基于 Keplerian 动力学，我们估计该系统最高轨道角频率为：

$$\omega_{\text{Kep}}|_{\max} = \frac{2\pi f_{GW}|_{\max}}{2} = 150\pi \text{ rad/s} \quad (5.5)$$

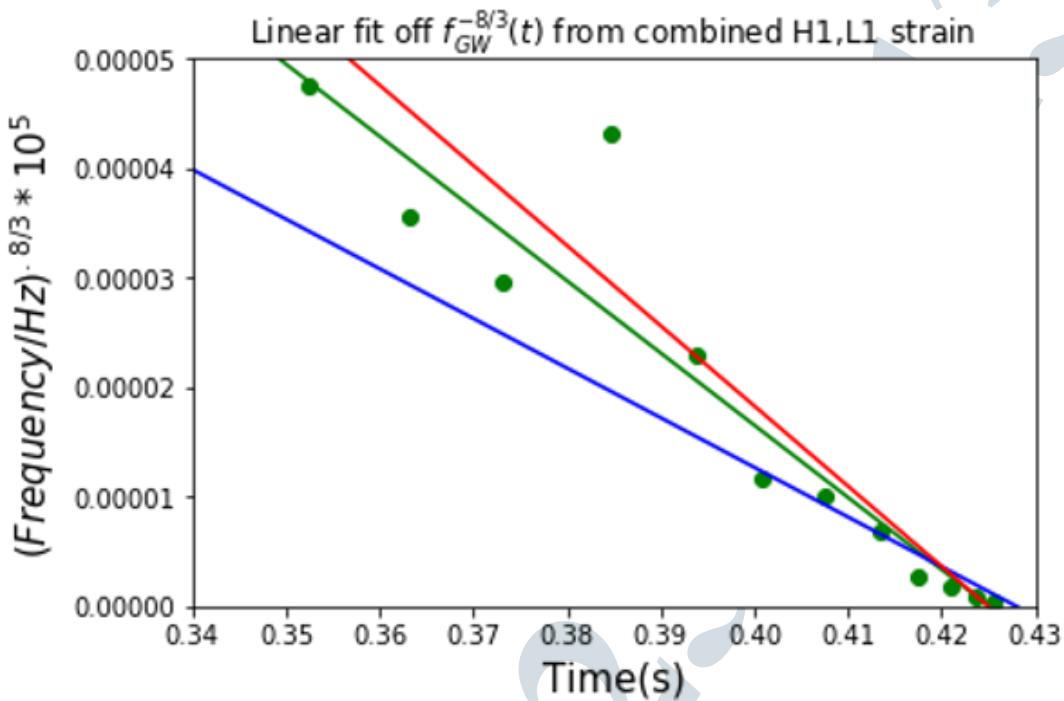


图 5-5 绿色的线是对数据点的拟合结果，利用 5.4 式可以估算出数据所对应的 chirp mass，大约为  $37M_{\odot}$ 。而红线和绿线则代表了，chirp mass 为  $40M_{\odot}$  与  $30M_{\odot}$  的情况下，时间与频率的对应关系

Figure 5-5 Fitting of frequency versus time and corresponding masses.

然后，我们进一步考虑啁啾质量范围。结合 Einstein 四极距公式以及 Newtonian 运动定律，可以简单地推导出啁啾质量与引力波频率、频率的一阶导的关系，

$$\mathcal{M} = \frac{c^3}{G} \left[ \left( \frac{5}{96} \right)^3 \pi^{-8} f_{\text{GW}}^{-11} \dot{f}_{\text{GW}}^3 \right]^{1/5} \quad (5.6)$$

此公式在 Newtonian 近似成立的条件下一直适用。且通过 5.6 式可以得知，利用现有观测数据例如图 5-4，我们可以通过在任何时间点处的引力波频率与频率一阶导数来直接求得系统的啁啾质量。我们发现双黑洞的旋近信号所处频率大致在  $30 < f_{\text{GW}} < 150 \text{ Hz}$  的范围内，而在  $f_{\text{GW}} < 150 \text{ Hz}$  区间，系统的啁啾质量近似保持不变，这是双星轨道运动自然而然得出的结果。同样的，探测数据当中表现出来的引力波振幅随着频率变大而变大的现象，也证实了之前计算当中采用的一些假设也是适用的。另一方面，我们知道了系统的啁啾质量和并合时间  $t_c$ ，就能得知频率随时间的变化方式。基于图 5-5 的结论，我们保守估计 GW150914 对应的双星系统的啁啾质量为  $30M_{\odot}$ 。在最简单的情况下，我们假设双星系统是等质量的，我们计算系统在辐射引力波频率最高处所对应的两个物体

之间的距离  $R$ :

$$R = \left( \frac{GM}{\omega_{\text{Kep}}^2 |_{\max}} \right)^{1/3} = 350 \text{ km} \quad (5.7)$$

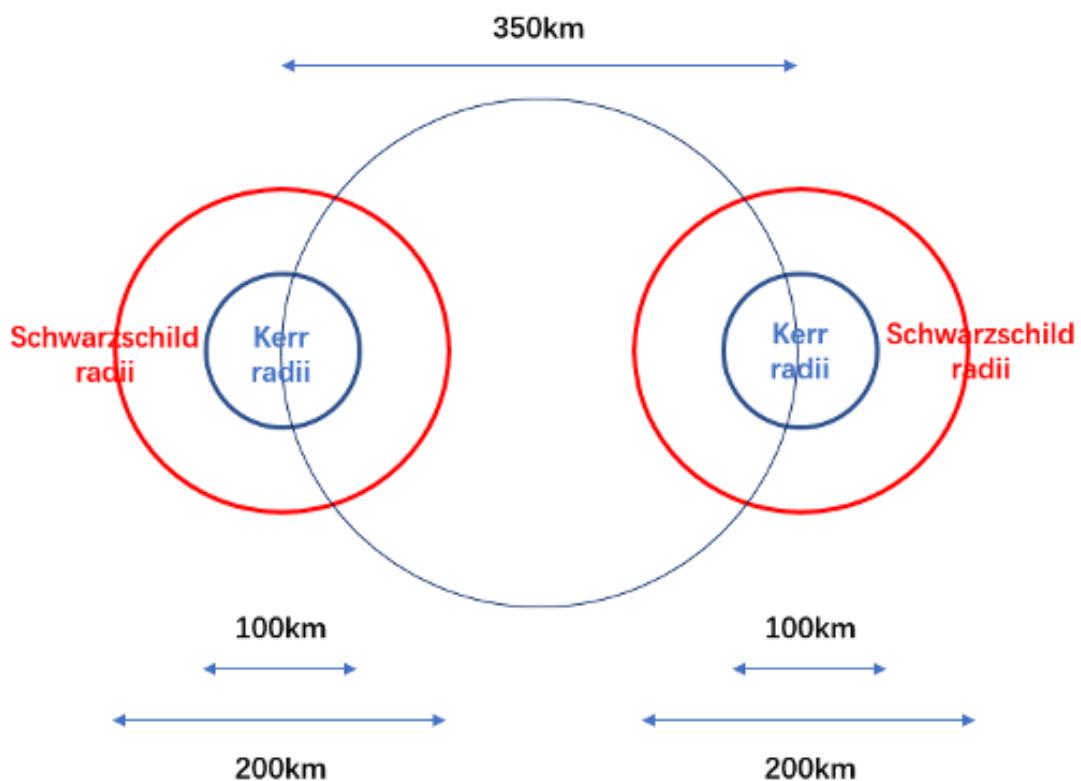


图 5-6 中间的蓝圈代表两物体间隔，蓝色小圈是物体的 Kerr 半径，而红圈是物体的 Schwarzschild 半径。后面会证明对于不等质量的系统来说，两个物体会挨得更近。

Figure 5-6 Illustration of scales.

同时为了量化相对于他们自身引力半径的两个物体的接近程度我们引入变量，致密度比  $\mathcal{R}$ 。它是由 Newtonian 轨道间隔与最小的自身半径之比来定义的。如图 5-6 所示，对于无自旋，圆轨道且等质量的情况下，致密性比  $\mathcal{R} = 350 \text{ km} / 206 \text{ km} \approx 1.7$ 。而 Newtonian 或 Keplerian 轨道演化会在致密比接近 1 的时候失效，这也证明了物体是相当致密的。

### 5.3 讨论假设的合理性以及对于结论的影响

接下来，我们对所做的假设——在无自旋、等质量和圆轨道假设下，两个并合的物体都是黑洞——进行一个检验。检验他们对于我所得到的结论的影响。

### 5.3.1 非零轨道偏心率

因为在非圆轨道下系统偏心率  $e > 0$ , 从而在5.7式当中的  $R$  就不再代表两个物体的间距, 而瞬时的间距  $r_{\text{sep}} \geq (1 - e)R$ , 因而存在比之前更加小的  $R$  值。所以在非零轨道偏心率的情况下, 双星系统的致密比会相比于圆轨道的情况更加小, 也就意味着双星更近。轨道的非零偏心率还会引入对于光度的影响因素, 尤其见于高偏心率的系统, 物体经过近心点的速度是大于在远心点的速度, 因而反映在引力波信号上就会呈现出一个, 在高峰值与低峰值之间变换的模式, 但是这并未在观测数据中体现。一个合理的解释是, 随着双星轨道运动演化, 引力波带走了角动量使得系统比并合更快的圆化自己的轨道, 因而我们观测到的数据并未显示出明显的高偏心率的特征。

### 5.3.2 不等质量

根据之前对于致密比  $\mathcal{R}$  的定义, 我们可以推导出质量比与致密比的关系:

$$\begin{aligned}\mathcal{R} &= \frac{r}{r_{\text{Schwarz}}(M)} = \frac{c^2}{2(\omega_{\text{Kep}}|_{\max} GM)^{2/3}} \\ &= \frac{c^2}{2(\pi f_{\text{GW}}|_{\max} G\mathcal{M})^{2/3}} \frac{q^{2/5}}{(1+q)^{4/5}} \\ &\approx \frac{3.0 q^{2/5}}{(1+q)^{4/5}}\end{aligned}\tag{5.8}$$

这其中, 我们讨论两个黑洞的 Schwarzschild 半径之和,  $r_{\text{Schwarz}}(M) = r_{\text{Schwarz}}(m_1) + r_{\text{Schwarz}}(m_2)$

从而, 根据提到的偏心率对光度的影响, 我们记之为  $g(e)$ 。

$$g(e) = (1 - e)^{-7/2} \left( 1 + \frac{73}{24}e^2 + \frac{37}{96}e^4 \right)\tag{5.9}$$

考虑到对轨道半径的影响, 我们可以写出非零偏心率系统的致密比与圆轨道的致密比之间的关系,

$$\mathcal{R}(e) = g^{2/5}(e)\mathcal{R}(e=0)\tag{5.10}$$

从图5-7我们可以看出, 总体趋势是随着双星越接近等质量的情况, 致密比越大, 也就意味着随着质量比的增加, 双星会越来越近且在  $q \sim 17$  的位置, 甚至小于了双星的 Schwarzschild 半径之和, 而此刻对应的较小质量物体的质量则  $\geq 11M_\odot$ , 这个质量是中子星质量上限的 3~4 倍, 因而 GW150914 信号里对应的双星系统预期是双黑洞系统。

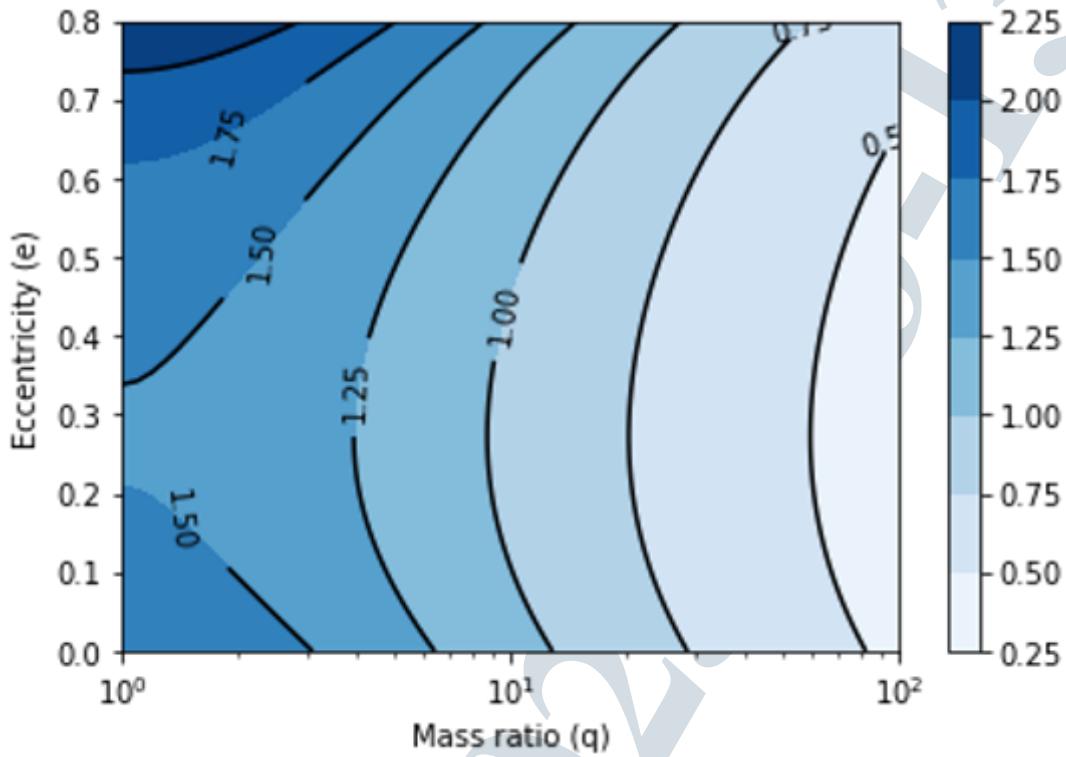


图 5-7 假设在 chirp mass 为  $30M_{\odot}$ , 最高频率为 150Hz 的情况下, 双星系统的质量比和偏心率与致密比的康拓图。

Figure 5-7 Contour for mass ratio and eccentricity, assuming chirp mass of  $30M_{\odot}$ , maximum frequency of 150Hz.

### 5.3.3 自旋效应

关于我们之前做的第三个无自旋假设, 对于一个质量为  $m$ , 自旋角动量为  $S$  的物体, 我们定义一个无量纲自旋参数:

$$\chi = \frac{c}{G} \frac{S}{m^2} \quad (5.11)$$

对于  $\chi = 1$  的情况, Kerr 黑洞的半径  $r_{EK} = 1/2r_{\text{Schwarz}}(m) = Gm/c^2$ 。我们通过以下方法来约束含偏心率, 不等质量和有自旋的轨道致密比:

$$R = \frac{r_{\text{sep}}(M)}{r_{EK}(M)} \leq \frac{R(M)}{r_{EK}(M)} = \frac{c^2}{(GM\omega_{\text{Kep}})^{2/3}} \leq \frac{c^2}{(2^{6/5}G\mathcal{M}\omega_{\text{Kep}})^{2/3}} = \frac{c^2}{(2^{6/5}G\mathcal{M}f_{\text{GW}}|_{\max})^{2/3}} \approx 3.4 \quad (5.12)$$

由于致密比  $\mathcal{R}$  可以由质量比来决定，因而在  $\mathcal{R}$  必须大于 1 的前提下，我们可以限制质量比  $q$ 。同时固定 chirp mass 与  $f_{\text{GW}}|_{\text{max}}$  的值，求得最大总质量。

$$\frac{M_{\text{max}}}{\mathcal{M}} \approx 3.4^{3/2} \times 2^{6/5} \approx 14.4 \quad (5.13)$$

对于 GW150914 信号，最大总质量  $M_{\text{max}} \approx 432M_{\odot}$ ,  $q \approx 83$ ，因此最小的物体质量也大于  $5M_{\odot}$ ，明显高于中子星质量上限。因此我们再次确认了双星系统一定为双黑洞。

### 5.3.4 Newtonian 动力学

现在我们讨论我们计算中利用的 Newtonian 动力学的适用性。Newtonian 动力学在物体运动接近光速或引力能量接近静质能是会失效。而对于受引力与轨道速度限制的双黑洞系统，我们采用 post Newtonian (PN) 近似，来描述他的波形。同样的当不止一个黑洞的自旋非常大时，轨道旋转速度也可以接近光速，这是我们引入自旋-轨道和自旋-自旋修正，这也是包含在 PN 波形当中。同样的对于我们之前采用的四极距公式以及利用致密比作为衡量标准，都是不是足够精确的，但这些假设带来的误差在轨道十分接近黑洞之前都是能够忽略的，并不足以使得我们放弃之前得到的结论。

### 5.3.5 哽啾质量测量

对于 5.6 式是否在双黑洞旋近、并合和铃宕阶段都适用的问题，答案为否。因为在接近最后的轨道的时候，就不存在 5.6 式所需求的稳定的圆轨道，根据广义相对论，存在一个最内稳定圆轨道 (ISCO)，致密物体（其中一个比另一个大得多）并合时，存在非圆轨道，并向内“倾斜”。现在，我们基于最终频率  $f_{\text{GW}}|_{\text{fin}}$ ，此时不用 Newtonian 近似处理，由于目前没有观测到质量  $> 3M_{\odot}$  的中子星，因而我们对他的质量上限采用一个保守估计  $4.76M_{\odot}$ ，选择这个值是因为已知图 5-1 中读出的哽啾质量的情况下，为了较小的物体质量小于阈值  $4.76M_{\odot}$ ，较大物体的质量需不小于  $476M_{\odot}$ 。这样质量比  $q \geq 100$  系统叫做极端质量比旋近 (EMRI)，在这种系统中小质量物体沿着测地线运动。轨道频率  $\omega_{\text{orb}}$  为：

$$\omega_{\text{orb}} = \frac{\sqrt{GM}}{r^{3/2} + \chi(\sqrt{GM}/c)^3} = \frac{c^3}{GM} \left( \chi + \left( \frac{c^2 r}{GM} \right)^{3/2} \right) \quad (5.14)$$

$\chi$  是无量纲自旋， $r$  是稳定轨道半径。由 5.14 式，对于 Schwarzschild 黑洞，它在 ISCO 附近  $r = 6GM/c^2$ ,  $f_{\text{GW}} = 4.4(M_{\odot}/M)\text{kHz}$ ，而对于 Kerr 黑洞，它在 ISCO 附近的频率  $f_{\text{GW}} = 32(M_{\odot}/M)\text{kHz}$ 。在物体最后并合时所发出的最高频引力波，应该来自 light ring。Light ring 的位置在：

$$r_{\text{LR}} = \frac{2GM}{c^2} \left[ 1 + \cos \left( \frac{2}{3} \cos^{-1}(-\chi) \right) \right] \quad (5.15)$$

结合上述两式，一个 Schwarzschild 双黑洞系统中小质量物体进入最后轨道的频率约是 67Hz，而对于 Kerr 双黑洞，他的 light right 更小，相对应的轨道频率更低。然而我们从数据中得到的最高频率是大于这个数的，因此这两种情况都可以被排除，从而较小物体的质量也必须高于中子星质量上限  $4.76M_{\odot}$ 。

### 5.3.6 估计波源的光度距离

引力波在穿越宇宙时被宇宙的膨胀拉伸，这就增加了在地球上观测到的引力波长，并降低了频率，而探测器探测到的质量会小  $(1+z)$  倍。下面我们会分析红移的大小。引力波振幅与光度距离  $d_L$  满足  $h \propto 1/d_L$  的关系，当距离增加时振幅也降低。从图 5-1 可见，引力波的 strain data 的极大值  $h|_{\max} \sim 10^{-21}$ ，当我们离源越近信号强度也越大，直到黑洞的 Schwarzschild 半径处，引力非线性效应开始显现。粗略估计光度距离的上限：

$$d_L < 10^{21} \times 200 \text{ km} \sim 6 \text{ Gpc} \quad (5.16)$$

我们继续采用一个更精确的计算方法，首先一个物体理论上的最大光度为普朗克光度  $L_{\text{Planck}} = c^5/G = 3.6 \times 10^{52} \text{ W}$ ，光度距离与引力波光度的关系如下：

$$L \sim \frac{c^3 d_L^2}{4G} |\dot{h}|^2 \sim \frac{c^5}{4G} \left( \frac{\omega_{\text{GW}} d_L h|_{\max}}{c} \right)^2 \quad (5.17)$$

因此我们有，

$$\frac{L_{\text{peak}}}{L_{\text{Planck}}} \equiv \frac{L|_{\max}}{L_{\text{Planck}}} \sim 0.2 \times 10^{-3} \sim \left( \frac{\omega_{\text{GW}} d_L h|_{\max}}{c} \right)^2 \quad (5.18)$$

我们估计了振幅达到最大值时的光度距离

$$d_L \sim 45 \text{ Gpc} \left( \frac{1 \text{ Hz}}{f_{\text{GW}}|_{\max}} \right) \left( \frac{10^{-21}}{h|_{\max}} \right) \quad (5.19)$$

对于 GW150914 信号，他的光度距离  $d_L \sim 300 \text{ Mpc}$ ，对应的红移  $z \sim 0.1$ ，因此探测器探测的双黑洞系统质量与源质量差 0.1%。

## 5.4 结论

经过以上的分析，GW150914 信号其实是来源于距地约 300Mpc 的两个互相旋绕并最终并合的双黑洞系统。我们一开始采用了 Newtonian 动力学、等质量双星、圆轨道和无自旋的假设来分析数据，后面进一步验证了这些假设并不会影响到我们所得到的结论，并最后讨论了信号的光度距离。根据这些基本的论证，我们还不能很准确地对质量比进行限制，之后在 PN 波形中我们可以进一步限制。且以上处理适用于类 GW150914 的双黑洞系统的引力波数据。

2023-03-13

# 6

## 引力波信号探测

实验学家艰苦卓绝的工作，就是不断地提高仪器的灵敏度；而数据处理团队的贡献，就是从复杂的噪声中提取信号。可以说，数据处理团队是整个引力波探测链条中费效比最低的：要想将位置噪声或者加速度噪声压低 10%，或者将轨道定轨误差降低一半，往往伴随着数额巨大的软硬件开发成本和实验方案修改，隐藏着对实验可行性的威胁，因而所费甚多。然而，一个聪明的硕士生、博士生，通过仔细思考引力波探测器的运行轨道、工作特征、数据特性，提出创新的数据处理方案，往往可以带来 10% 以上的敏感度提升。更棒的是，这样的提升往往是“免费”的，不需要对现有硬件作太多改动。万一出了差错，大不了把程序扔到回收站就是了。这么看起来，换算成研发经费的话，也许一个做数据处理的优秀研究生就能抵得上成百上千万的硬件研发投入呢！

在之前章节的学习中，我们认识到，引力波的幅度极小，对引力波信号的探测非常困难。这一方面是因为广义相对论框架下不存在引力的偶极辐射，最低阶的引力波辐射就来自质量四极矩变化，与电磁波的偶极辐射相比自然弱了不少。另一方面，在所有四种基础相互作用中，引力是最弱的，描述引力作用的引力常数  $G$  在 SI 单位制下只有  $10^{-11}$  的量级，而描述电磁力的 Coulomb 常数  $k$  在 SI 单位制下则有  $10^9$  的量级。

由于数据中一定会存在各类随机起伏的噪声，这就导致相应的数据处理方法一定依赖于概率论、统计学相关的知识。对引力波信号的探测与分析，需要依赖专门的数据处理方法实现。在本章中，我们将结合不同特性的引力波信号，介绍现有的探测方法。不管具体方法如何，本质上来说，其思想大致上都可以归结为将理论的波形模板与实测数据进行相关、匹配，利用所谓的匹配滤波技术，寻找数据中对波形的强响应。注意到，有一些无法精确建模的引力波波形，超新星爆发、宇宙弦碰撞等，上述描述的探测手段并不适用，在这里不作讨论。

$B$	$C$	$B, C$	$A = \overline{B, C}$	$D = \overline{B} + \overline{C}$	$B + C$	$\overline{B + C}$	$\overline{\overline{B}, \overline{C}}$
T	T	T	F	F	T	F	F
T	F	F	T	T	T	F	F
F	T	F	T	T	T	F	F
F	F	F	T	T	F	T	T

表 6-1 一些逻辑操作的真值表

Table 6-1 Truth table of some logic operation

## 6.1 概率初步

### 6.1.1 概率的性质

通常，我们可以用大写字母  $A, B, C, \dots$  来代表命题 (proposition)。比如，可以定义  $A \equiv$  “中山大学校珠海校区篮球比赛亚军是天琴中心队”。对应的否命题则为  $\overline{A} \equiv$  “ $A$  是假的”，也称为  $A$  命题的 逻辑非。我们定义命题的真值为该命题的正确性，即，若命题为真，则其真值为真，或用数字 1 表示；若命题为假，则其真值为假，或用数字 0 表示。对于多个命题，还可以定义操作“与”和“或”。

#### 材料 6.1 Compound propositions

$A, B$  表示  $A$  与  $B$  同时为真，通常也称为逻辑与或逻辑积 (logical product, conjunction)

$A + B$  表示  $A$  或  $B$  起码一个为真，通常也称为逻辑或或逻辑和 (logical sum, disjunction)

我们通常只对二分的命题感兴趣，即，命题非真即假。可以定义真值 1 或字母 T 表示命题为真，真值 0 或字母 F 表示命题为假。1854 年，Boole 首次指出，通过逻辑与、逻辑或、逻辑非三种操作，可以计算不同情况下的命题的真值，这就是 Boolean 代数。

如，可以考虑  $A = \overline{B, C}$  和  $D = \overline{B} + \overline{C}$ 。如果普遍地有  $A$  和  $D$  的真值一致，则两者相同。我们可以构造真值表，如表 6-1

注意到，不管  $B$  和  $C$  的真值为何， $A$  和  $D$  的真值总是相同，这就表明逻辑上两者等价。换言之，我们构建了逻辑关系

$$\overline{B, C} = \overline{B} + \overline{C} \quad \text{and} \quad \neq \overline{B, C} \quad (6.1)$$

### 材料 6.2 An adequate set of operations

我们可以通过逻辑积 (AND)、逻辑或 (OR) 和逻辑非 (NOT) 构造出不同命题的复杂组合。但是上述操作是否足够，是否能找到更简单的完备操作？答案是肯定的。可以构建逻辑与非 (NAND) 操作

$$\begin{aligned} A \uparrow B &\equiv \overline{A, B} = \overline{\overline{A}} + \overline{\overline{B}} \\ \overline{\overline{A}} &= A \uparrow A \\ A, B &= (A \uparrow B) \uparrow (A \uparrow B) \\ A + B &= (A \uparrow A) \uparrow (B \uparrow B) \end{aligned} \tag{6.2}$$

和逻辑或非 (NOR) 操作，

$$\begin{aligned} A \downarrow B &\equiv \overline{A + B} = \overline{\overline{A}, \overline{B}} \\ \overline{\overline{A}} &= A \downarrow A \\ A, B &= (A \downarrow A) \downarrow (B \downarrow B) \\ A + B &= (A \downarrow B) \downarrow (A \downarrow B) \end{aligned} \tag{6.3}$$

可以覆盖所有类型的 Boolean 代数操作。

概率，就是对不同命题可能性的量化。那么如何将定性的命题和定量的可能性结合起来呢？从合理性要求出发，我们可以预期，对概率的表述，需要满足以下三个条件：

- ▶ 定量的可能性由实数表述
- ▶ 必须与常识一致。如，若有新的信息加入，则对应的概率应该增加
- ▶ 自洽性，包括
  - Structural consistency: 如果可以用多种方式表述一个结论，每种方式都应该得到同一个结果。
  - Propriety: 对于相关的信息，理论应该全都考虑
  - Jaynes consistency: 相同的知识应该对应相同的概率。比如，如果  $A, B|C = B|C$ ，那么对应的就应该有个  $A, B|C$  对应的概率与  $B|C$  对应的概率一致。

我们可以定义  $p(A)$  为事件  $A$  的概率，对应的  $p(A|B)$  就表示  $B$  命题成立的条件下， $A$  命题成立的概率。可以证明<sup>1</sup>，通过上述的合理性要求，可以推导出概率的乘法法则 (product rule) 和加法法则 (sum rule)。

<sup>1</sup>可以参考 [48] 第 2.5 章内容

乘法法则表示，对于  $p(A, B|C)$ ，可以写成

$$\begin{aligned} p(A, B|C) &= p(A|B, C)p(B|C) \\ &= p(B|A, C)p(A|C) \end{aligned} \quad (6.4)$$

基于乘法法则，可以证明，完全明确对应的概率是 1。更进一步，可以证明，概率的定义可以选取值域  $[0, 1]$  的递增函数，或  $[+\infty, 1]$  的递减函数。通常，选取的结果是  $0 \leq p(x) \leq 1$ 。

加法法则表示，对于互斥的两个命题  $A$  和  $\bar{A}$ ，有  $p(A|B) + p(\bar{A}|B) = 1$ 。

乘法法则和加法法则是概率论的基石。我们看到，可以从非常基础的几个论断出发推导出概率的诸多性质，难怪 Laplace 说 “the theory of probabilities is basically just common sense reduced to calculus.”

### 6.1.2 频率学派 v.s.Bayesian

在上述描述中，我们并没有使用本科概率论和统计学相关课程中对于概率的基本定义，即重复  $N$  次相同、独立的实验，得到  $N_A$  次结果为  $A$ ，则概率  $p(A)$  的定义为  $p(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$ 。这样的定义中，一定牵动着所谓随机变量的定义，如果没有随机性，则无法解释在全同状态下，实验的结果会给出不同的结论。

实际上，关于概率的这两种不同的解读、定义方式，恰恰牵动了两种概率理论的学派，即频率学派 (frequentist) 和贝叶斯派 (Bayesian)。Bayesian 的观点则认为， $A$  可以只是一个逻辑命题，而不一定非得是随机变量。

运用概率的乘法法则，可以得到

$$p(A|B, C) = \frac{p(A|C)p(B|A, C)}{p(B|C)} \quad (6.5)$$

对应到我们通常会遇到的情形，可以写出

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)} \quad (6.6)$$

这其中， $H_i$  是某个假说 (hypothesis) 成立的命题； $I$  代表了描述我们先验信息的命题； $D$  表示了数据。我们通常把  $p(H_i|D, I)$  称为假说  $H_i$  的后验概率； $p(H_i|I)$  则是假说  $H_i$  的先验概率； $p(D|H_i, I)$  是在  $H_i$  和  $I$  成立的前提下，得到数据  $D$  的概率，通常也被叫做似然函数 likelihood，写成  $\mathcal{L}(H_i)$ ，最后， $p(D|I) = \sum_i p(H_i|I)p(D|H_i, I)$  是所谓的 Bayesian evidence，其作用是确保后验概率归一  $\sum_i p(H_i|D, I) = 1$ 。

### 6.1.2.1 最大熵原理

由于在 Bayesian 体系中，可以利用新的信息来更新对于某个命题的可信度，所以不同的信息可以更新对应命题的概率。Jaynes 在 1957 年证明，利用 Shannon 对于信息熵的定义，可以在有效信息的基础上唯一确定概率分布。

对于离散的概率分布， $p_1, p_2, \dots, p_n$ ，可以定义信息熵

$$S(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \ln(p_i). \quad (6.7)$$

最大熵原理指出，在给定信息的前提下，最合理的概率分布应该是拥有最大信息熵的概率分布。换言之，任何其他分布一定蕴藏着现有信息无法推导出的结论。

让我们举一个例子来说明。考虑离散分布满足  $\sum_{i=1}^M p_i = 1$  可以使用 Lagrange 未定乘子法，对于任意的  $j$ ，都有

$$\begin{aligned} \frac{\partial}{\partial p_j} \left[ - \sum_{i=1}^M p_i \ln p_i - \lambda \left( \sum_{i=1}^M p_i - 1 \right) \right] &= 0 \\ - \ln p_j - p_j \frac{\partial \ln p_j}{\partial p_j} - \lambda \frac{\partial p_j}{\partial p_j} &= 0 \\ - \ln p_j - 1 - \lambda &= 0 \end{aligned} \quad (6.8)$$

换言之， $\ln p_j = 1 + \lambda$  对于任意  $j$  都成立。这就意味着，通过最大熵原理，如果没有别的信息，应该认为不同的结果概率相同。

### 6.1.3 典型分布

在数据处理中，打交道最多的分布是二项式分布、Poisson 分布和 Gaussian 分布，并且他们之间也互相联系。

#### 6.1.3.1 Binomial

考虑  $n$  次独立的抽取，每次的成功几率为  $p$ ，则获得  $x$  次成功的概率由二项式分布给出。

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}, x \in \{0, 1, \dots, n\}, 0 \leq p \leq 1 \quad (6.9)$$

对应的，二项式分布的期望值  $\langle x \rangle = \sum_{i=1}^{\infty} x P(x) = np$ ，而方差为  $\langle x^2 \rangle - \langle x \rangle^2 = np(1-p)$

### 6.1.3.2 Poisson

将二项式分布推广到  $n \rightarrow \infty$  的情形，同时使得  $p \rightarrow 0$ ，使得  $Np$  趋向于常数  $\lambda$ ，则有

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x \in \mathbb{N}, \lambda \geq 0 \quad (6.10)$$

其描述的场景为，给定范围内发生某事件的次数  $x$ ，预期为  $\lambda$ ，其概率可以由 Poisson 分布给出，如单位面积、单位时间内的雨滴数目，或者全球每天的出生人口数目，等等。

Poisson 分布的期望值和方差均为  $\lambda$ 。

### 6.1.3.3 Gaussian

二项式分布和 Poisson 分布都是离散分布。考虑  $n \rightarrow \infty, x \rightarrow \infty$  的情形，则得到正态分布，或 Gaussian 分布  $N(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x, \mu < \infty, \sigma > 0 \quad (6.11)$$

需要注意的是，对于连续变量，概率分布对应的纵轴是所谓概率密度，而非概率。

中心极限定理保证了，大量独立的随机变量相加时，其和服从 Gaussian 分布，而生活中的各类连续测量量均包含了各种不同的测量误差，因此经常会假设连续型数据服从高斯分布。

Gaussian 分布下，期望值为  $\mu$ ，方差为  $\sigma^2$ 。

另一种有趣的推导出 Gaussian 分布的方式如下：考虑一个分布  $p(x)$ ，已知其分布的平均值为  $\mu$ ，这时，一个自然的想法是最靠近期望值的地方，概率密度越高，而越远离  $\mu$  的地方，概率密度越低。这一要求就表明了概率分布在  $x = \mu$  的地方有  $p'(x) = 0$ 。根据概率的非负性要求，可以将概率分布写成  $p(x) = \exp(f(x))$ 。利用 Taylor 展开，可以写成  $p(x) \approx \exp [f(x)|_{x=\mu} + f'(x)|_{x=\mu}(x - \mu) + \frac{1}{2}f''(x)|_{x=\mu}(x - \mu)^2 + O(x - \mu)^3]$  若  $f''(x)|_{x=\mu} > 0$ ，则有整个概率分布发散，与预期不符，因此要求  $f''(x)|_{x=\mu} < 0$ ，由此定义  $f''(x)|_{x=\mu} = -1/\sigma^2$  在最低阶近似下，可以得到  $p(x) \propto \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ ，这一形式恰恰就是 Gaussian 分布的函数形式。

实际上，也可以从最大熵原理出发，证明，在给定期望值和方差的前提下，信息熵最大的分布就是 Gaussian 分布。

## 6.2 时序列分析

### 6.2.1 相关函数

对于两个时间域函数  $s_1(t)$  和  $s_2(t)$ , 可以定义操作互相关 (cross-correlation)

$$s_1 \star s_2(\tau) = \int_{-\infty}^{\infty} s_1(t)s_2(t + \tau)dt \quad (6.12)$$

可以这样去理解: 将  $s_2$  错开时间  $\tau$  以后和  $s_1$  进行相关操作, 并在整个时间段上对两者的积进行求和。有些情况下, 对相关函数进行归一化会简化计算, 可以选取归一化条件  $1/s \star s(0)$ , 这样就能保证自相关函数的最大值为 1。

利用互相关的定义, 也可以衍生到自相关

$$s \star s(\tau) = \int_{-\infty}^{\infty} s(t)s(t + \tau)dt \quad (6.13)$$

不难发现, 当  $\tau = 0$  时, 数据与自身重合, 因此其自相关达到最大。

在实际操作中, 由于数据长度有限, 通常会将正负无穷的积分转换为较长的时间  $2T$ , 即  $\int_{-T}^T$ 。在数据长度有限的前提下, 也会采取另一种归一化条件  $1/2T$ , 即  $(1/2T)\int_{-T}^T$ 。

### 6.2.2 卷积

数据处理中, 另一个有用的操作是卷积 (convolution)。可以定义两个函数  $s_1(t)$  和  $s_2(t)$  的卷积

$$s_1 * s_2(\tau) = \int_{-\infty}^{\infty} s_1(t)s_2(\tau - t)dt \quad (6.14)$$

互相关和卷积的表达式中相差一个负号, 但对应的意义就发生了巨大的变化。

### 6.2.3 Fourier 变换 (Fourier Transform)

考虑时间序列  $s(t)$ , 可以得到其 Fourier 变换后的函数  $\mathcal{F}[s(t)] = \tilde{s}(f)$ ,

$$\tilde{s}(f) = \int_{-\infty}^{\infty} s(t)e^{-i2\pi ft}dt \quad (6.15)$$

经过 Fourier 变换后, 原来定义在时间域的函数  $s(t)$  变成了定义在频率域的函数  $\tilde{s}(f)$ 。也可以通过逆 Fourier 变换, 把频率域函数  $\tilde{s}(f)$  转换到时间域  $s(t)$ 。

$$s(t) = \int_{-\infty}^{\infty} \tilde{s}(f)e^{i2\pi ft}df \quad (6.16)$$

 **注 6.3.** 注意, 在不同的场景, 对于 Fourier 变换的形式定义上略有差异。如有时会以大写字母表示频率域函数  $\mathcal{F}[s(t)] = S(f)$ ; 有时会以  $\omega = 2\pi f$  表示频率域函数  $\mathcal{F}[s(t)] = \tilde{s}(\omega)$ ; 还有些文献以  $x$  和  $\xi$  分别表示时间和频率, 等等。这些标记的不同 (除了个别情况下会有常数倍差别外) 并不影响 Fourier 变换的性质。

Fourier 变换这一操作本身并不会丢失任何信息，所以可以通过逆 Fourier 变换，将频率域函数转换成时间域函数。在频率域的数据处理有着很大的便利，因此 Fourier 变换在引力波数据处理中起到了重要的作用。

#### 材料 6.4 Property of Fourier Transform

##### ► 线性性质

两函数之和的 Fourier 变换等于各自变换之和。数学描述是：若函数  $h(t)$  和  $g(t)$  的 Fourier 变换  $\mathcal{F}[h]$  和  $\mathcal{F}[g]$  都存在， $\alpha$  和  $\beta$  为任意常系数，则  $\mathcal{F}[\alpha h + \beta g] = \alpha \mathcal{F}[h] + \beta \mathcal{F}[g]$ 。

##### ► 平移性质

若函数  $h(t)$  存在 Fourier 变换  $\mathcal{F}[h(t)] = \tilde{h}(f)$ ，则对任意实数  $f_0$ ，函数  $h(t)e^{i2\pi f_0 t}$  也存在 Fourier 变换，且有  $\mathcal{F}[h(t)e^{i2\pi f_0 t}] = \tilde{h}(f - f_0)$ 。

##### ► 微分关系

若函数  $h(t)$  当  $|t| \rightarrow \infty$  时的极限为 0，而其导函数  $h'(t)$  的 Fourier 变换存在，则有  $\mathcal{F}[h'(t)] = i2\pi f \mathcal{F}[h(t)]$ ，即导函数的 Fourier 变换等于原函数的 Fourier 变换乘以因子  $i2\pi f$ 。更一般地，若  $h(\pm\infty) = h'(\pm\infty) = \dots = h^{(k-1)}(\pm\infty) = 0$ ，且  $\mathcal{F}[h^{(k)}(t)]$  存在，则  $\mathcal{F}[h^{(k)}(t)] = (i2\pi f)^k \mathcal{F}[h]$ ，即  $k$  阶导数的 Fourier 变换等于原函数的 Fourier 变换乘以因子  $(i2\pi f)^k$ 。

##### ► 卷积特性

若函数  $h(t)$  及  $g(t)$  都在  $(-\infty, +\infty)$  上绝对可积，则卷积函数  $h * g = \int_{-\infty}^{+\infty} h(t - \tau)g(\tau)d\tau$ （或者  $h * g = \int_{-\infty}^{+\infty} h(\tau)g(t - \tau)d\tau$ ）的 Fourier 变换存在，且  $\mathcal{F}[h * g] = \mathcal{F}[h] \cdot \mathcal{F}[g]$ 。卷积性质的逆形式为  $\mathcal{F}^{-1}[\tilde{h}(f) * \tilde{g}(f)] = 2\pi \mathcal{F}^{-1}[\tilde{h}(f)] \cdot \mathcal{F}^{-1}[\tilde{g}(f)]$ ，即两个函数卷积的 Fourier 逆变换等于它们各自的 Fourier 逆变换的乘积乘以  $2\pi$ 。

##### ► Parseval 定理

若函数  $h(t)$  可积且平方可积，则  $\int_{-\infty}^{+\infty} h^2(t)dt = \int_{-\infty}^{+\infty} |\tilde{h}(f)|^2 df$ 。更一般化而言，若函数  $h(t)$  和  $g(t)$  皆为平方可积函数，则  $\int_{-\infty}^{+\infty} h(t)g^*(t)dt = \int_{-\infty}^{+\infty} \tilde{h}(f)\tilde{g}^*(f)df$ 。其中  $\tilde{h}(f)$  和  $\tilde{g}(f)$  分别是  $h(t)$  和  $g(t)$  的 Fourier 变换， $*$  代表复共轭。

##### ► 翻转与共轭

若函数  $h(t)$  存在 Fourier 变换  $\mathcal{F}[h(t)] = \tilde{h}(f)$ ，则函数  $h(-t)$  的 Fourier 变换为  $\mathcal{F}[h(-t)] = \tilde{h}(-f)$ （翻转）。且函数  $h^*(t)$  的 Fourier 变换为  $\mathcal{F}[h^*(t)] = \tilde{h}^*(-f)$ （共轭）。

需要注意的是，我们通常处理的数据都是离散化采点的，所以并不适用积分形式的

Fourier 变换。如果在  $t_n = n\Delta t$  时刻规则采样数据  $s_n = s(t_n)$ , 通常会采用离散 Fourier 变换 (discrete Fourier transform, DFT)。

$$\tilde{s}_k = \sum_{n=0}^{N-1} s_n e^{-i2\pi k \frac{n}{N}}. \quad (6.17)$$

而其中, 高效的快速 Fourier 变换 (fast Fourier transform, FFT) 由于可以将原本  $O(N^2)$  复杂度的计算降为  $O(N \log N)$  复杂度的计算, 因此被各领域广泛采用。

 **注 6.5.** 对应地, 逆离散 Fourier 变换 (*inverse discrete Fourier transform, IDFT*) 为

$$s_n = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{s}_k e^{i2\pi k \frac{n}{N}}. \quad (6.18)$$

若需将离散 Fourier 变换结果与连续的 Fourier 变换进行比较, 往往需要引入归一化因子  $1/N$ 。

若时间域数据以  $\Delta t$  为间隔取点, 共取  $N$  个点, 即时间总长为  $T = (N - 1)\Delta t$ , 在经历离散 Fourier 变换后, 对应的频率域间隔  $\Delta f = \frac{1}{T}$ , 而最高频为  $f_{\text{Nyquist}} = \frac{1}{2\Delta t}$ , 是采样频率的一半。

#### 6.2.4 功率谱密度

我们首先定义随机过程 (random process) 为一个随机变量的序列。可以将引力波探测器的噪声视为随机过程, 虽然无法明确地预测噪声的具体数值, 但可以给出一定的统计学性质。考虑实随机过程  $x(t)$ , 假设分布服从  $p(x)$ , 其期望值则为系综平均  $\langle x \rangle \equiv \int xp(x)dx$ 。如果  $x(t)$  的统计学性质不随时间变化, 我们称其为稳态随机过程 (stationary random process)。这时, 系综平均可以等价为对时间的平均

$$\langle x \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)dt \quad (6.19)$$

可以定义信号的功率为单位时间  $T$  内  $x^2(t)/T$  的积分。如果  $x(t)$  是稳态的, 即时间序列的性质不随时间变化, 那么通过选取足够大的  $T$ , 就有

$$\langle x^2 \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt \quad (6.20)$$

不妨定义函数  $x_T(t)$

$$x_T(t) = \begin{cases} x(t) & -T/2 < t < T/2 \\ 0 & \text{otherwise} \end{cases} \quad (6.21)$$

则有

$$\begin{aligned}
 \langle x^2 \rangle &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} x_T^2(t) dt \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} |\tilde{x}_T(f)|^2 df \\
 &= \lim_{T \rightarrow \infty} \frac{2}{T} \int_0^{\infty} |\tilde{x}_T(f)|^2 df \\
 &= \int_0^{\infty} S_x(f) df
 \end{aligned} \tag{6.22}$$

这里的  $S_x(f)$  就是定义的**功率谱密度** (power spectral density, PSD)。特别地，功率谱密度分为**单边功率谱密度** (single-sided PSD) 和**双边功率谱密度** (double-sided PSD)，差别在定义域上，单边功率谱密度定义域为  $(0, \infty)$ ，而双边功率谱密度的定义域为  $(-\infty, \infty)$ 。由公式6.22定义的是单边功率谱密度，因此其强度是双边功率谱密度的两倍。这其中，第二行的等式源于 Parseval's theorem，也就是

$$\int_{-\infty}^{\infty} |x_T(t)|^2 dt = \int_{-\infty}^{\infty} |\tilde{x}_T(f)|^2 df. \tag{6.23}$$

又因为  $x(t)$  是实的，有  $\tilde{x}_T(-f) = \tilde{x}_T^*(f)$ ，由此得到第三行。并且， $\tilde{x}_T(f)$  是  $x_T(t)$  的 Fourier 变换。由此，我们得到，

$$S_x(f) \equiv \lim_{T \rightarrow \infty} \frac{2}{T} \left| \int_{-T/2}^{T/2} x(t) e^{-i2\pi f t} dt \right|^2 \tag{6.24}$$

并且，在稳态条件下，可以证明，功率谱密度是自相关函数  $x \star x(\tau)$  的 Fourier 变换的两倍。我们不妨展开

$$\begin{aligned}
 S_x(f) &= \lim_{T \rightarrow \infty} \frac{2}{T} \int_{-T/2}^{T/2} x(t) e^{-i2\pi f t} dt \int_{-T/2}^{T/2} x(t') e^{i2\pi f t'} dt' \\
 &= 2 \int_{-\infty}^{\infty} d\tau e^{-i2\pi f \tau} \left[ \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t') x(t' + \tau) dt' \right] \\
 &= 2 \int_{-\infty}^{\infty} d\tau e^{-i2\pi f \tau} x \star x(\tau)
 \end{aligned} \tag{6.25}$$

其中，第二行采用了变量代换  $t = t' + \tau$ ，且其中括号内部分即为  $x(t)$  的自相关函数。注意，由于自相关函数是偶函数，所以其 Fourier 变换后不包含虚数部分。这也被称为 **Wiener-Khintchine 定理**。

功率谱密度的另一种性质也可以推导得出：

$$\begin{aligned}
 \langle \tilde{x}^*(f') \tilde{x}(f) \rangle &= \left\langle \int_{-\infty}^{\infty} x(t') e^{i2\pi f' t'} dt' \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt \right\rangle \\
 &= \left\langle \int_{-\infty}^{\infty} x(t') e^{i2\pi f' t'} dt' \int_{-\infty}^{\infty} x(t' + \tau) e^{-i2\pi f(t'+\tau)} d\tau \right\rangle \\
 &= \int_{-\infty}^{\infty} dt' e^{-i2\pi(f-f')t'} \int_{-\infty}^{\infty} d\tau e^{-i2\pi f\tau} \langle x(t') x(t' + \tau) \rangle
 \end{aligned} \tag{6.26}$$

第二行再次采用变量代换  $t = t' + \tau$ 。注意第一个积分对应于 Dirac delta 函数，第二个积分对应于功率谱密度，由此可以总结出

$$\langle \tilde{x}^*(f') \tilde{x}(f) \rangle = \frac{1}{2} S_x(f) \delta(f - f') \tag{6.27}$$

### 6.2.5 周期图

在第6.2.4节中定义的功率谱密度具有良好的数学性质，但是在数据处理中，往往不具备操作性。通过时长为  $T$  的实测数据  $x(t)$ ，可以利用周期图 (periodogram) 的方法近似功率谱密度，即，首先对时序列  $x(t)$  作 Fourier 变换得到  $\tilde{x}(f)$ ，则  $|\tilde{x}(f)|^2/T$  则为周期图。可以证明在  $T \rightarrow \infty$  的极限下，周期图趋向于功率谱密度。实际上，可以证明  $|\tilde{x}(f)|^2/T$  的方差即为功率谱密度，换言之，周期图法并不是非常理想的估计方法，但在数据处理中，其依然受到了极为广泛的应用。对周期图法的一个简单改进，就是采用平均周期图法（又称 Welch 方法），即将数据分为  $N$  份，分别作周期图，并取平均。当然，这样做的代价就是损失频率分辨率，不过这一点对于数据处理来说往往并不是很大的问题。

## 6.3 信号探测的统计学原理

虽然在给定物理参数的前提下，可以确定地描述引力波信号，但信号通常蕴藏在随机的噪声中。因此，从噪声中提取信号必须依赖统计学方法。通过对仪器特性的理解，可以预期噪声的统计学特性，更进一步地，可以据此计算出得出相应数据  $x$  的概率  $p_0(x)$ 。而如果数据中存在引力波信号，但依然使用噪声假设进行计算，则会改变对应的概率分布。与此同时，如果更换模型，在引力波信号存在的前提下，计算出得到测量数据的概率，则可以得到一个不同的概率  $p_1(x)$ 。从引力波探测器数据中搜索出引力波信号存在的方法，就可以归结为对信号模型下的概率  $p_1(x)$  和噪声模型下的概率  $p_0(x)$  的比对。

### 6.3.1 假设检验 (hypothesis testing)

我们将首先界定几个术语。首先，信号(signal)是由引力波源产生的引力波信号，经引力波探测器的响应，而记录到探测器上的成分，对应于公式4.15

$$h(t) = h_+(t)F_+(\theta, \phi) + h_\times(t)F_\times(\theta, \phi)$$

。其次，噪声(noise)是由探测器过程产生的，不对应引力波信号的成分，通常可以由第6.2.4节中介绍的噪声功率谱密度描述。当然，功率谱密度可以是确定性的，而噪声本身是具有误差起伏的，通常用  $n(t)$  来表示。最后，探测器记录到的数据 (data) 是信号和噪声两部分的线性叠加。也就是，数据  $D(t)$  可以表示为  $D(t) = n(t) + h(t)$ 。

通常，可以将引力波信号的探测问题转化为一个统计学假设检验的问题。这其中，零假设 (null hypothesis)  $H_0$  中，假设了数据中不存在引力波信号，因此可以表示为  $D(t) = n(t)$ ；另一方面，替代假设 (alternative hypothesis)  $H_1$  则假设数据中同时包含信号和噪声，即  $D(t) = n(t) + h(t)$ 。在 frequentist 的框架下，如果要对两个不同的假说进行判断，通常需要进行所谓假设检验 (hypothesis testing 或者 decision rule) $\delta$ 。通过这一操作，可以将数据分类到两个互斥的子集中  $R$  和  $R'$ 。如果数据在  $R$  中，则接受零假设；反之则拒绝。

这个过程中，我们可能会犯下两类错误，即所谓 type I error，又称为误警 false alarm，即  $H_0$  真实成立而将其错误地拒绝；以及 type II error，又称为漏警 false dismissal，即  $H_1$  真实成立而错误地接受零假设。在假设检验的理论中，type I error 对应的概率被称为检验的显著性 (significance of the test)。反过来，正确地将  $H_1$  的数据归类到  $H_1$  为上述命题的反命题，由概率的加法法则可知其概率等于 1-type II error 对应的概率，该值被称为检验的能力 (power of the test)。

上述的描述较为抽象，那么如何从实际操作层面来看，得到一个最优化的检验方法呢？可以从不同的出发点出发，进行推演。下面介绍几种常见的推导思路。

#### 6.3.1.1 Bayesian 方法

在 Bayesian 方法中，可以对判断  $\delta$  分配不同的代价 cost。具体来说，可以引入正实数  $C_{ij}$ ，其中  $i, j = 0, 1$ ， $C_{ij}$  是假设  $H_j$  正确的前提下归类到  $H_i$  的代价。由此，可以定义对于假设检验  $\delta$  的条件风险 (conditional risk)  $R$ ：

$$R_j(\delta) = C_{0j}P_j(R) + C_{1j}P_j(R'), \quad j = 0, 1 \quad (6.28)$$

这其中， $P_j$  是当假设  $H_j$  成立时的概率分布。接下来，我们分别将  $H_0$  和  $H_1$  成立的先验概率 (a priori probability 或者 prior) 定位  $\pi_0$  和  $\pi_1 = 1 - \pi_0$ 。于是就可以定义出 Bayes 风

险 (Bayes risk) 为总体平均代价

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta) \quad (6.29)$$

最终选取可以使 Bayes 风险最小化的探测手段。

### 6.3.1.2 最小最大方法

在很多时候，没有办法准确的对于不同假设的先验概率进行赋值。因此在操作中，通常可以做的是，对于不同的  $\delta$ ，选择判断依据，使得条件风险  $R_0(\delta)$  和  $R_1(\delta)$  中最大值可以最小化。

### 6.3.1.3 Neyman-Pearson 方法

Neyman-Pearson 引理是统计学中的一条重要理论。很多时候，对于不同的选择带来的具体代价很难量化。Neyman-Pearson 方法指出，在给定检验的显著性（换言之，即误警率）的前提下，应该最大化检验的能力 power of the test。Neyman-Pearson 方法对于选取具体某个假设时，所带来的两种错误的某种调和。

### 6.3.1.4 似然函数比检验

在第6.3.1.1节-第6.3.1.3节中，我们描述了不同的逻辑出发点，但是这几种不同的出发点，无一例外地都指向了同一个探测统计量，即似然函数比 likelihood ratio test，即当信号存在时的似然函数  $\mathcal{L}(H_1)$  与信号不存在时的似然函数  $\mathcal{L}(H_0)$  之间的比值  $\Lambda$ 。

$$\Lambda(x) = \frac{p(D|H_1, I)}{p(D|H_0, I)} \quad (6.30)$$

可以证明，在不同的前提下，上述的出发点均可以等价于将似然函数比最大化。根据出发点的不同，依据代价  $C_{ij}$ ，先验  $\pi_i$  和显著度  $\alpha$  的具体取值，可以推导出探测阈值  $k$ ，当  $\Lambda > k$  时，选择假设  $H_1$ 。

### 6.3.1.5 Fisher 方法

通常和上述方法，尤其是和 Neyman-Pearson 方法进行比较的是 Fisher 方法。在这一方法中，是否接受替代假说  $H_1$  的结论，完全是基于对零假设的分析得到的，即，只关心 p 值，即当  $H_0$  成立时得到与实际数据一致或更极端的数据的概率。利用 p 值和预设的显著度进行比较，如果 p 值小于预设阈值，则拒绝零假设，并接受替代假设。近年来，对于 Fisher 方法的可靠性和合理性产生了大量的质疑。实际上，许多学科中对于 Fisher

方法的使用已经大大地违背了 Fisher 发明该方法的初衷，因此，在实际使用中，需谨慎使用。

### 6.3.1.6 似然函数比、内积与匹配滤波

在本小节中，我们将讨论似然函数比  $\Lambda$  的具体表达式。不妨考虑均匀的离散化的采点，则可以定义记录到的数据  $D = D_i$ ，其中，

$$\begin{aligned} D_i &= D(t_i) \\ t_i - t_j &= (i - j)\Delta t \\ \Delta t &= \frac{T}{N-1} \end{aligned} \quad (6.31)$$

$T$  为总观测时长， $N$  为采样总数。

对于噪声模型  $H_0$ ，不妨假设噪声是 Gaussian 的。这就意味着，得到数据  $D_i$  的概率可以由期望值  $\mu$  和方差-协方差矩阵  $\Sigma$  描述。 $k$  维 Gaussian 分布的概率分布可以写为

$$P(D|\mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(D - \mu)^T \Sigma^{-1} (D - \mu)\right)}{\sqrt{(2\pi)^k |\Sigma|}} \quad (6.32)$$

更进一步地，假设稳态 (stationary) 条件满足，则期望值  $\mu$  为常数。不失普遍性地，可以假设其值为 0，并且有协方差  $\Sigma_{ij} = R[(i - j)\Delta t]$ ，其中  $R(\tau)$  是噪声的自相关函数  $n \star n(\tau)$ 。我们同时定义  $\sum_{l=1}^N \Sigma_{jl} \Sigma_{lk}^{-1} = \delta_{jk}$ 。

由此可以计算

$$\begin{aligned} \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \sum_{j,k=1}^N \Sigma_{jk}^{-1} D_j D_k &= \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \frac{1}{\Delta t^2} \sum_{j,k=1}^N \Delta t^2 \Sigma^{-1}(t_j, t_k) D(t_j) D(t_k) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t^2} \int_{-\infty}^{\infty} dt_k D(t_k) \int_{-\infty}^{\infty} dt_j \Sigma^{-1}(t_j, t_k) D(t_j) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t^2} \int_{-\infty}^{\infty} dt_k D(t_k) \int_{-\infty}^{\infty} df \widetilde{\Sigma^{-1}}(f, t_k) \widetilde{D^*(f)} \end{aligned} \quad (6.33)$$

最后一行用到了 Parseval's theorem。为了进一步简化，需要推导  $\widetilde{\Sigma^{-1}}(f, t_k)$ 。回顾定义，结合  $T \rightarrow \infty$ ,  $\Delta t \rightarrow 0$  的前提，我们有

$$\begin{aligned} \delta_{jk} &= \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \frac{1}{\Delta t} \sum_{l=1}^N \Delta t \Sigma_{jl} \Sigma_{lk}^{-1} \\ &= \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \frac{1}{\Delta t} \int_{-\infty}^{\infty} dt_l R(t_j - t_l) \Sigma^{-1}(t_l, t_k) \end{aligned} \quad (6.34)$$

利用自相关函数的 Fourier 变换是功率谱密度这一特性，考虑对上式引入 Fourier 变换从而得到  $\widetilde{\Sigma^{-1}}(t_j, f)$ ，

$$\begin{aligned} \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \sum_j e^{-i2\pi f t_j} \delta_{jk} &= \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \frac{1}{\Delta t^2} \int_{-\infty}^{\infty} dt_l \int_{-\infty}^{\infty} dt_j e^{-i2\pi f t_j} R(t_j - t_l) \Sigma^{-1}(t_l, t_k) \\ &= \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \frac{1}{\Delta t^2} \int_{-\infty}^{\infty} dt_l e^{-i2\pi f t_l} \Sigma^{-1}(t_l, t_k) \int_{-\infty}^{\infty} d\tau e^{-i2\pi f \tau} R(\tau) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t^2} \widetilde{\Sigma^{-1}}(f, t_k) \frac{1}{2} S_n(|f|) \end{aligned} \quad (6.35)$$

同时，需要注意， $\lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \sum_j e^{-i2\pi f t_j} \delta_{jk} = e^{-i2\pi f t_k}$ ，因此可以得到

$$\widetilde{\Sigma^{-1}}(f, t_k) = \lim_{\Delta t \rightarrow 0} \Delta t^2 \frac{2e^{-i2\pi f t_k}}{S_n(|f|)} \quad (6.36)$$

代入到公式6.33，就可以推导出

$$\begin{aligned} \lim_{\substack{\Delta t \rightarrow 0 \\ T \rightarrow \infty}} \sum_{j,k=1}^N \Sigma_{jk}^{-1} D_j D_k &= 2 \int_{-\infty}^{\infty} df \frac{\tilde{D}^*(f)}{S_n(|f|)} \int_{-\infty}^{\infty} dt_k e^{-i2\pi f t_k} D(t_k) \\ &= 2 \int_{-\infty}^{\infty} df \frac{\tilde{D}^*(f) \tilde{D}(f)}{S_n(|f|)} \\ &= 4 \int_0^{\infty} df \frac{\tilde{D}^*(f) \tilde{D}(f)}{S_n(f)} \end{aligned} \quad (6.37)$$

为了简化表达式，可以定义两个函数  $g(t)$  和  $h(t)$  的内积 inner production

$$\begin{aligned} (g(t) | h(t)) &\equiv 2 \int_0^{\infty} df \frac{\tilde{g}^*(f) \tilde{h}(f) + \tilde{h}^*(f) \tilde{g}(f)}{S_n(f)} \\ &= 4 \Re \int_0^{\infty} df \frac{\tilde{g}^*(f) \tilde{h}(f)}{S_n(f)} \end{aligned} \quad (6.38)$$

其中  $\Re$  表示实部。进一步地，如果  $g$  与  $h$  都是实函数，则有  $\tilde{g}^*(f) = \tilde{g}(-f)$ ，而内积可以简化为

$$(g | h) = 4 \int_0^{\infty} df \frac{\tilde{g}^*(f) \tilde{h}(f)}{S_n(f)} \quad (6.39)$$

通过定义内积这一公式的简化形式，可以将噪声模型的似然函数写为

$$P(D | H_0) \propto \exp \left[ -\frac{1}{2} (D | D) \right]. \quad (6.40)$$

另一方面，由于信号与噪声是线性相加的，我们可以建立替代模型的似然函数为，数据中抽出信号后，残余的部分应该符合噪声模型。换句话说，假设信号  $h$  可以由参数  $\theta$  决

定性地写出  $h(\theta)$ 。则替代模型的似然函数为

$$\begin{aligned} P(D|H_1(\theta)) &\propto \exp\left[-\frac{1}{2}(D - h(\theta)|D - h(\theta))\right] \\ &\propto \exp\left\{-\frac{1}{2}\left[(D|D) + (h(\theta)|h(\theta)) - 2(h(\theta)|D)\right]\right\} \end{aligned} \quad (6.41)$$

将公式6.40与公式6.40 代入似然函数比  $\Lambda$  的表达式<sup>1</sup>，则有

$$\Lambda(\theta) = \frac{P(D|H_1(\theta))}{P(D|H_0)} \propto \exp\left[(h(\theta)|D) - \frac{1}{2}(h(\theta)|h(\theta))\right]. \quad (6.42)$$

其中的系数与参数  $\theta$  无关，因此不会影响最终的探测结果。所谓匹配滤波 (matched filter)，就是指通过不断变更波形模板  $h(\theta)$ ，与数据进行匹配，从而寻找出最佳的参数组合  $\theta$ 。

### 材料 6.6 Wiener optimal filter

可以证明，上述匹配滤波的手段也是 Wiener 最优滤波的结果。考虑对数据  $D$  以未定模板  $K(t)$  进行卷积操作，有

$$D * K(\tau) = \int_{-\infty}^{\infty} [h(t) + n(t)]K(t - \tau)dt \approx S + N \quad (6.43)$$

其中，有

$$S = \int_{-\infty}^{\infty} dt h(t)K(t - \tau) = \int_{-\infty}^{\infty} df \tilde{h}(f)\tilde{K}^*(f) \quad (6.44)$$

$$N = \int_{-\infty}^{\infty} df \tilde{n}(f)\tilde{K}^*(f) \quad (6.45)$$

而噪声的贡献项  $N$  的期望值为 0。同时，可以计算出，其方差为

$$\begin{aligned} \text{Var}(N) &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' K(t)K(t')\langle n(t)n(t') \rangle \\ &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' K(t)K(t') \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \langle \tilde{n}(f)\tilde{n}^*(f') \rangle \exp[-i2\pi(ft - f't')] \\ &= \int_{-\infty}^{\infty} df \frac{1}{2} S_n(f) \tilde{K}(f) \tilde{K}^*(f) \end{aligned} \quad (6.46)$$

由此，可以定义出信噪比  $\rho$

$$\begin{aligned} \rho^2 &= \frac{S^2}{\text{Var}(N)} = \frac{\left(\frac{1}{2} S_n(f) \tilde{K}(f) |\tilde{h}(f)\right)^2}{\left(\frac{1}{2} S_n(f) \tilde{K}(f) \left|\frac{1}{2} S_n(f) \tilde{K}(f)\right.\right)} \\ &\leq \frac{\left(\frac{1}{2} S_n(f) \tilde{K}(f) \left|\frac{1}{2} S_n(f) \tilde{K}(f)\right.\right) (\tilde{h}(f) | \tilde{h}(f))}{\left(\frac{1}{2} S_n(f) \tilde{K}(f) \left|\frac{1}{2} S_n(f) \tilde{K}(f)\right.\right)} \\ &\leq (\tilde{h}(f) | \tilde{h}(f)) \end{aligned} \quad (6.47)$$

<sup>1</sup>有些文献会用符号  $\mathcal{L}$  来表示似然函数比，请注意区分。

上式第一行到第二行利用了 Cauchy-Schwarz 不等式，由此，只有当

$$\tilde{K}(f) = \frac{\tilde{h}(f)}{S_n(f)} \quad (6.48)$$

时，等式才会成立。最优滤波，指的是对于输入信号  $h(t)$  而言，只有用  $\frac{\tilde{h}(f)}{S_n(f)}$  对应的逆 Fourier 变换的模板去响应，才能在统计上预期获得最大的信噪比，而这时的最优信噪比

$$\rho_{\text{opt}}^2 = (\tilde{h}(f)|h(f)) \quad (6.49)$$

### 材料 6.7 内积操作的线性代数理解

对于内积操作，还可以从比较单纯的线性代数角度理解。若矩阵在每个斜线方向上有相同的值，

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \dots & \dots & a_2 & a_1 & a_0 \end{bmatrix} \quad (6.50)$$

则称这种性质的矩阵称为 Toeplitz 矩阵，又称常对角矩阵。对于高斯稳态的时序列噪声，其协方差矩阵  $\Sigma$  满足  $\Sigma_{ij} = R[(i - j)\Delta t]$ ，因此是 Toeplitz 矩阵。

有一种特殊的 Toeplitz 矩阵，若每一列（行）都为前一列（行）挪动一个元素得到，

$$\mathbf{C} = \begin{bmatrix} c_0 & c_{n-1} & c_{n-2} & \dots & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ c_2 & c_1 & c_0 & c_{n-1} & \\ \vdots & & & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \dots & c_0 \end{bmatrix} \quad (6.51)$$

则称为循环矩阵。

可以证明，循环矩阵的特征值  $\lambda_j$  可以由行向量的 Fourier 变换求得

$$\lambda_j = c_0 + c_{n-1} \exp\left(-i2\pi j \frac{1}{n}\right) + c_{n-2} \exp\left(-i2\pi j \frac{2}{n}\right) + \dots + c_1 \exp\left(-i2\pi j \frac{n-1}{n}\right) \quad (6.52)$$

对于维度足够高的循环矩阵，可以将其对角化

$$\mathbf{C} \approx \mathbf{F}^* \mathbf{\Lambda} \mathbf{F} \quad (6.53)$$

其中  $\mathbf{F}$  是 Fourier 变换矩阵， $F_{jk} = \frac{1}{n} \exp\left(-i2\pi \frac{jk}{n}\right)$ ， $\mathbf{\Lambda} = \text{diag}(\lambda_j)$ 。

虽然协方差矩阵并不自然地满足循环矩阵条件，但是  $\Sigma \approx \mathbf{F}^* \mathbf{\Lambda} \mathbf{F}$  的近似依然足够准确。因此，似然函数的对数可以近似为

$$\log P(\mathbf{D} | \boldsymbol{\mu} = \mathbf{0}, \Sigma) \approx -\frac{1}{2} \left[ n \log 2\pi + \log |\Sigma| + \tilde{\mathbf{D}}^T \mathbf{\Lambda}^{-1} \tilde{\mathbf{D}} \right] \quad (6.54)$$

不难看出，这一表达式与内积的表达式（公式 6.37）一致。公式 6.54 又称为 Whittle likelihood 或者 Whittle 近似。

## 6.4 连续引力波探测

### 6.4.1 $\mathcal{F}$ -统计

让我们考虑最简单的连续引力波情形，在第 3.5 节中我们进行过一些讨论。由于探测器相对地月系质心有运动，而地月系又有相对太阳系质心的运动，这些运动引起的 Doppler 频移会使引力波的波形进一步复杂化。因此，一般将讨论放在太阳系质心（solar system barycentre, SSB）考虑，再转换到具体的探测器坐标系中。如第 3.5 节所述，连续引力波源往往可以由振幅不变，频率近似不变的三角函数去描述。这时，源信号可以表述为

$$\begin{aligned} h_+(\tau) &= A_+ \cos \Phi(\tau) \\ h_\times(\tau) &= A_\times \sin \Phi(\tau) \end{aligned} \quad (6.55)$$

将瞬时频率  $f(\tau)$  定义为  $2\pi f(\tau) \equiv d\Phi(\tau)/d\tau$ 。这样，在选定频率零点以后，就可以将相位  $\Phi$  在零点  $\tau_{\text{ref}}$  附近进行 Taylor 展开，即

$$\begin{aligned} \Phi(\tau) &\equiv \phi_0 + \phi(\Delta\tau) \\ \phi_0 &\equiv \Phi(\tau_{\text{ref}}) \\ \phi(\Delta\tau) &\equiv 2\pi \sum_{s=0} \frac{f^{(s)}(\tau_{\text{ref}})}{(s+1)!} (\Delta\tau)^{s+1} \end{aligned} \quad (6.56)$$

其中,  $f^{(s)}(\tau) \equiv \frac{d^s f(\tau)}{d\tau^s}|_{\tau_{\text{ref}}}$  是瞬时频率  $f(\tau)$  在零点  $\tau_{\text{ref}}$  时刻的  $s$  阶导数。在  $t$  时刻, 对单位方向向量为  $\hat{n}$  的信号来说, SSB 和探测器  $X$  之间的转换主要集中于时间  $\tau$  的转换关系

$$\Delta\tau^X(t; \hat{n}) \equiv \tau^X - \tau_{\text{ref}} \approx t - \tau_{\text{ref}} + \frac{\vec{r}^X(t) \cdot \hat{n}}{c} \quad (6.57)$$

其中,  $c$  表示光速。

注意到, 上述的表述中, 我们将信号的来源  $\hat{n}$  作为参数进行考虑, 但式4.23表述, 探测器对于给定信号的响应还依赖于极化角  $\psi$  这一参数。可以将天线响应函数分解为含时和不含时的两个部分

$$\begin{aligned} F_+^X &= a^X(t; \hat{n}) \cos 2\psi + b^X(t; \hat{n}) \sin 2\psi \\ F_\times^X &= b^X(t; \hat{n}) \cos 2\psi - a^X(t; \hat{n}) \sin 2\psi \end{aligned} \quad (6.58)$$

由此, 探测器记录到的波形总共可以由四个不含时间变化的参数, 即  $A_+$ ,  $A_\times$ , 初始相位  $\phi_0$  和偏振角  $\psi$ 。可以定义四个与探测器无关的新参数,

$$\begin{aligned} \mathcal{A}^1 &\equiv A_+ \cos \phi_0 \cos 2\psi - A_\times \sin \phi_0 \sin 2\psi \\ \mathcal{A}^2 &\equiv A_+ \cos \phi_0 \sin 2\psi + A_\times \sin \phi_0 \cos 2\psi \\ \mathcal{A}^3 &\equiv -A_+ \sin \phi_0 \cos 2\psi - A_\times \cos \phi_0 \sin 2\psi \\ \mathcal{A}^4 &\equiv -A_+ \sin \phi_0 \sin 2\psi + A_\times \cos \phi_0 \cos 2\psi \end{aligned} \quad (6.59)$$

对应地, 可以定义参数集  $\lambda \equiv \{\hat{n}, f^{(s)}(\tau_{\text{ref}})\}$ , 并将探测器记录到的信号分解

$$h^X(t; \mathcal{A}, \lambda) = \sum_{\mu=1}^4 \mathcal{A}^\mu h_\mu^X(t; \lambda) \quad (6.60)$$

其中

$$\begin{aligned} h_1^X(t; \lambda) &= a^X(t) \cos \phi(\Delta\tau^X) \\ h_2^X(t; \lambda) &= b^X(t) \cos \phi(\Delta\tau^X) \\ h_3^X(t; \lambda) &= a^X(t) \sin \phi(\Delta\tau^X) \\ h_4^X(t; \lambda) &= b^X(t) \sin \phi(\Delta\tau^X) \end{aligned} \quad (6.61)$$

由于可以将信号做如式6.60般的拆分, 因此, 代入似然函数比 (公式6.42) 后, 可以得到,

$$\begin{aligned} \log \mathcal{L}(D; \mathcal{A}, \lambda) &= \left(h \middle| D\right) - \frac{1}{2} \left(h \middle| h\right) \\ &= \mathcal{A}^\mu D_\mu - \frac{1}{2} \mathcal{A}^\mu \mathcal{M}_{\mu\nu} \mathcal{A}^\nu \end{aligned} \quad (6.62)$$

这其中，我们定义

$$D_\mu(\lambda) \equiv (h_\mu | D) \quad (6.63)$$

$$\mathcal{M}_{\mu\nu}(\lambda) \equiv (h_\mu | h_\nu) = (\partial_\mu h | \partial_\nu h) \quad (6.64)$$

上述表述用到了6.60中不同分量间线性独立的特性。

通过对式子6.62求最大值，不难发现似然函数比最大时，其对  $\mathcal{A}$  的偏导皆为 0，于是可以定义所谓  $\mathcal{F}$ -统计，也就是

$$\mathcal{F}(D; \lambda) \equiv \max_{\mathcal{A}} \log \mathcal{L}(D; \mathcal{A}, \lambda) = \frac{1}{2} D_\mu \mathcal{M}^{\mu\nu} D_\nu \quad (6.65)$$

注意，我们这里定义  $\mathcal{M}^{\mu\sigma} \mathcal{M}_{\sigma\nu} = \delta_\nu^\mu$ ，也就是两者互为逆矩阵。

### 材料 6.8 F-statistic

$\mathcal{F}$ -统计是专门针对引力波数据处理开发的处理方法，由 Jaranowski、Królak 和 Schutz 在 1998 年提出，因此常将该论文称为 JKS98[53]。它与统计学中的 F-统计有区别。统计学中常用的 F-统计主要用于所谓的 F-检验 (F-test)，通过将数据分组后，计算出组间标准差与组内标准差的比值，这一比值也被称为 F-统计量。通过理论推导，可以计算出在高斯噪声的前提下 F-统计的分布，如果不同组别的数据有显著的分布差异，则 F-统计量将倾向于超出预设阈值，从而拒绝零假设。

我们发现，最大似然函数比条件下，可以推导出振幅  $\mathcal{A}^\mu$  的估计值为

$$\mathcal{A}_{\text{MLR}}^\mu = \mathcal{M}^{\mu\nu} D_\nu \quad (6.66)$$

换言之，也可以将  $\mathcal{F}$ -统计转换为

$$2\mathcal{F} = \mathcal{A}_{\text{MLR}}^\mu \mathcal{M}_{\mu\nu} \mathcal{A}_{\text{MLR}}^\nu \quad (6.67)$$

#### 6.4.1.1 振幅调制

由于探测器相对太阳系质心的运动，与探测器绑定的探测器坐标系随时间变化，由此，探测器对给定引力波源的响应也随时间变化。这种响应的改变引起了振幅的调制，但若观测时长远大于调制周期，则可以用平均化的结果代替。注意到，由于连续引力波信号的频率  $f_s$  基本不随时间变化，而功率谱密度在较小的范围内可以看成常数，所以相应的内积可以近似为

$$(a|b) = 4\Re \int_0^\infty df \frac{\tilde{a}(f)^* \tilde{b}(f)}{S_n(f)} \simeq \frac{2}{S_n(f_s)} \int_0^T dt a(t)b(t). \quad (6.68)$$

基于天琴星座平面固定指向的特点，可以给出天琴对双白矮星引力波信噪比的解析表达式，信噪比的定义为信号与自身的内积

$$\rho^2 = (h|h) \simeq \frac{2\langle A^2 \rangle T}{S_n(f_s)}, \quad (6.69)$$

其中平均振幅定义为

$$\langle A^2 \rangle = \frac{1}{T} \int_0^T h^2(t) dt. \quad (6.70)$$

同时有

$$\langle A^2 \rangle \approx \frac{3}{16} \mathcal{A}^2 [(1 + \cos^2 \iota)^2 \langle F_+^2 \rangle + 4 \cos^2 \iota \langle F_x^2 \rangle] \quad (6.71)$$

$$\langle F_+^2 \rangle = \frac{1}{4} (1 + \cos^2 \theta_S)^2 \cos^2 2\psi_S + \cos^2 \theta_S \sin^2 2\psi_S \quad (6.72)$$

$$\langle F_x^2 \rangle = \frac{1}{4} (1 + \cos^2 \theta_S)^2 \sin^2 2\psi_S + \cos^2 \theta_S \cos^2 2\psi_S \quad (6.73)$$

这些近似的表达式可以直接用来在时域上解析的计算信噪比，而不用在频域上计算数值积分，这样有助于缩短计算时间。在推导公式 6.71 时，我们舍去了积分后的时间的一阶小量  $\sim O(T)^{-1}$ 。

#### 6.4.1.2 Doppler 频移和相位调制

如前所述，探测器相对于地月系质心和 SSB 都会有运动，这种运动会引入 Doppler 频移和相位调制。

$$f = f_0 \left( 1 + \frac{\mathbf{v} \cdot \hat{\mathbf{n}}}{c} \right) \quad (6.74)$$

其中， $\mathbf{v}$  是探测器的速度。随着时间变化，频率的变化可以表述为速度变化的函数

$$(\Delta f)_{\text{Doppler}} = f_0 \frac{(\Delta v)_T}{c} \quad (6.75)$$

这种相位调制的引入，会引起原本集中在某个频率处的信号向边带（sideband）溢出能量，从而削弱信号。对于给定方向的源，可以通过重采样（resample）的方法人为地在数据中消除 Doppler 频移的影响，从而得到完全单频的信号。即，不对原时间  $t$  均匀采样，而是对  $t' = t + \frac{\mathbf{r} \cdot \hat{\mathbf{n}}}{c}$  采样。将一个时域数据进行离散 Fourier 变换到频率域后，在频率域上离散的数据点对应的频率为  $\Delta f, 2\Delta f, 3\Delta f, \dots, N/2\Delta f$ ，其中  $N$  是时域采样数， $\Delta f = 1/T$ ，不难得出， $N\Delta f = N/T = f_{\text{sample}}$  是时间域的采样率。

如果 Doppler 频移在频率域上的影响低于频率分辨率，则意味着所有的效应都在同一个频率间隔（frequency bin）内，可以不用考虑。反之，则同一频率的信号分散在多个频率间隔中，需要考虑 Doppler 频移的效应。

### 6.4.1.3 未知源搜索

一些中子星或者双白矮星系统，可能已经被电磁波手段探测到，并精确测量出其空间位置、频率、频率演化等各类参数。针对这类已知源，通过  $\mathcal{F}$ -统计的方法在数据中搜索信号时，实际上就是在外禀参数的参数空间上解析地求似然函数比的极值。然而，我们预期数据中包含着许多之前未知的引力波信号。对这类未知源，则需要通过不断改变内禀参数的组合，对数据进行匹配，进而实现对信号的探测。

我们考虑空间上两个相距  $\Delta\theta$  的点  $A$  和  $B$ 。由于位置不一样，导致用  $A$  点的波形去匹配  $B$  点的数据时，会产生  $|\Delta f| = fv/c \cdot (\hat{n} - \hat{n}') \lesssim fv/c\Delta\theta$  的频率调制。考虑数据的频率分辨率，即可估算出在搜索未知源时对空间参数的允许偏移量。而若要覆盖全天，则需要共  $N_{\text{sky}}$  个独立波形的尝试

$$N_{\text{sky}} \sim \frac{4\pi}{(\Delta\theta)^2} \sim 4\pi T^2 f_{\max}^2 (v/c)^2 \quad (6.76)$$

代入一年的观测时间，1Hz 的采样频率，和地球运动速度  $v \sim 30 \text{ km s}^{-1}$  后，可以计算出  $N_{\text{sky}} \sim 10^8$ 。考虑到对每个方向，都需要将数据转换到 SSB 后作 Fourier 变换，而 Fourier 变换的计算量  $\propto T f_{\max}$ ，故总计算量可估算出  $\propto T^3 f_{\max}^3 (v/c)^2$ 。注意上述计算基于单频近似，即  $f^{(s)} = 0$ ，如果考虑频率演化，则总的计算数会进一步以正比于  $T^2$  的速率增长，最终使计算量  $\propto T^5$ 。

我们可以看到，积分时间越长，搜寻信号所需的计算量越大。因此，比较实际的方法是只对一小段数据进行  $\mathcal{F}$ -统计的计算，然后相加，即非相干搜索；或者再对不同数据段之间进行自洽性检验，即所谓的半相干搜索方法。

对于非相干搜索/半相干搜索而言，信号频率自身的变化会影响搜寻能力。如果信号存在频率漂移  $\dot{f}$ ，则对不同时间中同一频率处对应的  $2\mathcal{F}$  求和，可能会将信号与噪声叠加而错失信号。可以利用 Hough 变换，实现对有频率变化的信号的探测。Hough 变换为 Paul Hough 在 1959 年发明，主要为了分析 CERN 的气泡室中的粒子轨迹。在 Hough 变换中，首先得到时频图，然后对每一个超出阈值的视频图像素标黑，然后用模式识别算法判断是否“黑”的像素均落在某个曲线附近。对于 Gaussian 噪声，Hough 变换的方法会丢失信息，因此损失灵敏度；但真实数据会存在非 Gaussian 性质，因此实际应用中常常会用到 Hough 变换。

### 6.4.1.4 拟合度检验

在数据处理中，经常需要进行拟合好坏程度（goodness-of-fit）的判断的，通常会使用的判断是  $\chi^2$  检验，和 Kolmogorov-Smirnov 检验，或者简称为 KS 检验。

定义  $\chi^2 = \sum_{i=1}^{\nu} \frac{(x_i - \mu_i)^2}{\sigma^2}$  其中  $\mu_i$  是每个点的期望值，是当数据  $x_i$  源自独立的 Gaussian 分布时，则计算得到的统计量  $\chi^2$  符合自由度为  $\nu$  的中心  $\chi^2$  分布。一个实用的场景是，将数据按照不同的数值进行分隔，然后对每个间隔内的数据数进行计数。假设在第  $i$  个间隔内，期望数是  $e_i$ ，而实际计数是  $o_i$ ，则构造出的统计量

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (6.77)$$

由于中心极限定理的存在，可以证明，统计量  $\chi^2$  符合自由度为  $n-1$  的  $\chi^2$  分布。

Kolmogorov-Smirnov 检验通常用于比较两个分布之间（在存在随机起伏的前提下）是否一致。假设有累计分布函数  $F$ ，并有中采样的随机变量  $X_i (i = 1, \dots, n)$ ，则通过采样点重构的累计分布函数  $F_n(x)$ ：

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \quad (6.78)$$

这其中， $I_{(-\infty, x]}$  是定义在  $(-\infty, x]$  上的指示函数（indicator function）

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \quad (6.79)$$

由此可以构造出检验统计量

$$D_n \equiv \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \quad (6.80)$$

其中  $\sup S$  是在集合  $S$  上的最小上界。换言之，检验统计量是数据构造出的累计分布与零假设的累计分布之间的最大距离。根据 Glivenko-Cantelli 定理，可以证明，当  $n \rightarrow \infty$  时， $D_n$  趋向于 0。

更进一步地，Kolmogorov 还给出了  $D_n \rightarrow 0$  的收敛速度。可以证明，在零假设下，当  $n \rightarrow \infty$  时，统计量  $\sqrt{n}D_n$  收敛到随机变量  $K$  其中， $B(t)$  是 Brownian 桥随机过程。

$$K \equiv \sup_{t \in [0, 1]} |B(t)| \quad (6.81)$$

对应的累计概率分布为

$$\Pr(K \leq x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} e^{-(2k-1)^2 \pi^2 / (8x^2)}, \quad (6.82)$$

在置信度  $\alpha$  的水平上，当  $\sqrt{n}D_n > K_\alpha$  时，拒绝零假设。其中

$$P(K \leq K_\alpha) = 1 - \alpha \quad (6.83)$$

### 6.4.2 误警率和探测概率

给定零假设下，概率分布  $p_0(\mathcal{F})$ ，可以计算出误警率 (false alarm probability)。根据定义，误警率就是数据符合零假设的前提下，超出某预设探测阈值  $\mathcal{F}_0$  的概率：

$$p_F(\mathcal{F}_0) = \int_{\mathcal{F}_0}^{\infty} d\mathcal{F} p_0(\mathcal{F}) \quad (6.84)$$

零假设下， $p_0(2\mathcal{F})$  服从自由度为 4 的中心  $\chi^2$  分布。

类似的，探测概率可以定义为，替代假设（即，数据中包含信噪比为  $\rho$  的引力波信号）成立的前提下，探测概率  $p_D$  可以定义为

$$p_D(\rho, \mathcal{F}_0) = \int_{\mathcal{F}_0}^{\infty} d\mathcal{F} p_1(\rho, \mathcal{F}) \quad (6.85)$$

而替代假设的概率分布  $p_1(\rho, 2\mathcal{F})$  为自由度为 4，非中心参数为  $\rho^2$  的非中心  $\chi^2$  分布。

### 6.4.3 次优滤波

通常而言，在使用匹配滤波方法从数据中提取引力波信号时，用于搜索的模板并不能完美地匹配真实信号，而是会有一定的偏差。可以定义匹配因子 fitting factor  $FF$ ，对于输入信号  $h(\theta)$  而言，有

$$FF(\theta) \equiv \max_{\theta'} \frac{(h(\theta) | h(\theta'))}{\sqrt{(h(\theta) | h(\theta))(h(\theta') | h(\theta'))}} \quad (6.86)$$

以及错配 mismatch  $\mu = 1 - FF$ 。在低阶近似下，可以预期，采用次优滤波时，预期得到的信噪比与最优信噪比之间的比值为  $FF$ 。

在设计波形模板库进行匹配滤波的数据分析过程时，通常以匹配因子作为重要考量。原则上，对于任一输入波形  $h(\theta)$ ，都存在与之相距不远的次优滤波波形，使得  $FF(\theta) \geq 0.97$ 。0.97 的阈值的选取是一个约定俗成的取值，对应的探测能力的损失为  $\sim 10\%$ 。

## 6.5 喳啾信号探测

### 6.5.1 new SNR

在真实数据中，噪声不一定真的具有 Gaussian、稳态的性质，因此，通过匹配滤波手段获得较高的信噪比的情况下，也不能直接得出信号真实存在的结论。

对于啁啾信号（chirp signal，波形描述见第3.2.1节）而言，由于其信号在可探测频带中频率分布较广，因此可以利用  $\chi^2$  检验这一简单有效的手段，对探测结果的可靠性进行校验。

通过理论波形，结合探测器的噪声曲线，就可以计算出给定信号的信噪比在频率上的分布。定义出理论上预期的频率域上的分段信噪比  $\rho_i$

$$\rho_i^2 = 4 \int_{f_{i-1}}^{f_i} df \frac{\tilde{h}^*(f)\tilde{h}(f)}{S_n(f)} \quad (6.87)$$

据此，可以将频率分割为  $p$  份，计算出相应的  $f_0 < f_1 < f_2 < \dots < f_p$ ，使得每一段的信噪比都相同，即

$$\rho_i^2 = \frac{\rho^2}{p} \quad (6.88)$$

可以将实际数据代入，得到对应的  $z_i$

$$z_i = 4 \Re \int_{f_{i-1}}^{f_i} df \frac{\tilde{D}^*(f)\tilde{h}(f)}{S_n(f)} \quad (6.89)$$

更进一步的，可以定义出统计量  $z$ ，

$$z = \sum_{i=1}^p z_i = (D|h) \quad (6.90)$$

注意，在  $\rho^2$  和  $\rho_i$  的计算中，只涉及到波形和理论的功率谱密度，而不涉及真实数据。而在统计量  $z$  的计算中，需要涉及具体的数据，而其和  $\rho_i$  的联系主要集中于  $f_i$  的计算中。

根据上述计算，可以在每个频率段  $[f_{i-1}, f_i)$  上计算出  $\Delta z_i = z_i - z/p$ ，这样一定有  $\sum_{i=1}^p \Delta z_i = 0$ ，而其统计学期望为

$$\langle \Delta z_i \rangle = 0 \quad (6.91)$$

不难发现，在信号的信噪比均匀分布、噪声的性质符合 Gaussian、稳态时， $\chi^2 = p \sum_{i=1}^p (\Delta z_i)^2$  符合自由度为  $p-1$  的  $\chi^2$  分布。反过来，如果数据出现异常，由噪声给出了极高的似然函数比，在绝大部分情况下，其频率上的分布和正常的引力波波形有着较大的差别，那么利用上述过程构造的统计量  $\chi^2$  将不符合自由度为  $p-1$  的  $\chi^2$  分布。利用这种方法，定义出的  $\rho_{\text{new}}$ ，

$$\rho_{\text{new}} = \begin{cases} \rho, & \text{if } \chi^2 \leq 1 \\ \rho [\frac{1}{2} (1 + (\chi^2)^3)]^{-1/6}, & \text{if } \chi^2 > 1 \end{cases}$$

可以有效地降低非 Gaussian、非稳态噪声对探测过程的影响。实际上，[16] 首先在 1996 年提出这个方法，随后在实践中做了很多优化，而大体上这种简单的策略依然保留了下来，成为 LIGO 等地面引力波探测器对啁啾信号的数据处理过程的重要一步。

### 6.5.2 匹配滤波的统计学性质

注意到，似然函数比的定义式为

$$\ln \Lambda(\theta) = \left( h(\theta) \middle| D \right) - \frac{1}{2} \left( h(\theta) \middle| h(\theta) \right).$$

因此，我们有必要讨论数据与波形模板内积  $x = \left( h(\theta) \middle| D \right)$  的统计学性质。注意到，通常会采用归一化的模板  $g$  进行分析，其中最优信噪比  $\rho_{\text{opt}}^2 = (g|g) = 1$ 。基于内积，可以讨论对数据  $D$  用波形  $g$  进行匹配滤波  $x = (D|g)$  时，统计量  $x$  的统计学性质。在没有信号的时候，数据  $D$  等于噪声  $n$ ，因此有  $\langle x \rangle = \langle (n|g) \rangle = 0$ ，并且有

$$\begin{aligned} \langle x^2 \rangle &= \left\langle 2 \int_{-\infty}^{\infty} df \frac{\tilde{n}^*(f)\tilde{g}(f)}{S_n(f)} 2 \int_{-\infty}^{\infty} df' \frac{\tilde{n}(f')\tilde{g}^*(f')}{S_n(f')} \right\rangle \\ &= 4 \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \frac{\langle \tilde{n}^*(f)\tilde{n}(f') \rangle \tilde{g}(f)\tilde{g}^*(f')}{S_n(f)S_n(f')} \\ &= 4 \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \frac{\frac{1}{2}S_n(f)\delta(f-f')\tilde{g}(f)\tilde{g}^*(f')}{S_n(f)S_n(f')} \\ &= 2 \int_{-\infty}^{\infty} df \frac{\tilde{g}(f)\tilde{g}^*(f')}{S_n(f)} \\ &= (g|g) \end{aligned} \quad (6.92)$$

由此，当没有信号的时候，统计量  $x$  服从期望值为 0，方差  $\sigma^2 = (g|g) = 1$  的 Gaussian 分布。

考虑数据中包含真实信号，其波形与归一化的模板一致，仅在振幅上相差  $A$  倍，则有信号  $h(t) = Ag(t)$ ，数据  $D(t) = n(t) + h(t) = n(t) + Ag(t)$ 。则  $x$  的期望值为

$$\langle x \rangle = \langle (D|g) \rangle = \langle (n|g) \rangle + (h|g) = A(g|g) = A \quad (6.93)$$

更进一步地，有

$$\begin{aligned} \langle x^2 \rangle &= \langle (D|g)^2 \rangle = \langle [(n|h) + (g|h)]^2 \rangle \\ &= \langle (n|h)^2 \rangle + 2(g|h)\langle (n|h) \rangle + (g|h)^2 \\ &= 1 + A^2 \end{aligned} \quad (6.94)$$

对应，可以计算出  $x$  的方差

$$\text{Var}(x) = \langle x^2 \rangle - \langle x \rangle^2 = 1 + A^2 - A^2 = 1 \quad (6.95)$$

我们习惯将信号的信噪比定义为信号的期望值比上噪声的标准差。可以发现，由此定义出来的信噪比正是  $A$ ，与公式 6.49  $\rho_{\text{opt}} = \sqrt{(h(f)|h(f))}$  一致。

### 6.5.3 时间未知的匹配滤波

考虑真实的信号  $h(t)$  与模板波形之间相差  $t_0$ , 亦即,  $h(t) = Ag(t - t_0)$ 。不难知,

$$\tilde{h}(f) = A\tilde{g}(f)e^{-i2\pi f t_0} \quad (6.96)$$

由此, 可以得到, 进行模板匹配时, 对应的是

$$(D|h) = 4A\Re \int_0^\infty df \frac{\tilde{D}(f)\tilde{g}^*(f)}{S_n(f)} e^{i2\pi f t_0} = Ax(t_0) \quad (6.97)$$

其中,

$$x(t) \equiv 4\Re \int_0^\infty df \frac{\tilde{D}(f)\tilde{g}^*(f)}{S_n(f)} e^{i2\pi f t} \quad (6.98)$$

在最大似然函数处, 估算得到的振幅为  $A_{\max} = \rho(t_{\text{peak}})$ 。形式上来看,  $x(t)$  是一个频率域函数的 Fourier 变换, 利用 FFT 算法可以以较快的速度对不同时间参数进行搜索, 因此这一事实可以加快数据处理过程。

### 6.5.4 匹配滤波模板库

对于一组参数  $\theta$ , 可以计算出一个似然函数比的数值, 但真实的信号对应的物理参数对数据处理器而言是未知的, 怎样才能避免错过探测? 对此, 一个重要的手段就是构建模板库 (template bank) 来进行匹配滤波。在参数空间上, 选取一组格点, 每一个格点对应一组参数  $\theta$ , 这组格点覆盖整个参数空间。根据这些格点, 可以分别计算出归一化的波形模板  $g(\theta)$ , 使得  $(g|g) = 1$ 。可以预期, 能够给出最高响应的模板  $\theta'$  应该与原参数最为接近, 设两者有偏离

$$\Delta\theta = \theta' - \theta \quad (6.99)$$

两者的内积定义为混淆函数 (ambiguity function)

$$A = (h(\theta)|h(\theta')) \quad (6.100)$$

注意到, 对波形做 Taylor 展开, 可以得到

$$h(\theta') = h(\theta) + \Delta\theta_i \frac{\partial h(\theta)}{\partial \theta_i} + \frac{1}{2} \Delta\theta_i \Delta\theta_j \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j} + O(\Delta\theta^3) \quad (6.101)$$

代入公式 6.100 后, 可以得到

$$\begin{aligned} A &= \left( h(\theta) \left| h(\theta + \Delta\theta_i \frac{\partial h(\theta)}{\partial \theta_i} + \frac{1}{2} \Delta\theta_i \Delta\theta_j \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j}) \right. \right) + O(\Delta\theta^3) \\ &= 1 + \frac{1}{2} \Delta\theta_i \Delta\theta_j \left( h(\theta) \left| \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j} \right. \right) + O(\Delta\theta^3) \end{aligned} \quad (6.102)$$

注意，其中线性项不贡献，因为  $\boldsymbol{\theta}'$  能够给出最高响应，所以  $\left(h(\boldsymbol{\theta})\left|\frac{\partial h(\boldsymbol{\theta})}{\partial \theta_i}\right.\right) = 0$ 。

可以将混淆函数写为

$$A = 1 - ds^2 \quad (6.103)$$

这其中

$$ds^2 = g^{ij} d\theta_i d\theta_j \quad (6.104)$$

描述了信噪比相对降低值的期望。<sup>1</sup> 而  $g^{ij}$  则是在这个参数空间上定义的度规

$$g^{ij} = -\frac{1}{2} \left( h(\boldsymbol{\theta}) \left| \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right. \right) \quad (6.105)$$

在参数空间  $\Omega$  上，波形模板的分布应该符合让  $FF(\boldsymbol{\theta}) \geq 0.97$ ，或者让 mismatch  $ds^2 \leq 0.03$ ，换言之，定义最大 mismatch( $ds_{\max}$ )<sup>2</sup> = 3% 若要对波形模板的数量作量级估算，可以采用

$$N \sim \frac{1}{(ds_{\max})^{\dim \Omega}} \int_{\Omega(\boldsymbol{\theta})} \sqrt{\det g^{ij}} d\boldsymbol{\theta} \quad (6.106)$$

### 6.5.5 同时性

由于用于匹配滤波的模板库数量巨大，即使数据中完全不包含噪声，也可能会产生较高的响应。除了利用  $\chi^2$  检验生成  $\rho_{\text{new}}$ ，以降低非 Gaussian 噪声对探测过程的干扰外，还可以利用一些额外信息提高探测结果的可信度。首先就是探测结果的同时性。

噪声和信号的一个重要特征是，物理上相隔较远的两个探测器，噪声通常可以视为不相关。而信号由于本质上是天文学起源，所以通常抵达多个探测器时，信号有极强的相关性。LIGO 建立两个独立的探测器（距离极远，且指向相近）的重要动机，就是借助这一特性，而仅对在两个探测器上同时留下响应的候选体进行下一步的分析。

上述对信号进行搜索的过程，基本上都针对一个探测器进行。通过对两个探测器所报告的探测结果进行比对，特别是对报告的并合时间  $t_c$  参数比对，可以排除大量由噪声引发的误警。实际上，由于探测器之间存在一定距离  $\vec{r}$ ，因此，两个探测器之间测量到的  $t_c$  之间的时间差应该为  $\Delta t = \vec{r} \cdot \hat{n}/c$ ，其中  $\hat{n}$  是引力波源方向的单位向量。考虑到地球的尺寸这个时间最长可以设置为 40ms。除了时间信息以外，质量参数的一致性也是一个重要判断标准。即使两个探测器同时报告探测结果，如果最高响应的噪声参数间存在较大的不一致，同样应该将其排除。

对于空间引力波探测器而言，如第4.3.1节描述，一个三角形构型的激光干涉引力波探测器，可以等效为两个正交的 Michelson 构型的干涉仪。因此，上述的同时性的概念同样可以运用于空间引力波探测中。

<sup>1</sup>注意，Allen et al. 2019 [17] 指出， $ds^2 = \sin^2 \sqrt{g^{ij} d\theta_i d\theta_j}$  是一个更好的近似。

### 6.5.6 显著度分析

如同第6.3.1节所述，通常把 type I error 对应的概率被称为检验的显著性。对于引力波数据而言，如何计算？

对于一般的物理实验而言，通常的做法是，首先将信号屏蔽，使仪器只记录噪声，这时计算出的探测统计量可以明确来自噪声，对应可以求出相应分布。由于显著性计算的是 type I error，所以只需要噪声分布的信息即可。接下来将屏蔽撤除，记录数据。这时候将候选信号对应的探测统计量和噪声的分布进行比对、计算，可以计算出相应的 p 值，也就是对应的 Fisher 方法（见第6.3.1.5节）。将计算出的 p 值与预设的显著度阈值  $\alpha$  进行比较，一旦  $p < \alpha$ ，则拒绝噪声的零假设；反之，则接受零假设。对于常见的情形，通常以  $3 - \sigma$  或 3 sigma 来确定显著度，也就是

$$\alpha = 1 - \int_{\mu-3\sigma}^{\mu+3\sigma} dx P(x|\mu, \sigma) \quad (6.107)$$

作为判据。具体数值为 0.27%，换言之， $1 - \alpha = 99.73\%$ 。在物理学领域，许多重要的探测都以  $5\sigma$  或 5sigma 作为判据。需要注意的是，这里的定义略有不同，为

$$\alpha = 1 - \int_{-\infty}^{\mu+5\sigma} dx P(x|\mu, \sigma) \quad (6.108)$$

对应的数值为  $2.8 \times 10^{-7}$ 。

对于引力波信号而言，地面引力波探测器无法实现对引力波信号的屏蔽，所以必须用替代的方法获得噪声对应的探测统计量分布。一个常用的方法就是时间移动 (time slide)，通过人为调动探测器的时间，利用前述的同时性判据，可以获得一组新的信号候选。具体来说，通过将探测器数据平移如 5s 的时间，利用同时性判据，这时获得的同时事件，不可能来自同一个信号时间（注意，该平移量必须大于地球尺寸引起的 40ms），因而成为背景事件。这样，等效地对引力波信号实现了屏蔽。通过不断的将数据平移，可以创造出等效时间相当长的有效数据来。在引力波事件 GW150914 的探测中报告的误警率低于 20300 年一次，正是由此而来。

对于引力波事件候选体的显著性计算而言，最大的争端在于在统计“背景”的探测统计量分布时，是否需要将被测试的候选体剔除。剔除这一过程必定会将显著度提升，使报告结果偏向非保守的一侧。另一方面，既然这些候选体被作为信号考虑，将其放到背景的估计中，会导致对 p 值的估计不再是无偏的。

面对这种两难的问题，可以通过设置模拟数据挑战 (mock data challenge, MDC) 的方法解决。<sup>[33]</sup> 就通过设立 MDC，对这一问题进行了定量的回答，最后结论是，由于对实际的探测结果而言，感兴趣的 p 值都非常小（显著度在 5sigma 上下），而 p 值又是

非负的，所以虽然移除候选体的操作会使得对  $p$  值的估计不再具有无偏特性，但其代价是，从统计上来说，平衡一次保守方向的偏差，需要许多次激进的估计来平衡，也就意味着，大多数概率下，估计出来的  $p$  值都是偏激进的。而实际上，不移除候选体的时候，虽然在线性下  $p$  值的估计有偏，但是倾向于保守，且  $p$  值的概率分布中位数是无偏的。因此，对于引力波探测而言，在保守偏好的前提下，通常倾向于选择在估计背景时不移除候选体。

### Exercise 6.1

作真值表，证明逻辑上  $A, (B + C) = A, B + A, C$ ，亦即逻辑与的分配律。

### Exercise 6.2

证明 Parseval 定理。

### Exercise 6.3

给定特定引力波源以及对应的物理参数，可以计算出一个引力波探测器对此类源的极限探测距离，这一距离通常被定义为视界距离 **horizon distance**。对应的情形为，信号处于探测器坐标系中的天顶位置处，且轨道倾角取  $\iota = 0$ ，极化角  $\psi = 0$ 。换言之，响应后的信号为  $h = F^+ h_+ = h_+$ 。通常情况下，选取单个 Michelson 干涉仪的信噪比  $\rho_{\text{opt}} = 8$  作为探测阈值。

请编写程序，利用非全天平均的天琴灵敏度曲线（见公式4.98），计算不同质量双黑洞并合事件的视界距离。具体画图中，横坐标取总质量  $M$ ，纵坐标取视界距离  $D$ ，注意请以 log-log 形式画图。计算过程中，假设质量比为 1 : 1，假设对所有的并合事件观测时长均为 3 个月（即，天琴名义连续工作时长，且双黑洞在天琴结束观测的瞬间刚好并合）。

### Exercise 6.4

已知在噪声前提下， $p_0(2\mathcal{F})$  服从自由度为 4 的中心  $\chi^2$  分布。而对于信噪比为  $\rho$  的信号而言， $p_1(\rho, 2\mathcal{F})$  服从自由度为 4，非中心参数为  $\rho^2$  的非中心  $\chi^2$  分布。已知单次测量误警率  $p_{\text{FA}}$  和多次测量误警率之间  $p_{\text{FA}}^N$  的关系为

$$p_{\text{FA}}^N = 1 - (1 - p_{\text{FA}})^N$$

考虑模板数为  $N = 10^{10}$ ，求问在误警率  $p_{\text{FA}}^N = 0.01$  时，多高信噪比的信号，其漏警率不高于 10%。

### Exercise 6.5

考虑恒星级双黑洞旋近信号的参数空间为：质量  $m_1, m_2 \in [5, 100]M_\odot$  范围内， $t_c \in [0, 5]\text{yr}$  范围内，采用 PN 波形，结合[数值差分方法](#)，利用[Monte Carlo 积分](#)对公式[6.106](#)积分，计算：天琴对恒星级双黑洞旋近进行探测所需要的波形模板数。

2023-03-13

# 7

## 引力波信号测量

### 7.1 参数估计

在第六章中，我们将重点放在对概率  $p(H_1|D, I)$  的分析上。其中， $H_1$  是替代模型， $D$  是数据，而  $I$  是先验信息。

通常，我们可以用一组参数  $\theta$  来描述具体的模型。在实现对信号的探测之后，我们还关心对响应参数的测量。具体来说，我们关心给定数据  $D$ ，对概率密度分布  $p(\theta|D, I)$  的测量。在本章中，我们结合引力波数据、波形和噪声模型，尝试对引力波源的参数实现反推。

#### 7.1.1 测量精度

首先，如果要对参数  $\theta$  实现点估计，可以采用的做法有很多，如最高概率点  $\theta_{\max}$ ，对应概率分布上的峰（peak, mode），

$$\frac{\partial}{\partial \theta_i} p(\theta|D, I) \Big|_{\theta=\theta_{\max}} = 0 \quad (7.1)$$

再比如平均点  $\bar{\theta}$ ，等等。根据不同的情形，这些点估计可能会相距较近，但也可能相距甚远。考虑到后验概率  $p(\theta|D, I)$  是先验概率  $p(\theta|I)$  和似然函数  $p(D|\theta, I)$  的相乘，若假设先验概率在较广的参数空间上都平坦，利用似然函数比的定义（公式6.42），可以得到，公式7.1等价于

$$\left( D - h(\theta) \Big| \frac{\partial}{\partial \theta_i} h(\theta) \right) \Big|_{\theta=\theta_{\max}} = 0 \quad (7.2)$$

再考虑数据  $D(t)$  本身，包含噪声  $n(t)$  和信号  $h(t|\theta_0)$ ，其中，真实的源参数为  $\theta_0$ ，即  $D(t) = n(t) + h(t|\theta_0)$ 。那么，公式 7.2 进一步等价于

$$\left( h(\theta_0) - h(\theta_{\max}) \middle| \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \right) = - \left( n \middle| \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \right) \quad (7.3)$$

在 Gaussian，稳态的假设下，等式右边可以看作是一组期望为零的高维 Gaussian 随机变量

$$v^i \equiv \left( n \middle| \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \right) \quad (7.4)$$

要理解对参数的测量误差，我们需要知道  $v$  的分布。

由于高维 Gaussian 随机变量  $v$  的期望为零，即  $\langle v^i \rangle = 0$ ，因此  $v$  的分布完全取决于 Fisher 信息矩阵：

$$\begin{aligned} \Gamma^{ij} &\equiv \langle v^i v^j \rangle = \left\langle \left( n \middle| \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \right) \left( n \middle| \frac{\partial}{\partial \theta_j} h(\theta_{\max}) \right) \right\rangle \\ &= \left\langle \left( \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \middle| \frac{\partial}{\partial \theta_j} h(\theta_{\max}) \right) \right\rangle \end{aligned} \quad (7.5)$$

这其中，我们运用了关系

$$\langle (n|g)(n|h) \rangle = \langle (g|h) \rangle \quad (7.6)$$

由此，我们得到， $k$  维随机变量  $v$  的概率密度分布为

$$p(v) = \frac{1}{\sqrt{(2\pi)^k \det(\Gamma^{mn})}} \exp \left( -\frac{1}{2} (\Gamma^{-1})_{ij} v^i v^j \right) \quad (7.7)$$

其中， $(\Gamma^{-1})_{ij}$  是 Fisher 信息矩阵的逆矩阵。

不妨考虑信号很强的情形，即，信噪比  $\rho$  较大。这时，最大似然函数法所得出的点估计  $\theta_{\max}$  与真实的源参数  $\theta_0$  足够接近，误差  $\Delta\theta = \theta_{\max} - \theta_0$  很小。可以将真实波形写作 Taylor 展开

$$h(\theta_0) = h(\theta_{\max}) - \Delta\theta_i \frac{\partial h(\theta_{\max})}{\partial \theta_i} + O(\Delta\theta^2) \quad (7.8)$$

将其代入到公式 7.3 后，可以得到，

$$\left( \frac{\partial}{\partial \theta_j} h(\theta_{\max}) \middle| \frac{\partial}{\partial \theta_i} h(\theta_{\max}) \right) \Delta\theta_j + O(\Delta\theta^2) = v^i \quad (7.9)$$

换言之，有

$$\Gamma^{ij} \Delta\theta_j \approx v^i \quad (7.10)$$

对应地，测量误差  $\Delta\theta$  的概率分布为

$$p(\Delta\theta) \approx \sqrt{\frac{\det(\Gamma^{mn})}{(2\pi)^k}} \exp \left( -\frac{1}{2} \Gamma^{ij} \Delta\theta_i \Delta\theta_j \right) \quad (7.11)$$

不难看出，在高信噪比极限下，Fisher 信息矩阵包含了参数的方差与协方差的信息。

$$(\Delta\theta_i)_{\text{rms}} = \left\langle (\Delta\theta_i)^2 \right\rangle^{1/2} = \left( \text{Var } \Delta\theta_i \right)^{1/2} = \sqrt{(\Gamma)^{-1}_{ii}}. \quad (7.12)$$

换言之，源参数间的方差-协方差矩阵  $\Sigma_{ij}$  就是 Fisher 信息矩阵的逆矩阵。

$$\Gamma^{ij}\Sigma_{jk} = \delta_k^i \quad (7.13)$$

可以利用 Fisher 信息矩阵对误差进行量级估算。考虑仅振幅  $A$  未知，其相对不确定度是  $\frac{\sigma_A}{A}$ ，其中，

$$\sigma_A \approx \sqrt{\Gamma^{AA}} \quad (7.14)$$

而对应的有

$$\Gamma^{AA} = \left( \frac{\partial h}{\partial A} \middle| \frac{\partial h}{\partial A} \right) = (g|g) = 1 \quad (7.15)$$

$A$  的误差为  $\sigma_A = \sqrt{(\Gamma^{AA})^{-1}} = 1$  所以，可以得到，振幅的相对不确定度  $\frac{\sigma_A}{A} = \sigma_{\ln A}$

$$\sigma_{\ln A} \approx 1/A = \frac{1}{\rho} \quad (7.16)$$

可以发现，对于信噪比为  $\rho$  的信号而言，其测量精度反比于信噪比。对于信噪比为 10 的信号，其振幅测量精度约为 10%；对于信噪比为 100 的信号，其振幅测量精度约为 1%。

## 7.1.2 参数估计中的系统误差

波形中可能会存在系统误差，比如，双中子星的引力波信号携带了潮汐信息，若采用双黑洞信号来匹配，就会对参数的测量带来系统偏差。考虑真实波形  $h(t|\boldsymbol{\theta}_0)$ ，以及用于匹配滤波的波形  $h'(t|\boldsymbol{\theta}_0)$ ，则有残差部分为  $\delta h(t|\boldsymbol{\theta}_0) = h'(t|\boldsymbol{\theta}_0) - h(t|\boldsymbol{\theta}_0)$ 。这时，公式7.3化为

$$\left( h(\boldsymbol{\theta}_0) - h'(\boldsymbol{\theta}_{\max}) \middle| \frac{\partial}{\partial \boldsymbol{\theta}_i} h'(\boldsymbol{\theta}_{\max}) \right) = - \left( n \middle| \frac{\partial}{\partial \boldsymbol{\theta}_i} h'(\boldsymbol{\theta}_{\max}) \right) \quad (7.17)$$

类似地，我们可以得到

$$\left( \delta h(\boldsymbol{\theta}_0) \middle| \frac{\partial}{\partial \boldsymbol{\theta}_i} h'(\boldsymbol{\theta}_{\max}) \right) + \Gamma^{ij} \Delta\theta_j \approx v^i \quad (7.18)$$

与公式7.10比较，可以发现多出一项，对应地，即为总体的系统误差

$$\begin{aligned} (\Delta\theta_i)_{\text{sys}} &= (\Gamma^{-1})_{ij} \left( \delta h(\boldsymbol{\theta}_0) \middle| \frac{\partial}{\partial \boldsymbol{\theta}_j} h'(\boldsymbol{\theta}_{\max}) \right) \\ &\approx (\Gamma^{-1})_{ij} \left( \delta h(\boldsymbol{\theta}_0) \middle| \frac{\partial}{\partial \boldsymbol{\theta}_j} h'(\boldsymbol{\theta}_0) \right) \end{aligned} \quad (7.19)$$

注意到，系统偏差的大小与信号的强度并无关系。

为更好地理解，不妨假设以下情形：系统偏差由单一参数  $\theta$  描述，

$$h(t|\theta) = h(t|\theta=0) + \theta\delta h \quad (7.20)$$

其中， $\delta h = h'(t|\theta=0) - h(t|\theta=0)$ 。不难得出， $\Gamma = (\delta h|\delta h)$ ，以及  $(\Delta\theta)_{\text{sys}} = 1$ 。这就意味着，如果要求误差中，统计误差占主导，就有  $(\Delta\theta)_{\text{rms}} > (\Delta\theta)_{\text{sys}}$ ，而  $(\Delta\theta)_{\text{rms}} = 1/\Gamma$ ，则等价于要求

$$(\delta h|\delta h) < 1 \quad (7.21)$$

如果上述条件满足，我们说两个波形无法区分 (indistinguishable)。公式 7.21 也被称为分辨判据 (indistinguishability criterion)。

值得注意的是，分辨判据是强度依赖的。在低信噪比下不可分辨的两个波形，随着信噪比的提升，可能被分辨出。另一个与强度无关的测量系统误差大小的方法是定义

$$\epsilon_{\text{sys}} \equiv \frac{(\delta h|\delta h)}{(h|h)} \quad (7.22)$$

当  $\epsilon_{\text{sys}} \geq \frac{1}{\rho^2}$  时，系统误差开始显著。换言之，对于信噪比为 10 的信号，我们要求波形的准确度  $\epsilon_{\text{sys}} < 1\%$ ，否则系统误差会影响对物理参数的估计影响。

另一种推导方式是由公式 7.19 出发，应用 Cauchy-Schwarz 不等式，就有

$$(\Delta\theta_i)_{\text{sys}} \leq (\Delta\theta_i)_{\text{rms}}(\delta h|\delta h)^{1/2} \quad (7.23)$$

我们再次发现，这一限制是与强度无关的。

### 7.1.3 置信区域

通常情况下，我们除了参数的点估计，也希望能知道后验概率分布。特别的，我们希望能给出以一定概率包含真实参数的参数空间范围。这一范围被称为置信区域 (confidence region)，对于一维情形，也叫做置信区间 (confidence interval)。需要注意的是，在 Bayesian 的语境下，通常把 Bayesian confidence interval 称为 credible interval 可信区间，以区分 frequentist 的定义。在本讲义中不作特别区分。

首先，定义置信水平  $\alpha$ ，通过求解，

$$\alpha = \int_{\theta_1}^{\theta_2} p(h(\theta)|D)d\theta \quad (7.24)$$

可以求得  $\theta_1$  和  $\theta_2$ ，这样  $\theta_1 < \theta < \theta_2$  的概率为  $\alpha$ 。这时可以说， $(\theta_1, \theta_2)$  是在置信水平  $\alpha$  下， $\theta$  的置信区间。

注意，给定  $\alpha$ ，可能可以对应许多不同的置信区间。通常可以选取测度最小的区域为置信区间。常见的选择还包括：以峰值左右对称选取置信区间；以期望值左右对称选取置信区间； $\theta_1 = -\infty; \theta_2 = \infty$  等。

在频率学派的语境下，由于只有似然函数有合法定义，无法定义后验概率，对其积分更无从谈起。严格的定义应该是，对某一真实参数  $\theta$ （一定存在但不一定明确可知），重复足够多次的实验，那么其中比例为  $\alpha$  的实验中，真实参数  $\theta$  将包含于置信区间中。

### 7.1.4 冗余参数

很多时候，我们关心的模型中不只包含一个参数，但可能我们只关心其中某一部分参数。比如，对于双星并合的引力波信号，我们关心其质量，但是对于并合时的相位并不关心；对于连续引力波信号，我们通常关心其频率和振幅，但是对于其初始相位并不关心。

这些通常不感兴趣，但又不得不包含的参数，称为冗余参数（nuisance parameter）。频率学派的主要工具是似然函数，而似然函数  $p(D|\theta, I)$  是给定参数的条件概率。因此，若想获得与冗余参数具体选取无关的结论，在频率学派的框架下处理冗余参数是较为复杂的。而 Bayesian 则可以很好地处理这一问题。考虑将参数分为两类  $\theta = \{\phi, \psi\}$ ，其中  $\psi$  是冗余参数。则可以定义边缘化（marginalization）

$$p(\phi|D, I) = \int d\psi p(\phi, \psi|D, I) \quad (7.25)$$

## 7.2 Markov 链蒙特卡洛

频率学派（frequentist）的方法曾经长期主导了概率领域，这是有其原因的：为了开展 Bayesian 推断，需要计算参数空间上的后验分布。如果无法通过解析得到结果，则相关计算必须通过数值方法得到，通常需要大量计算资源。另一方面，频率学派似乎提供了一种自然的工具来实现模型选择，即 *p-value*（p-值）。但是，需要注意的是，p 值的实际含义并不直观，并且很大一部分用户在研究中滥用了它。这并不是说频率学派方法是错误的，而只是意味着对结果的正确阐述需要付出不小的努力。

执行 Bayesian 推理的一种简单方法是网格方法，即将网格点均匀地放置在参数空间上。不妨假设在每个维度中分隔 10 个点，那就意味着，对于  $N$  个维度，总计有  $10^N$  点。在这种情况下，总体计算需求将随着讨论问题的参数维数指数增长，这很快会使格点搜索的策略难以承受。有一些巧妙的采样方法可以避开这种“维度诅咒”（curse of dimensionality），其中，最著名的两种方法是 Markov 链蒙特卡洛（Markov Chain Monte Carlo, MCMC）和嵌套采样（nested sampling）。

Markov 链的是这样一条链：下一个点的位置仅取决于前一个点的位置；而蒙特卡洛方法是采用随机数进行计算的一种通用方法。

最常见、也是最简单的实现 MCMC 的算法是 Metropolis-Hastings 算法。在这个算法中，首先在参数空间中选择一个随机位置。对于第  $i+1$  步， $\theta_{i+1}$  的位置由上一个点，即第  $i$  步的位置  $\theta_i$ ，结合转移概率 (*transition probability*) 或转移核 (*transition kernel*)  $P(\theta_{i+1}|\theta_i)$  而确定。通常，通过构建 *proposal density*  $q(\theta^*|\theta_i)$  和相应的 *acceptance probability*，可以从提议的概率密度函数  $q(\theta^*|\theta_i)$  中随机抽取尝试点，这一提议的参数点将按照接受概率接受，否则被拒绝，在这种情况下， $\theta_{i+1} = \theta_i$ 。我们后面将证明，以此构建的采样点密度将正比于后验概率分布 (*probability density function, PDF*)。

具体来说，首先需要计算出尝试点的后验概率，将其与前一个点的后验概率相比，可以得到 *Metropolis ratio*  $r$ ，

$$r = \frac{P(\theta^*|D, I) q(\theta^*|\theta_i)}{P(\theta_i|D, I) q(\theta_i|\theta^*)}. \quad (7.26)$$

新的点被接受的概率为接受概率  $\alpha$ ，具体定义为

$$\alpha(\theta_i, \theta^*) \equiv \min(1, r) = \min\left(1, \frac{P(\theta^*|D, I) q(\theta^*|\theta_i)}{P(\theta_i|D, I) q(\theta_i|\theta^*)}\right) \quad (7.27)$$

通过上述构建，MCMC 采样算法可以从先验概率允许的参数空间中的任何位置开始。通常来说，起始点很可能位于后验概率相对较低的区域，经过不断更新 Markov 链的采点，新的采样点不断在后验概率分布上“爬坡”，偶尔也可能向下“下坡”。经过“煲机”*burn-in* 阶段后，整个 MCMC 关于起点的“记忆”将被清除。因此，为了排除起点选择对最终结果的影响，通常会将 *burn-in* 阶段的点抛弃。

MCMC 算法会向目标分布  $P(\theta|D, I)$  采样，这也被称为这个 Markov 链的稳态分布 (*stationary distribution*)。MCMC 的采样工作需要持续到整个链条收敛。为了实现收敛，采样器必须满足三个属性：

### 1. 不可约性 Irreducible

对于后验概率不为零的点，采样器在该处采样的概率不应该为零，特别是采样器不应该卡在一个特定区域中。

### 2. 非周期性 Aperiodic

采样器不应在有限状态之间振荡。

### 3. 正复现性 Positive recurrent

存在一个稳态分布，否则讨论收敛问题是没有意义的。

由于对于常见的后验概率分布而言，Metropolis-Hastings 采样器满足上述所有要求，因此可以保证收敛到这一稳态分布。

细致平衡条件是 MCMC 收敛的充分非必要条件，通常可以用来判别采样器是否可以得到预期结果。细致平衡条件为可以总结为

$$P(\theta_i)q(\theta_{i-1}|\theta_i) = P(\theta_{i-1})q(\theta_i|\theta_{i-1}) \quad (7.28)$$

细致平衡条件的概念来源于热力学。在热力学中，如果我们将粒子处于状态  $\theta_i$  的概率定义为  $P(\theta_i)$ ，将其从状态  $\theta_i$  转移到  $\theta_{i+1}$  的概率定义为  $q(\theta_{i+1}|\theta_i)$ ，细致平衡条件要求，在任一时期内，从状态  $\theta_i$  跳到状态  $\theta_{i+1}$  的粒子数目应与从状态  $\theta_{i+1}$  跳到状态  $\theta_i$  的数目完全相同。

由于细致平衡要比收敛性强，因此某些不满足详细平衡要求的马尔可夫链仍然可以收敛到目标分布。但是，如果选择了详细的平衡标准，则检查和实施要容易得多。我们可以证明，详细的平衡可以导致目标分布的收敛，如下所示：

$$\begin{aligned} \text{Joint probability}(\theta_i, \theta_{i+1}) &= P(\theta_i)P(\theta_{i+1}|\theta_i) \\ &= P(\theta_i)q(\theta_{i+1}|\theta_i)\alpha(\theta_i, \theta_{i+1}) \\ &= P(\theta_i)q(\theta_{i+1}|\theta_i) \min\left(1, \frac{P(\theta_{i+1})q(\theta_i|\theta_{i+1})}{P(\theta_i)q(\theta_{i+1}|\theta_i)}\right) \\ &= \min(P(\theta_i)q(\theta_{i+1}|\theta_i), P(\theta_{i+1})q(\theta_i|\theta_{i+1})) \\ &= \alpha(\theta_{i+1}, \theta_i)P(\theta_{i+1})q(\theta_i|\theta_{i+1}) \\ &= P(\theta_{i+1})P(\theta_i|\theta_{i+1}) \end{aligned} \quad (7.29)$$

如果我们对所有可能的  $\theta_i$  进行边缘化，则

$$\begin{aligned} \int d\theta_i P(\theta_{i+1}|\theta_i)P(\theta_i) &= \int d\theta_i P(\theta_i|\theta_{i+1})P(\theta_{i+1}) \\ &= P(\theta_{i+1}) \int d\theta_i P(\theta_i|\theta_{i+1}) \\ &= P(\theta_{i+1}) \end{aligned} \quad (7.30)$$

在第二行中，积分自然地归一化。

公式7.30保证了采样得到任何参数  $\theta$  的概率（密度）恰好是该参数的后验概率（密度）。这一特性非常重要，因为它使地对概率密度的任何数值积分都变得非常容易。对于任何函数  $f(\theta)$ ，参数空间上的积分可以通过数值求和来近似

$$\int d\theta f(\theta)P(\theta) \approx \frac{1}{N} \sum_{i=1}^N f(\theta_i) \quad (7.31)$$

### 7.2.1 并行回火 MCMC (Parallel Tempering MCMC)

MCMC 的关键问题之一是多峰问题，即在后验分布中，参数空间内里存在多个不连通的区域，均可以很好地被数据支持，形成多个“高峰”。对于简单的 Metropolis-Hastings 算法，如果采样器想要从一种峰跳转到另一种峰，就必须经过后验概率的一个“谷”，这通常意味着经历一系列向低概率处行进的向下跳跃。这一过程需要花很长时间才会出现，因此尽管从理论上讲是可能的，但实践中并不实际。

并行回火 MCMC 的原理是对多条链同时采样，其中一些链采样的后验概率更平坦，而另一些则对目标分布进行采样。通过不同链之间的通信，它可以引导采样器实现参数空间上不同区域之间的跳转。

一个更平坦的分布具有如下后验

$$P_T(\theta|D, I) \propto P(\theta|I)P(D|\theta, I)^\beta$$

其中参数  $T \equiv \frac{1}{\beta}$  被定义为链的“温度”。在并行回火方案中，通常会构造一个温度阶梯，一系列（设为  $n$ ）条链对应不同的温度，即  $T_0 < T_1 < \dots < T_n$ ，其中最低温度  $T_0 = 1$ 。不难发现，较高温度的链与先验更相似，温度为  $\infty$  的链将完全退化为先验分布。

需要注意的是，相邻链之间，温度比率恒定。对于温度较高的链，由于其后验概率比较平坦，采样器更有可能实现不同高峰之间的跳跃。时不时地，两个相邻链之间可以发生交换，这种交换的接受度以概率

$$r = \min \left\{ 1, \frac{P_{T_j}(\theta_{i+1}|D, I)P_{T_{j+1}}(\theta_i|D, I)}{P_{T_j}(\theta_i|D, I)P_{T_{j+1}}(\theta_{i+1}|D, I)} \right\}. \quad (7.32)$$

给出。通过这种方案，可以让低温链根据目标分布进行采样，而高温链则辅助引导在不同高峰之间探索，并通过与低温链条的交换实现采样区域的切换。

### 7.2.2 延迟拒绝 MCMC (Delayed Rejection MCMC)

在大多数情况下，MCMC 算法的计算瓶颈在于对后验概率的数值计算。因此，根据接受率而最终拒绝掉点  $\theta_{(1)}^*$  时，在某种程度上是对计算能力的浪费。我们将方程 7.27 稍微修改为

$$\alpha_1(\theta_i, \theta_{(1)}^*) = \min \left( 1, \frac{P(\theta_{(1)}^*)}{P(\theta_i)} \frac{q_1(\theta_{(1)}^* | \theta_i)}{q_1(\theta_i | \theta_{(1)}^*)} \right).$$

尽管 MCMC 过程总是需要通过拒绝，才能为我们提供有关后验的信息，但是不立即丢弃计算的信息，而是通过糅合其信息的新的候选分布  $q_2(\theta_{(2)}^* | \theta_{(1)}^*, \theta_i)$  得到新的点。

对应的，新提议的点  $\theta_{(2)}^*$  的接受率为

$$\alpha_1(\theta_i, \theta_{(1)}^*, \theta_{(2)}^*) = \min \left( 1, \frac{P(\theta_{(2)}^*)}{P(\theta_i)} \frac{q_1(\theta_{(2)}^* | \theta_{(1)}^*)}{q_1(\theta_i | \theta_{(1)}^*)} \frac{q_2(\theta_{(2)}^* | \theta_{(1)}^*, \theta_i)}{q_2(\theta_i | \theta_{(1)}^*, \theta_{(2)}^*)} \frac{1 - \alpha_1(\theta_{(2)}^*, \theta_{(1)}^*)}{1 - \alpha_1(\theta_i, \theta_{(1)}^*)} \right) \quad (7.33)$$

最后一项是将提议的  $\theta_{(1)}^*$  不被接受的情形纳入考虑而引入的修正。这一思想可以进一步衍生到更高的延迟拒绝数。

### 7.2.3 可逆跳 MCMC(Reversible Jump MCMC)

由于普通的 MCMC 通常难以对贝叶斯证据给出准确估算，通常不会将 MCMC 应用于模型选择问题。然而，可逆跳 MCMC 可以产生在两个不同模型之间的跳转建议。假设有两个对应不同维度的模型，在这一算法下，除了通常的 Metropolis-Hastings 采样外，会设计一个跨维度的跳跃提议，可以保证在模型之间的运动满足细致平衡。

可逆跳 MCMC 的核心问题是如何构建不同模型之间的跳跃提议。一旦跨维度的跳跃建议设计不佳，会导致不同模型之间的交换效率低下。

### 7.2.4 自动化 MCMC(Automated MCMC)

对于一条 Markov 链，由于新的点的位置取决于上一个点，因此每一个点都会与相邻点有一定相关性，且在链上距离越近，它们之间的相关性就越高。可以用自相关函数来表征一个点和另一个间隔  $\tau$  个数的点之间的相关性

$$R(\tau) = \frac{\text{E}[(\theta(i) - \mu_\theta)(\theta(i + \tau) - \mu_\theta)]}{\sigma_\theta^2}, \quad (7.34)$$

其中  $\mu_\theta$  和  $\sigma_\theta$  是分布的均值和标准差，并且平均值接管指数  $i$ 。我们通常将自相关函数的特征长度记作自相关长度，在一个自相关长度内，只保留一个采样点，从而保证不同点之间没有相关性。

在优化 MCMC 算法时，确保跳转提议的设计合理并非易事。如果建议的跳动太大，则接受率可能太低而无法有效；另一方面，如果建议跳跃太小，则采样点将具有非常高的自相关长度，因此 MCMC 需要花费很长时间，才能探索足够的参数空间。最佳建议将位于中间的某个位置，这样可以实现足够高的接受概率，而自相关长度也足够短，以使采样过程更加有效。

对于这一类问题，不少研究对 MCMC 过程的自动化作出了尝试，例如，Gregory 提出了一种方法 [48]，该方法在运行 MCMC 算法之前设置了目标接受率。有时，将实际的容忍率与参考进行比较，当现有的平均接收率高于预设值时，表示步长设置过低，相

关长度太低，则降低步长；反过来，如果接受率低于预设值，则表明步长太大，要降低步长。另一方面，在高维参数空间上，后验概率分布往往具有一定的内在相关性，通常可以计算所有已有点的协方差矩阵，以为跳转提议作为参考。

### 7.2.5 仿射不变的 MCMC (Affine Invariant MCMC)

近年来，一种特定的 MCMC 实现，即仿射不变的 MCMC，引起了极大的关注（又称系综采样器 ensemble sampler）[29]。它从差分进化算法中借用了提出新提议点的想法。相较单纯地从前一个点提出新的提一点，这一算法从样本池中随机抽取另外两个上一个点  $\theta_1$  和  $\theta_2$ ，并将这些点之间的差与系数  $\gamma$  相乘。在前一点上添加此差异向量以构建建议

$$\theta^* = \theta + \gamma(\theta_2 - \theta_1). \quad (7.35)$$

我们可以注意到，严格来说，该算法不是 Markov 算法，因为该点将取决于之前的三个不同点。这种非 Markov 提议的优点是，它可以以相对较高的接受率实现较低的自相关长度 [29]。因此，从原理上说，该方法可以具备极高的效率。此外，需要注意，通常的 MCMC 实现需要调整多个运行参数，才能使其按预期运行。在这种仿射不变式 MCMC 中，由于该建议的构建可以反映现有点的几何特性，调参问题会被大大简化，因此这种方法很容易上手。它的 python 实现 emcee 已成为最近极受欢迎的 MCMC 工具之一 [43]。

### 7.2.6 组合

本节中提到的方法可以组合在一起以构造新方法。例如，可以将延迟拒绝和可逆跳跃结合起来，甚至将 MCMC 与后文提到的嵌套抽样算法结合起来，或者与重要性采样结合起来。重要性采样是一种从比目标分布更简单的分布形式中提出独立样本的策略，因此可以高效地进行数值积分。

## 7.3 模型选择与 Occam 刃刀

我们经常会面临这样的问题：对于相同的物理现象，或者实验观测数据  $D$ ，可以有不同的多个模型  $M_i$  和  $M_j$  来解释。举例而言，面临未知的现象，如 GRB, FRB 等，可以由不同的物理机制来解释；SNe Ia 也有单白矮星吸积和双白矮星并合两种不同的模型。这种情况就需要利用数据，对竞争的模型进行判别。另外的情况是，在研究中，会有一些广泛使用的、非常基础的假设，比如 Galileo 和 Eötvös 先后通过实验等效原理进行检验，可以通过数据理论进行检验。LIGO、Virgo 等地面引力波探测器探测到致密双星并合信号后，也利用观测数据对广义相对论进行了检验。

在 Bayes 理论的框架下，需要通过模型选择 (model selection)，定量地分析数据对模型的支持程度，从而实现对不同的理论进行区分的目的。

### 7.3.1 Occam 剃刀

通常情况下，会把 Bayesian 模型选择和 Occam's razor 连结起来描述。Occam 剃刀是由方济各会修士 William of Ockham 提出来的，在英文中通常表述为 “Entities should not be multiplied without necessity”，通常翻译为“如无必要，勿增实体”。这一哲学的指导思想并不对应任何物理法则，但对推动自然科学的发展上至关重要。Einstein 也有过类似的表述：“Everything should be as simple as possible, but not simpler”。

通常，可以这样去理解 Occam 剃刀：如果两个模型  $M_i$  和  $M_j$  同时可以解释观测到的数据  $D$ ，如果可以同样好地解释数据，那么我们会更希望选择更简洁的模型。比如，在古典的地心说里，是利用大量的本轮、均轮来解释日月星辰的运行。通过不断增加本轮、均轮的数量，总可以不断提高对观测数据的拟合结果；而将太阳放到太阳系中心，将行星轨道换作椭圆轨道，精度上并不一定比古典地心说更高，但由于模型简洁程度更高，因此通常会偏向于选取后者作为解释理论。

 **注 7.1.** 实际上，在 Nicolaus Copernicus 第一次提出日心说时，由于并没有采用椭圆轨道，因此也必须依赖于本轮、均轮的辅助，从模型而言，并不比古典的地心说更简洁，也难怪日心说理论在初期难以受到支持。

但是 Occam 剃刀本身有自己的局限性。比如，通常情况下，更简洁的模型并不总能更好地解释数据。据称，von Neumann 曾说：“With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.” 当模型组建地足够复杂时，往往能比简洁模型更好地解释数据。然而这通常只是意味着模型对数据的过拟合，并不意味着模型的有效性。在面对更复杂而更有效的模型和更简洁却解释能力更差的模型时，Occam 剃刀是无力的。幸运的是，Bayesian 统计的框架下，自然地会对更复杂的模型引入“处罚”，这样，自动平衡复杂模型带来的对拟合度的提升，而只在必要的时候选择更为复杂的模型。

### 7.3.2 Odds ratio

我们考虑  $N$  个互斥的模型  $M_i$ ，其中有且只有一个理论可以正确地解释数据，换言之，有  $I = M_1 + M_2 + \dots + M_N$ ，其中的加号对应逻辑或。由此，我们可以利用 Bayes 定理，计算出

$$p(M_i|D, I) = \frac{p(M_i|I)p(D|M_i, I)}{P(D|I)} \quad (7.36)$$

注意，其中  $p(D|M_i, I)$  就是 Bayesian evidence，(Bayes evidence 也被称为 marginalised likelihood，因为它是对似然函数的 marginalisation)

$$p(D|M_i, I) = \int d\theta p(\theta|M_i, I)p(D|\theta, M_i, I) \quad (7.37)$$

同时，与之前类似，分母上的  $p(D|I)$  是与具体模型无关的归一化常数。

具体对两个模型  $M_i$  和  $M_j$  进行比较时，我们计算出数据对两个模型的支持度之比

$$O_{ij} = \frac{p(M_i|D, I)}{p(M_j|D, I)} \quad (7.38)$$

这一比值通常被称为 **odds ratio**。更进一步地，将定义式代入后，可以得到

$$\begin{aligned} O_{ij} &= \frac{p(M_i|I)}{p(M_j|I)} \frac{p(D|M_i, I)}{p(D|M_j, I)} \\ &\equiv \frac{p(M_i|I)}{p(M_j|I)} B_{ij} \end{aligned} \quad (7.39)$$

这其中，第一项为 prior odds ratio，第二项为 Bayes factor。需要注意的是，由于 odds ratio 关注的是两个模型的 Bayesian evidence 的比值，因此，与具体模型无关的分母  $p(D|I)$  在计算中上下抵消，并不会出现在表达式中。

一般而言，可以假设在获取实验数据之前，不对任何模型有偏好，即 prior odds ratio 为 1。因此，odds ratio 的取值由 Bayes factor 决定。

### 7.3.3 定量的 Occam 剃刀：简单推导

我们以下列简单情形做参考，以定性推导来证明 odds ratio 对简洁模型的偏好。不妨考虑两个模型： $M_1$  有一个自由参数  $\theta$ ，而  $M_0$  则将参数  $\theta$  选取固定值  $\theta_0$ 。为了计算 Bayes factor  $B_{10}$ ，可以首先计算模型一下的 Bayesian evidence  $p(D|M_1, I)$ 。由于数据中包含信息，因此似然函数  $\mathcal{L}(\theta) = p(D|\theta, M_1, I)$  相较先验  $p(\theta|M_1, I)$  会更为集中。

如图 7-1 所示，考虑似然函数在  $\hat{\theta}$  达到峰值，并服从 Gaussian 分布。与此同时，假设先验 prior 的宽度为  $\Delta\theta$ 。对于均匀分布的先验，有

$$\int_{\Delta\theta} d\theta p(\theta|M_1, I) = p(\theta|M_1, I)\Delta\theta = 1 \quad (7.40)$$

换言之，有

$$p(\theta|M_1, I) = \frac{1}{\Delta\theta} \quad (7.41)$$

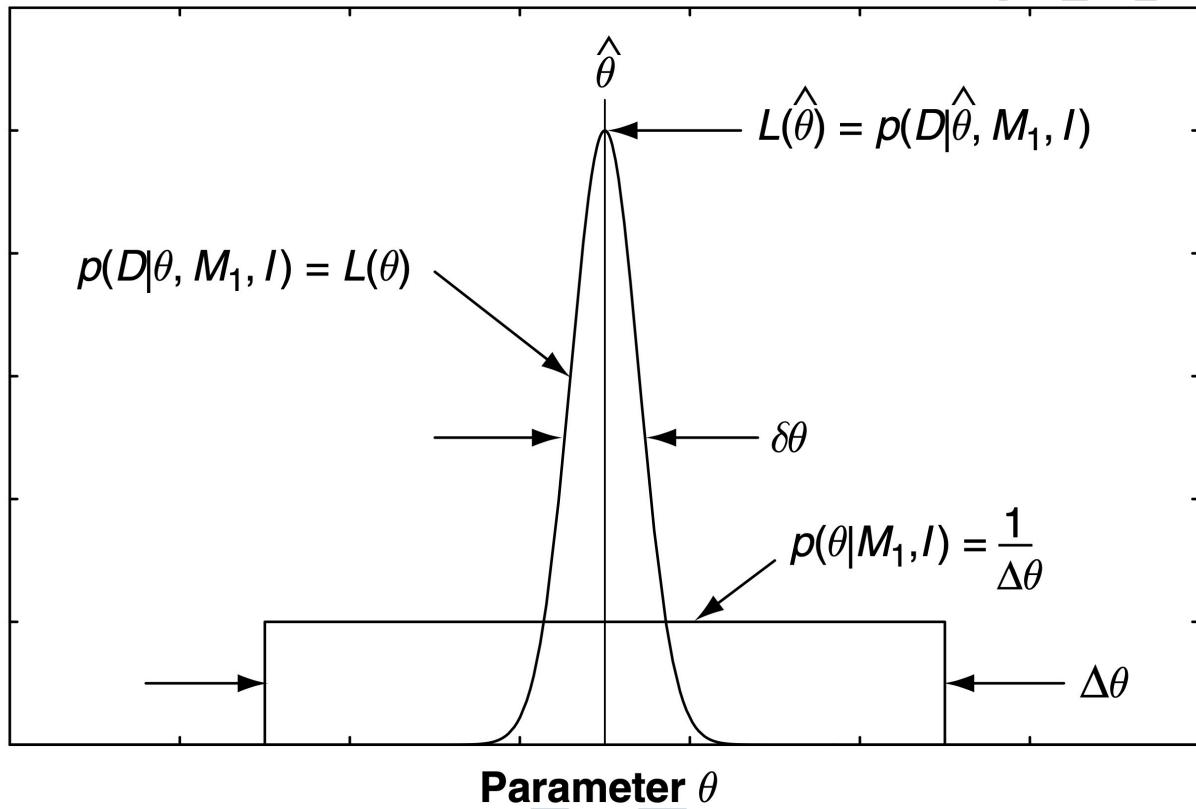
图 7-1 似然函数峰特征宽度  $\delta\theta$  与先验的宽度  $\Delta\theta$ .

Figure 7-1 The characteristic width  $\delta\theta$  of the likelihood peak and  $\Delta\theta$  of the prior. Credit: [48]

另一方面，我们可以定义似然函数的特征宽度  $\delta\theta$

$$\int_{\Delta\theta} d\theta p(D|\theta, M_1, I) = p(D|\hat{\theta}, M_1, I)\delta\theta \quad (7.42)$$

利用上述等式，可以将 Bayesian evidence(公式7.37) 重新表述

$$\begin{aligned} p(D|M_1, I) &= \int d\theta p(\theta|M_1, I)p(D|\theta, M_1, I) \\ &= \frac{1}{\Delta\theta} \int d\theta p(D|\theta, M_1, I) \\ &\approx p(D|\hat{\theta}, M_1, I) \frac{\delta\theta}{\Delta\theta} \end{aligned} \quad (7.43)$$

反观模型 0  $M_0$ ，由于不包含未定参数，因此其 Bayesian evidence 的计算不涉及积分，直接就是  $M_1$  在  $\theta = \theta_0$  时的似然函数值。

$$p(D|M_0, I) = p(D|\theta_0, M_1, I) \quad (7.44)$$

综合而言，可以计算两个模型之间的 Bayes factor

$$B_{10} \approx \frac{p(D|\hat{\theta}, M_1, I)}{p(D|\theta_0, M_1, I)} \frac{\delta\theta}{\Delta\theta} \quad (7.45)$$

在公式7.45中，等式右边的第一项一般总是会偏好更复杂的模型  $M_1$ ，因为  $M_0$  只不过是  $M_1$  中的一个特例，而允许增加一个新的自由参数意味着模型总能更好地拟合数据。但是，在第二项中由于后验的宽度  $\delta\theta$  比  $\Delta\theta$  更窄，因此，可以将第二项的比值视为对因无法良好匹配数据而“浪费”的参数空间的惩罚。因此，只有当似然函数的比值足够大时，才能补偿第二项的偏差，否则 Bayes factor 总是会偏好更简单的模型。

## 7.4 嵌套抽样

尽管 MCMC 是一种非常有用的方法，可以在各种问题（尤其是高维问题）中开展 Bayes 推断，但它也有其自身的缺点。例如，大多数 MCMC 采样点都将集中在后验概率较高的区域，无法有效探索参数空间的绝大部分。这一特性有助于使 MCMC 集中在有意义的区域，从而解决计算需求扩展速度远快于维数增长的问题，即所谓的“curse of dimensionality”问题。但是，这也意味着参数空间中，后验概率低的部分将在很大程度上被忽略，因此，通常的 MCMC 采样算法很难准确地估计 Bayes 证据（尽管一些研究已经努力缓解这一问题）。另一方面，嵌套抽样算法虽然以估算 Bayes 证据出发，但自身还具有执行参数估算的能力。[\[89\]](#)。

首先，让我们看一下统计力学中正则配分函数的计算。配分函数  $Z(\beta)$  定义为

$$Z(\beta) = \int d\theta \exp(-\beta E(\theta)). \quad (7.46)$$

一旦了解状态密度的信息

$$g(E) = \int d\theta \delta(E - E(\theta)), \quad (7.47)$$

其中  $\delta$  是 Dirac delta 函数，就可以将公式 7.46 转换为

$$Z(\beta) = \int dE g(E) \exp(-\beta E). \quad (7.48)$$

现在让我们回到 Bayes 的观点，我们将  $-\log P(D|\theta, M, I)$  视为  $E(\theta)$ ，并且

$$g(E) = \int d\theta P(\theta|M, I) \delta(E - E(\theta)).$$

在这种视角下，当  $\beta = 1$  时，Bayes 证据就是配分函数。

嵌套采样的原理是通过数值实现积分  $Z = \int_{-\infty}^{\infty} dE g(E) \exp(-E)$ 。

可以将先验质量 定义为

$$X(L) = \int_{P(D|\theta, M, I) > L} d\theta P(\theta|M, I) \quad (7.49)$$

也就是对参数空间中似然函数的值大于预设的  $L$  进行积分得到的结果。 $X$  的值限制在 0 和 1 之间 [89]。

Bayes 证据的积分可以转化为

$$\begin{aligned} Z &= \int_{-\infty}^{\infty} dE g(E) \exp(-E) \\ &= \int_0^1 dX L(X) \\ &\approx \sum_{i=2}^N L_i (X_i - X_{i-1}). \end{aligned} \quad (7.50)$$

嵌套采样算法将从先验中提取多个点  $\theta_i$ ，并计算相应的似然函数  $L_i$ ，随后将其排序，使得  $L_1 < L_2 < \dots < L_N$ 。自然，可以得出  $X_1 > X_2 > \dots > X_N$  的结论，如图7-2所示 [89]。

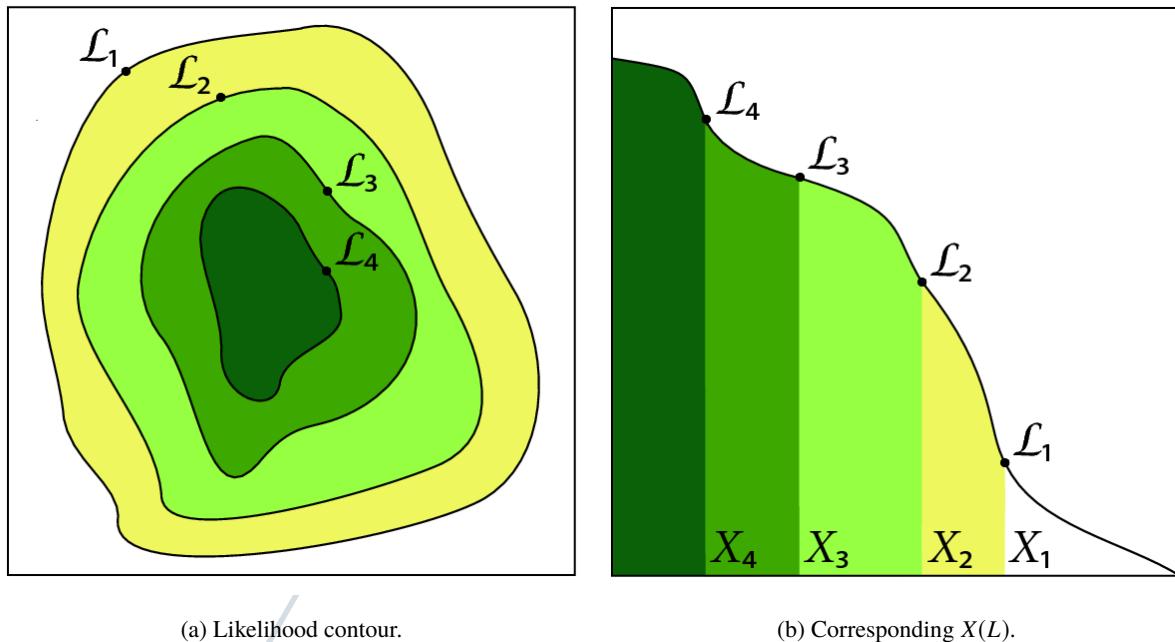


图 7-2 似然函数的等高线和对应  $X(L)$  的示意图。

Figure 7-2 An illustration of likelihood contours and corresponding  $X(L)$ . Credit: [89]

要开展嵌套采样算法，需要先撒开  $n$ （例如  $n \sim 1,000$ ）个活跃点（live point）。具有最小似然值的点将被记录为第一个采样点  $\theta_1$ 。该活跃点将被一个随机挑选的点进行更

新，更新后点的后验概率值需要大于更新前的点。这样，采样器将不断向上，到更高后验概率的区域，就像一层层剥洋葱一样。对于步骤  $k$  处的每次更新，将质量  $X_k$  定义为，根据所有活跃点中具有最低后验概率的点，所计算出的质量  $X(L)$ 。对于一次完全随机的更新，收缩率  $t = X_{k+1}/X_k$  遵循分布

$$P(t) = nt^{n-1}$$

因此有

$$\langle \log t \rangle = (-1 \pm 1)/n$$

由于  $X_0 = 1$ ，而  $X_k = \prod_{j=1}^k t_j$ ，则可以证明

$$\langle \log X_k \rangle = (-k \pm \sqrt{k})/n \quad (7.51)$$

结合方程7.50，可以显式计算出 Bayes 证据的估计值。还可以计算定义为

$$\mathcal{H} = \int dX P(X) \log [P(X)] \approx \sum_k \frac{L_k (X_k - X_{k-1})}{Z} \log \left[ \frac{L_k}{Z} \right]. \quad (7.52)$$

的信息  $\mathcal{H}$ 。通过信息  $\mathcal{H}$  中，可以估计当前对 Bayes 证据估计的标准差，因此

$$\log Z \approx \log \left( \sum_k L_k (X_k - X_{k-1}) \right) \pm \sqrt{\frac{\mathcal{H}}{n}}. \quad (7.53)$$

该算法被设计为在后验概率较低的区域中进行较为稀疏的采样，由于该部分对证据的贡献较小，可以容忍比较高的相对误差；而对于后验概率较高的区域，将会被密集采样，这样可以将最终估计中的不确定性降低 [89]。

#### 7.4.1 MultiNest

嵌套采样的实现也有不同的变体。由于原则上应该根据先验质量均匀地随机分配活跃点的更新，这意味着随着采样的进行，活跃点将变得越来越集中，直到它们缩小到参数空间中的一个小区域为止，此时提议得到一个合格点的概率会变低，计算效率随之降低 [89]。

MultiNest 算法利用活跃点采用高斯近似，用多个高维椭球来近似后验概率分布，并从椭球中拟议新的活跃点，可以提高采样效率。因此，该方法对于多峰问题尤其有效 [42]。

# 8

## 引力波数据处理实例



2023-03-13

2023-03-13

# A

## 常用物理量

物理含义	符号	数值	量纲	备注
光速	$c$	299,792,458	m/s	引力波速度亦为光速 $c$
引力常数	$G$	$6.67430(15) \times 10^{-11}$	$\text{m}^3 \cdot \text{kg}^{-1} \cdot \text{s}^{-2}$	来自 CODATA 数据 <sup>1</sup>
太阳质量	$m_{\odot}$	$1.98847 \times 10^{30}$	kg	
百万秒差距	Mpc	$3.08567758128 \times 10^{22}$	m	

<sup>1</sup>其中最新、误差最小的测量来自天琴中心主任罗俊院士所带领团队 [60]

2023-03-13

# 参考文献

- [1] B. P. Abbott et al. “A gravitational-wave standard siren measurement of the Hubble constant”. 刊于: *Nature* 551.7678 (Nov. 2017), pp. 85–88. doi: [10.1038/nature24471](https://doi.org/10.1038/nature24471). arXiv: [1710.05835 \[astro-ph.CO\]](https://arxiv.org/abs/1710.05835).
- [2] B. P. Abbott et al. “Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A”. 刊于: *the Astrophysical Journal Letters* 848.2, L13 (Oct. 2017), p. L13. doi: [10.3847/2041-8213/aa920c](https://doi.org/10.3847/2041-8213/aa920c). arXiv: [1710.05834 \[astro-ph.HE\]](https://arxiv.org/abs/1710.05834).
- [3] B. P. Abbott et al. “GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral”. 刊于: *Physical Review Letters* 119.16, 161101 (Oct. 2017), p. 161101. doi: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101). arXiv: [1710.05832 \[gr-qc\]](https://arxiv.org/abs/1710.05832).
- [4] B. P. Abbott et al. “Multi-messenger Observations of a Binary Neutron Star Merger”. 刊于: *the Astrophysical Journal Letters* 848.2, L12 (Oct. 2017), p. L12. doi: [10.3847/2041-8213/aa91c9](https://doi.org/10.3847/2041-8213/aa91c9). arXiv: [1710.05833 \[astro-ph.HE\]](https://arxiv.org/abs/1710.05833).
- [5] B. P. Abbott et al. “Properties of the Binary Neutron Star Merger GW170817”. 刊于: *Physical Review X* 9.1, 011001 (Jan. 2019), p. 011001. doi: [10.1103/PhysRevX.9.011001](https://doi.org/10.1103/PhysRevX.9.011001). arXiv: [1805.11579 \[gr-qc\]](https://arxiv.org/abs/1805.11579).
- [6] B. P. Abbott et al. “Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA”. 刊于: *Living Reviews in Relativity* 21.1, 3 (Apr. 2018), p. 3. doi: [10.1007/s41114-018-0012-9](https://doi.org/10.1007/s41114-018-0012-9). arXiv: [1304.0670 \[gr-qc\]](https://arxiv.org/abs/1304.0670).
- [7] B.P. Abbott et al. “GW190425: Observation of a Compact Binary Coalescence with Total Mass  $\sim 3.4M_{\odot}$ ”. 刊于: *Astrophys. J. Lett.* 892.1 (2020), p. L3. doi: [10.3847/2041-8213/ab75f5](https://doi.org/10.3847/2041-8213/ab75f5). arXiv: [2001.01761 \[astro-ph.HE\]](https://arxiv.org/abs/2001.01761).

- [8] R. Abbott et al. “Gravitational-wave Constraints on the Equatorial Ellipticity of Millisecond Pulsars”. 刊于: *Astrophys. J.* 902.1 (2020), p. L21. doi: [10.3847/2041-8213/abb655](https://doi.org/10.3847/2041-8213/abb655). arXiv: [2007.14251 \[astro-ph.HE\]](https://arxiv.org/abs/2007.14251).
- [9] R. Abbott et al. “GW190521: A Binary Black Hole Merger with a Total Mass of  $150 M_\odot$ ”. 刊于: *Phys. Rev. Lett.* 125.10 (2020), p. 101102. doi: [10.1103 / PhysRevLett.125.101102](https://doi.org/10.1103/PhysRevLett.125.101102). arXiv: [2009.01075 \[gr-qc\]](https://arxiv.org/abs/2009.01075).
- [10] R. Abbott et al. “GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object”. 刊于: *Astrophys. J. Lett.* 896.2 (2020), p. L44. doi: [10.3847/2041-8213/ab960f](https://doi.org/10.3847/2041-8213/ab960f). arXiv: [2006.12611 \[astro-ph.HE\]](https://arxiv.org/abs/2006.12611).
- [11] R. Abbott et al. “Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences”. 刊于: *Astrophys. J. Lett.* 915.1 (2021), p. L5. doi: [10.3847/2041-8213/ac082e](https://doi.org/10.3847/2041-8213/ac082e). arXiv: [2106.15163 \[astro-ph.HE\]](https://arxiv.org/abs/2106.15163).
- [12] R. Abbott et al. “The population of merging compact binaries inferred using gravitational waves through GWTC-3”. 刊于: (Nov. 2021). arXiv: [2111.03634 \[astro-ph.HE\]](https://arxiv.org/abs/2111.03634).
- [13] P. a R Ade, R. W. Aikin, et al. “Detection of B-mode polarization at degree angular scales by BICEP2”. 刊于: *Physical Review Letters* 112 (2014), pp. 1–26. ISSN: 10797114. doi: [10.1103/PhysRevLett.112.241101](https://doi.org/10.1103/PhysRevLett.112.241101). arXiv: [1403.3985](https://arxiv.org/abs/1403.3985).
- [14] European Space Agency. “Planck: The Scientific Programme”. 刊于: *arXiv preprint arXiv:astro-ph/0604069* (2006), pp. 1–152. arXiv: [0604069 \[astro-ph\]](https://arxiv.org/abs/0604069). URL: <http://www.informaworld.com/index/905664920.pdf>.
- [15] Kazunori Akiyama et al. “First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole”. 刊于: *Astrophys. J.* 875.1 (2019), p. L1. doi: [10.3847/2041-8213/ab0ec7](https://doi.org/10.3847/2041-8213/ab0ec7). arXiv: [1906.11238 \[astro-ph.GA\]](https://arxiv.org/abs/1906.11238).
- [16] Bruce Allen. “ $\chi^2$  time-frequency discriminator for gravitational wave detection”. 刊于: *Phys. Rev. D* 71 (2005), p. 062001. doi: [10.1103/PhysRevD.71.062001](https://doi.org/10.1103/PhysRevD.71.062001). arXiv: [gr-qc/0405045 \[gr-qc\]](https://arxiv.org/abs/gr-qc/0405045).
- [17] Bruce Allen. “Spherical ansatz for parameter-space metrics”. 刊于: *Physical Review D* 100.12, 124004 (Dec. 2019), p. 124004. doi: [10.1103/PhysRevD.100.124004](https://doi.org/10.1103/PhysRevD.100.124004). arXiv: [1906.01352 \[gr-qc\]](https://arxiv.org/abs/1906.01352).

- [18] Pau Amaro-Seoane and Miguel Pretor. “The impact of realistic models of mass segregation on the event rate of extreme-mass ratio inspirals and cusp re-growth”. 刊于: *Classical and Quantum Gravity* 28.9 (Apr. 2011), p. 094017. issn: 1361-6382. doi: [10.1088/0264-9381/28/9/094017](https://doi.org/10.1088/0264-9381/28/9/094017). url: <http://dx.doi.org/10.1088/0264-9381/28/9/094017>.
- [19] Andrea Antonelli et al. “Quasi-circular inspirals and plunges from non-spinning effective-one-body Hamiltonians with gravitational self-force information”. 刊于: *arXiv e-prints*, arXiv:1907.11597 (July 2019), arXiv:1907.11597. arXiv: [1907 . 11597 \[gr-qc\]](https://arxiv.org/abs/1907.11597).
- [20] R. F. Archibald et al. “A High Braking Index for a Pulsar”. 刊于: *the Astrophysical Journal Letters* 819.1, L16 (Mar. 2016), p. L16. doi: [10.3847/2041-8205/819/1/L16](https://doi.org/10.3847/2041-8205/819/1/L16). arXiv: [1603 . 00305 \[astro-ph.HE\]](https://arxiv.org/abs/1603.00305).
- [21] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, The NANOGrav Collaboration, et al. “The NANOGrav 11-year Data Set: High-precision Timing of 45 Millisecond Pulsars”. 刊于: *The Astrophysical Journal Supplement Series* 235, 37 (Apr. 2018), p. 37. doi: [10.3847/1538-4365/aab5b0](https://doi.org/10.3847/1538-4365/aab5b0). arXiv: [1801 . 01837 \[astro-ph.HE\]](https://arxiv.org/abs/1801.01837).
- [22] Zaven Arzoumanian et al. “The NANOGrav 12.5-year Data Set: Search For An Isotropic Stochastic Gravitational-Wave Background”. 刊于: (Sept. 2020). arXiv: [2009 . 04496 \[astro-ph.HE\]](https://arxiv.org/abs/2009.04496).
- [23] Stanislav Babak et al. ““Kludge” gravitational waveforms for a test-body orbiting a Kerr black hole”. 刊于: *Physical Review D* 75.2, 024005 (Jan. 2007), p. 024005. doi: [10.1103/PhysRevD.75.024005](https://doi.org/10.1103/PhysRevD.75.024005). arXiv: [gr-qc/0607007 \[gr-qc\]](https://arxiv.org/abs/gr-qc/0607007).
- [24] Stanislav Babak et al. “Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals”. 刊于: *Physical Review D* 95.10, 103012 (May 2017), p. 103012. doi: [10.1103/PhysRevD.95.103012](https://doi.org/10.1103/PhysRevD.95.103012). arXiv: [1703 . 09722 \[gr-qc\]](https://arxiv.org/abs/1703.09722).
- [25] Jiahui Bao et al. “Constraining modified gravity with ringdown signals: an explicit example”. 刊于: (2019). arXiv: [1905 . 11674 \[gr-qc\]](https://arxiv.org/abs/1905.11674).
- [26] Leor Barack and Curt Cutler. “LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy”. 刊于: *Physical Review D* 69.8, 082005 (Apr. 2004), p. 082005. doi: [10.1103/PhysRevD.69.082005](https://doi.org/10.1103/PhysRevD.69.082005). arXiv: [gr-qc/0310125 \[gr-qc\]](https://arxiv.org/abs/gr-qc/0310125).

- [27] Leor Barack and Norichika Sago. “Gravitational self-force on a particle in eccentric orbit around a Schwarzschild black hole”. 刊于: *Physical Review D* 81.8, 084021 (Apr. 2010), p. 084021. doi: [10.1103/PhysRevD.81.084021](https://doi.org/10.1103/PhysRevD.81.084021). arXiv: [1002.2386 \[gr-qc\]](https://arxiv.org/abs/1002.2386).
- [28] M. Bignotto et al. “A Cross-correlation method to search for gravitational wave bursts with AURIGA and Virgo”. 刊于: *Class. Quant. Grav.* 25 (2008), p. 114046. doi: [10.1088/0264-9381/25/11/114046](https://doi.org/10.1088/0264-9381/25/11/114046). arXiv: [0801.3801 \[gr-qc\]](https://arxiv.org/abs/0801.3801).
- [29] Cajo J. F. Ter Braak. “A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces”. 刊于: *Statistics and Computing* 16.3 (Sept. 2006), pp. 239–249. issn: 0960-3174. doi: [10.1007/s11222-006-8769-1](https://doi.org/10.1007/s11222-006-8769-1). URL: <http://www.springerlink.com/index/10.1007/s11222-006-8769-1>.
- [30] Richard Brito et al. “Stochastic and resolvable gravitational waves from ultralight bosons”. 刊于: *Phys. Rev. Lett.* 119.13 (2017), p. 131101. doi: [10.1103/PhysRevLett.119.131101](https://doi.org/10.1103/PhysRevLett.119.131101). arXiv: [1706.05097 \[gr-qc\]](https://arxiv.org/abs/1706.05097).
- [31] W. R. Brown et al. “VizieR Online Data Catalog: The ELM survey. VII. 15 new ELM white dwarf cand. (Brown+, 2016)”. 刊于: *VizieR Online Data Catalog* 181 (May 2016).
- [32] Philip Bull et al. “Beyond  $\Lambda$ CDM: Problems, solutions, and the road ahead”. 刊于: *Phys. Dark Univ.* 12 (2016), pp. 56–99. doi: [10.1016/j.dark.2016.02.001](https://doi.org/10.1016/j.dark.2016.02.001). arXiv: [1512.05356 \[astro-ph.CO\]](https://arxiv.org/abs/1512.05356).
- [33] Collin Capano et al. “Systematic errors in estimation of gravitational-wave candidate significance”. 刊于: *Phys. Rev.* D96.8 (2017), p. 082002. doi: [10.1103/PhysRevD.96.082002](https://doi.org/10.1103/PhysRevD.96.082002). arXiv: [1708.06710 \[astro-ph.IM\]](https://arxiv.org/abs/1708.06710).
- [34] Zu-Cheng Chen, Chen Yuan, and Qing-Guo Huang. “Non-tensorial gravitational wave background in NANOGrav 12.5-year data set”. 刊于: *Sci. China Phys. Mech. Astron.* 64.12 (2021), p. 120412. doi: [10.1007/s11433-021-1797-y](https://doi.org/10.1007/s11433-021-1797-y). arXiv: [2101.06869 \[astro-ph.CO\]](https://arxiv.org/abs/2101.06869).
- [35] Alvin J. K. Chua and Jonathan R. Gair. “Improved analytic extreme-mass-ratio inspiral model for scoping out eLISA data analysis”. 刊于: *Classical and Quantum Gravity* 32.23, 232002 (Dec. 2015), p. 232002. doi: [10.1088/0264-9381/32/23/232002](https://doi.org/10.1088/0264-9381/32/23/232002). arXiv: [1510.06245 \[gr-qc\]](https://arxiv.org/abs/1510.06245).

- [36] V. Connaughton et al. “Fermi GBM Observations of LIGO Gravitational-wave Event GW150914”. 刊于: *the Astrophysical Journal Letters* 826.1, L6 (July 2016), p. L6. doi: [10.3847/2041-8205/826/1/L6](https://doi.org/10.3847/2041-8205/826/1/L6). arXiv: [1602.03920 \[astro-ph.HE\]](https://arxiv.org/abs/1602.03920).
- [37] Jolien D. E. Creighton and Warren G. Anderson. *Gravitational-wave physics and astronomy. An introduction to theory, experiment and data analysis*. Hoboken, NJ: John Wiley & Sons, 2011, pp. xiv + 375. ISBN: 978-3-527-40886-3/hbk.
- [38] Curt Cutler. “Angular resolution of the LISA gravitational wave detector”. 刊于: *Phys. Rev. D* 57 (12 June 1998), pp. 7089–7102. doi: [10.1103/PhysRevD.57.7089](https://doi.org/10.1103/PhysRevD.57.7089). URL: <https://link.aps.org/doi/10.1103/PhysRevD.57.7089>.
- [39] G. Desvignes, R. N. Caballero, L. Lentati, et al. “High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array”. 刊于: *Monthly Notices of the Royal Astronomical Society* 458 (May 2016), pp. 3341–3380. doi: [10.1093/mnras/stw483](https://doi.org/10.1093/mnras/stw483). arXiv: [1602.08511 \[astro-ph.HE\]](https://arxiv.org/abs/1602.08511).
- [40] Hui-Min Fan et al. “Science with the TianQin observatory: Preliminary result on extreme-mass-ratio inspirals”. 刊于: *Phys. Rev. D* 102.6 (2020), p. 063016. doi: [10.1103/PhysRevD.102.063016](https://doi.org/10.1103/PhysRevD.102.063016). arXiv: [2005.08212 \[astro-ph.HE\]](https://arxiv.org/abs/2005.08212).
- [41] N. Farrow, X.-J. Zhu, and E. Thrane. “The Mass Distribution of Galactic Double Neutron Stars”. 刊于: *the Astrophysical Journal* 876, 18 (May 2019), p. 18. doi: [10.3847/1538-4357/ab12e3](https://doi.org/10.3847/1538-4357/ab12e3). arXiv: [1902.03300 \[astro-ph.HE\]](https://arxiv.org/abs/1902.03300).
- [42] F. Feroz, M. P. Hobson, and M. Bridges. “MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics”. 刊于: *Monthly Notices of the Royal Astronomical Society* 398.4 (Oct. 2009), pp. 1601–1614. issn: 00358711. doi: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x). URL: <http://mnras.oxfordjournals.org/cgi/doi/10.1111/j.1365-2966.2009.14548.x>.
- [43] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. “emcee : The MCMC Hammer”. 刊于: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), pp. 306–312. issn: 00046280. doi: [10.1086/670067](https://doi.org/10.1086/670067). arXiv: [1202.3665](https://arxiv.org/abs/1202.3665). URL: <http://www.jstor.org/stable/10.1086/670067>.
- [44] Morgane FORTIN. *Neutron star equation of state and mass-radius measurements*. <https://users.camk.edu.pl/fortin/>. 2020.

- [45] Jonathan R. Gair, Christopher Tang, and Marta Volonteri. “LISA extreme-mass-ratio inspiral events as probes of the black hole mass function”. 刊于: *Physical Review D* 81.10 (May 2010). issn: 1550-2368. doi: [10.1103/physrevd.81.104014](https://doi.org/10.1103/physrevd.81.104014). url: <http://dx.doi.org/10.1103/PhysRevD.81.104014>.
- [46] Alister W. Graham and Nicholas Scott. “THEMBH-LSPHEROIDRELATION AT HIGH AND LOW MASSES, THE QUADRATIC GROWTH OF BLACK HOLES, AND INTERMEDIATE-MASS BLACK HOLE CANDIDATES”. 刊于: *The Astrophysical Journal* 764.2 (Feb. 2013), p. 151. issn: 1538-4357. doi: [10.1088/0004-637x/764/2/151](https://doi.org/10.1088/0004-637x/764/2/151). url: <http://dx.doi.org/10.1088/0004-637X/764/2/151>.
- [47] M. J. Graham et al. “Candidate Electromagnetic Counterpart to the Binary Black Hole Merger Gravitational-Wave Event S190521g\*”. 刊于: *Physical Review Letters* 124.25, 251102 (June 2020), p. 251102. doi: [10.1103/PhysRevLett.124.251102](https://doi.org/10.1103/PhysRevLett.124.251102). arXiv: [2006.14122 \[astro-ph.HE\]](https://arxiv.org/abs/2006.14122).
- [48] Phil Gregory. *Bayesian logical data analysis for the physical sciences*. Cambridge: Cambridge University Press, 2005. isbn: 9780511791277. doi: [10.1017/CBO9780511791277](https://doi.org/10.1017/CBO9780511791277). url: <http://ebooks.cambridge.org/ref/id/CB09780511791277>.
- [49] Kayhan Gultekin et al. “THE M- $\sigma$  AND M-LRELATIONS IN GALACTIC BULGES, AND DETERMINATIONS OF THEIR INTRINSIC SCATTER”. 刊于: *The Astrophysical Journal* 698.1 (May 2009), pp. 198–221. issn: 1538-4357. doi: [10.1088/0004-637x/698/1/198](https://doi.org/10.1088/0004-637x/698/1/198). url: <http://dx.doi.org/10.1088/0004-637X/698/1/198>.
- [50] Wen-Biao Han. “Fast Evolution and Waveform Generator for Extreme-Mass-Ratio Inspirals in Equatorial-Circular Orbits”. 刊于: *arXiv e-prints*, arXiv:1609.06817 (Sept. 2016), arXiv:1609.06817. arXiv: [1609.06817 \[gr-qc\]](https://arxiv.org/abs/1609.06817).
- [51] Stephen W. Hawking, Werner Israel, and William G. Unruh. “300 Years of Gravitation”. 刊于: *Physics Today* 41.7 (1988), p. 69. issn: 00319228. doi: [10.1063/1.2811500](https://doi.org/10.1063/1.2811500). url: <http://link.aip.org/link/PHTOAD/v41/i7/p69/s1%5C&Agg=doi>.

- [52] Shun-Jia Huang et al. “Science with the TianQin Observatory: Preliminary results on Galactic double white dwarf binaries”. 刊于: *Phys. Rev. D* 102.6 (2020), p. 063021. doi: [10.1103/PhysRevD.102.063021](https://doi.org/10.1103/PhysRevD.102.063021). arXiv: [2005.07889 \[astro-ph.HE\]](https://arxiv.org/abs/2005.07889).
- [53] Piotr Jaranowski, Andrzej Krolak, and Bernard F. Schutz. “Data analysis of gravitational - wave signals from spinning neutron stars. 1. The Signal and its detection”. 刊于: *Phys. Rev. D* 58 (1998), p. 063001. doi: [10.1103/PhysRevD.58.063001](https://doi.org/10.1103/PhysRevD.58.063001). arXiv: [gr-qc/9804014 \[gr-qc\]](https://arxiv.org/abs/gr-qc/9804014).
- [54] Antoine Klein et al. “Science with the space-based interferometer eLISA: Supermassive black hole binaries”. 刊于: *Physical Review D* 93.2 (Jan. 2016). issn: 2470-0029. doi: [10.1103/physrevd.93.024003](https://doi.org/10.1103/physrevd.93.024003). url: <http://dx.doi.org/10.1103/PhysRevD.93.024003>.
- [55] John Kormendy and Luis C. Ho. “Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies”. 刊于: *Annual Review of Astronomy and Astrophysics* 51.1 (Aug. 2013), pp. 511–653. issn: 1545-4282. doi: [10.1146/annurev-astro-082708-101811](https://doi.org/10.1146/annurev-astro-082708-101811). url: <http://dx.doi.org/10.1146/annurev-astro-082708-101811>.
- [56] V. Korol et al. “Prospects for detection of detached double white dwarf binaries with Gaia, LSST and LISA”. 刊于: *MNRAS* 470 (Sept. 2017), pp. 1894–1910. doi: [10.1093/mnras/stx1285](https://doi.org/10.1093/mnras/stx1285). arXiv: [1703.02555 \[astro-ph.HE\]](https://arxiv.org/abs/1703.02555).
- [57] T. Kupfer et al. “LISA verification binaries with updated distances from Gaia Data Release 2”. 刊于: *MNRAS* 480 (1 2018), pp. 302–309. doi: [10.1093/mnras/sty1545](https://doi.org/10.1093/mnras/sty1545). url: <http://dx.doi.org/10.1093/mnras/sty1545>.
- [58] Laser Labs. *Space-Time Quest*. <https://www.laserlabs.org/spacetimequest.php>. 2017.
- [59] H. Li et al. “Probing Primordial Gravitational Waves: Ali CMB Polarization Telescope”. 刊于: *ArXiv e-prints* (Oct. 2017). arXiv: [1710.03047](https://arxiv.org/abs/1710.03047).
- [60] Q. Li et al. “Measurements of the gravitational constant using two independent methods”. 刊于: *Nature* 560 (Aug. 2018), pp. 582–588. doi: [10.1038/s41586-018-0431-5](https://doi.org/10.1038/s41586-018-0431-5).

- [61] Xiang Li et al. “Neutron Star–Black Hole Coalescence Rate Inferred from Macronova Observations”. 刊于: *Astrophys. J.* 844.2 (2017), p. L22. doi: [10.3847/2041-8213/aa7fb2](https://doi.org/10.3847/2041-8213/aa7fb2). arXiv: [1611.01760 \[astro-ph.HE\]](https://arxiv.org/abs/1611.01760).
- [62] LIGO and Virgo collaboration. *LIGO Document P1200087-v19*. <https://dcc.ligo.org/LIGO-P1200087-v19/public>. 2014.
- [63] LIGO and Virgo collaboration. *LIGO Document T1800044*. <https://dcc.ligo.org/LIGO-T1800044/public>. 2018.
- [64] Jifeng Liu et al. “A wide star-black-hole binary system from radial-velocity measurements”. 刊于: *Nature* 575.7784 (Nov. 2019), pp. 618–621. doi: [10.1038/s41586-019-1766-2](https://doi.org/10.1038/s41586-019-1766-2). arXiv: [1911.11989 \[astro-ph.SR\]](https://arxiv.org/abs/1911.11989).
- [65] Shuai Liu, Yi-Ming Hu, Jian-dong Zhang, and Jianwei Mei. “Science with the TianQin observatory: Preliminary results on stellar-mass binary black holes”. 刊于: *Phys. Rev. D* 101.10 (2020), p. 103027. doi: [10.1103/PhysRevD.101.103027](https://doi.org/10.1103/PhysRevD.101.103027). arXiv: [2004.14242 \[astro-ph.HE\]](https://arxiv.org/abs/2004.14242).
- [66] Shuai Liu et al. “Capability for detection of GW190521-like binary black holes with TianQin”. 刊于: *Phys. Rev. D* 105.2 (2022), p. 023019. doi: [10.1103/PhysRevD.105.023019](https://doi.org/10.1103/PhysRevD.105.023019). arXiv: [2110.05248 \[astro-ph.HE\]](https://arxiv.org/abs/2110.05248).
- [67] Jun Luo et al. “TianQin: a space-borne gravitational wave detector”. 刊于: *Class. Quant. Grav.* 33.3 (2016), p. 035010. doi: [10.1088/0264-9381/33/3/035010](https://doi.org/10.1088/0264-9381/33/3/035010). arXiv: [1512.02076 \[astro-ph.IM\]](https://arxiv.org/abs/1512.02076).
- [68] LVC. “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”. 刊于: *arXiv preprint arXiv:2010.14527* (2020), p. 1. arXiv: [2010.14527 \[astro-ph\]](https://arxiv.org/abs/2010.14527). URL: <https://arxiv.org/abs/2010.14527>.
- [69] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford Master Series in Physics. Oxford University Press, 2007. ISBN: 978-0-19-857074-5, 978-0-19-852074-0.
- [70] Michele Maggiore. *Gravitational Waves. Vol. 2: Astrophysics and Cosmology*. Oxford University Press, 2018. ISBN: 9780198570899. URL: <https://global.oup.com/academic/product/gravitational-waves-9780198570899?cc=de&lang=en&#>.

- [71] R. N. Manchester, G. Hobbs, M. Bailes, W. A. Coles, et al. “The Parkes Pulsar Timing Array Project”. 刊于: *Publications of the Astronomical Society of Australia* 30, e017 (Jan. 2013), e017. doi: [10.1017/pasa.2012.017](https://doi.org/10.1017/pasa.2012.017). arXiv: [1210.6130 \[astro-ph.IM\]](https://arxiv.org/abs/1210.6130).
- [72] David Merritt. “Loss-cone Dynamics”. 刊于: *Class. Quant. Grav.* 30 (2013), p. 244005. doi: [10.1088/0264-9381/30/24/244005](https://doi.org/10.1088/0264-9381/30/24/244005). arXiv: [1307.3268 \[astro-ph.GA\]](https://arxiv.org/abs/1307.3268).
- [73] J. Ming et al. “Results from an Einstein@Home search for continuous gravitational waves from Cassiopeia A, Vela Jr., and G347.3”. 刊于: *Physical Review D* 100.2, 024063 (July 2019), p. 024063. doi: [10.1103/PhysRevD.100.024063](https://doi.org/10.1103/PhysRevD.100.024063). arXiv: [1903.09119 \[gr-qc\]](https://arxiv.org/abs/1903.09119).
- [74] C. J. Moore, R. H. Cole, and C. P. L. Berry. “Gravitational-wave sensitivity curves”. 刊于: *Classical and Quantum Gravity* 32.1, 015014 (Jan. 2015), p. 015014. doi: [10.1088/0264-9381/32/1/015014](https://doi.org/10.1088/0264-9381/32/1/015014). arXiv: [1408.0740 \[gr-qc\]](https://arxiv.org/abs/1408.0740).
- [75] Rendong Nan et al. “The Five-Hundred Aperture Spherical Radio Telescope (fast) Project”. 刊于: *International Journal of Modern Physics D* 20.6 (Jan. 2011), pp. 989–1024. doi: [10.1142/S0218271811019335](https://doi.org/10.1142/S0218271811019335). arXiv: [1105.3794 \[astro-ph.IM\]](https://arxiv.org/abs/1105.3794).
- [76] G. Nelemans, L. R. Yungelson, and S. F. Portegies Zwart. “The gravitational wave signal from the Galactic disk population of binaries containing two compact objects”. 刊于: *Astronomy and Astrophysics* 375 (Sept. 2001), pp. 890–898. doi: [10.1051/0004-6361:20010683](https://doi.org/10.1051/0004-6361:20010683). eprint: [astro-ph/0105221](https://arxiv.org/abs/astro-ph/0105221).
- [77] P. C. Peters and J. Mathews. “Gravitational Radiation from Point Masses in a Keplerian Orbit”. 刊于: *Physical Review* 131 (July 1963), pp. 435–440. doi: [10.1103/PhysRev.131.435](https://doi.org/10.1103/PhysRev.131.435).
- [78] P. C. Peters and J. Mathews. “Gravitational Radiation from Point Masses in a Keplerian Orbit”. 刊于: *Phys. Rev.* 131 (1 July 1963), pp. 435–440. doi: [10.1103/PhysRev.131.435](https://doi.org/10.1103/PhysRev.131.435). URL: <https://link.aps.org/doi/10.1103/PhysRev.131.435>.

- [79] M. Pitkin. “psrqpy: a python interface for querying the ATNF pulsar catalogue”. 刊于: *Journal of Open Source Software* 3.22 (Feb. 2018), p. 538. doi: [10.21105/joss.00538](https://doi.org/10.21105/joss.00538). URL: <https://doi.org/10.21105/joss.00538>.
- [80] E. Poisson and C.M. Will. *Gravity: Newtonian, Post-Newtonian, Relativistic*. Cambridge University Press, 2014. ISBN: 9781107032866. URL: <https://books.google.com.tw/books?id=PZ5cAwAAQBAJ>.
- [81] Wolfgang Rindler. 相对论精义——狭义、广义和宇宙学相对论. Trans. by 方励之江山. 合肥: 安徽科学技术出版社, 1977.
- [82] G. H. A. Roelofs et al. “Hubble Space Telescope Parallaxes of AM CVn Stars and Astrophysical Consequences”. 刊于: *the Astrophysical Journal* 666 (Sept. 2007), pp. 1174–1188. doi: [10.1086/520491](https://doi.org/10.1086/520491). arXiv: [0705.3855](https://arxiv.org/abs/0705.3855).
- [83] R. Ruffini and M. Sasaki. “On a Semirelativistic Treatment of the Gravitational Radiation From a Mass Thrusted Into a Black Hole”. 刊于: *Prog. Theor. Phys.* 66 (1981), pp. 1627–1638. doi: [10.1143/PTP.66.1627](https://doi.org/10.1143/PTP.66.1627).
- [84] B. S. Sathyaprakash and Bernard F. Schutz. “Physics, Astrophysics and Cosmology with Gravitational Waves”. 刊于: *Living Reviews in Relativity* 12.1 (Mar. 2009), p. 2. ISSN: 1433-8351. doi: [10.12942/lrr-2009-2](https://doi.org/10.12942/lrr-2009-2). URL: <https://doi.org/10.12942/lrr-2009-2>.
- [85] V. Savchenko et al. “INTEGRAL Upper Limits on Gamma-Ray Emission Associated with the Gravitational Wave Event GW150914”. 刊于: *the Astrophysical Journal Letters* 820.2, L36 (Apr. 2016), p. L36. doi: [10.3847/2041-8205/820/2/L36](https://doi.org/10.3847/2041-8205/820/2/L36). arXiv: [1602.04180 \[astro-ph.HE\]](https://arxiv.org/abs/1602.04180).
- [86] B. Schutz. *A First Course in General Relativity*. Cambridge University Press, 2009. ISBN: 9781139479004.
- [87] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi. “LINKING THE SPIN EVOLUTION OF MASSIVE BLACK HOLES TO GALAXY KINEMATICS”. 刊于: *The Astrophysical Journal* 794.2 (Sept. 2014), p. 104. ISSN: 1538-4357. doi: [10.1088/0004-637X/794/2/104](https://doi.org/10.1088/0004-637X/794/2/104). URL: [http://dx.doi.org/10.1088/0004-637X/794/2/104](https://dx.doi.org/10.1088/0004-637X/794/2/104).

- [88] Changfu Shi et al. “Science with the TianQin observatory: Preliminary results on testing the no-hair theorem with ringdown signals”. 刊于: *Phys. Rev.* D100.4 (2019), p. 044036. doi: [10.1103/PhysRevD.100.044036](https://doi.org/10.1103/PhysRevD.100.044036). arXiv: [1902.08922 \[gr-qc\]](https://arxiv.org/abs/1902.08922).
- [89] Devinderjit S. Sivia and John Skilling. *Data analysis. A Bayesian tutorial. 2nd ed.* English. Oxford: Oxford University Press., 2006.
- [90] M. Soares-Santos et al. “First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo BinaryBlack-hole Merger GW170814”. 刊于: *the Astrophysical Journal Letters* 876.1, L7 (May 2019), p. L7. doi: [10.3847/2041-8213/ab14f1](https://doi.org/10.3847/2041-8213/ab14f1). arXiv: [1901.01540 \[astro-ph.CO\]](https://arxiv.org/abs/1901.01540).
- [91] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, et al. “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters”. 刊于: *ApJS* 148 (Sept. 2003), pp. 175–194. doi: [10.1086/377226](https://doi.org/10.1086/377226). eprint: [astro-ph/0302209](https://arxiv.org/abs/astro-ph/0302209).
- [92] The LIGO Scientific Collaboration et al. “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run”. 刊于: *arXiv e-prints*, arXiv:2111.03606 (Nov. 2021), arXiv:2111.03606. arXiv: [2111.03606 \[gr-qc\]](https://arxiv.org/abs/2111.03606).
- [93] The LIGO Scientific Collaboration et al. “Narrow-band search for gravitational waves from known pulsars using the second LIGO observing run”. 刊于: *arXiv e-prints*, arXiv:1902.08442 (Feb. 2019), arXiv:1902.08442. arXiv: [1902.08442 \[gr-qc\]](https://arxiv.org/abs/1902.08442).
- [94] Maarten van de Meent. “Gravitational self-force on generic bound geodesics in Kerr spacetime”. 刊于: *Physical Review D* 97.10, 104033 (May 2018), p. 104033. doi: [10.1103/PhysRevD.97.104033](https://doi.org/10.1103/PhysRevD.97.104033). arXiv: [1711.09607 \[gr-qc\]](https://arxiv.org/abs/1711.09607).
- [95] Hai-Tian Wang et al. “Science with the TianQin observatory: Preliminary results on massive black hole binaries”. 刊于: *Phys. Rev.* D100.4 (2019), p. 043003. doi: [10.1103/PhysRevD.100.043003](https://doi.org/10.1103/PhysRevD.100.043003). arXiv: [1902.04423 \[astro-ph.HE\]](https://arxiv.org/abs/1902.04423).
- [96] J. M. Weisberg and J. H. Taylor. “Relativistic Binary Pulsar B1913+16: Thirty Years of Observations and Analysis”. 刊于: *Astronomical Society of the Pacific Conference Series* 328 (2005), p. 25. arXiv: [0407149 \[astro-ph\]](https://arxiv.org/abs/0407149). URL: <http://arxiv.org/abs/astro-ph/0407149>.

- [97] Liang-Gui Zhu et al. “Constraining the cosmological parameters using gravitational wave observations of massive black hole binaries and statistical redshift information”. 刊于: *Phys. Rev. Res.* 4.1 (2022), p. 013247. doi: [10.1103/PhysRevResearch.4.013247](https://doi.org/10.1103/PhysRevResearch.4.013247). arXiv: [2104.11956 \[astro-ph.CO\]](https://arxiv.org/abs/2104.11956).
- [98] Liang-Gui Zhu et al. “Constraining the Hubble constant to a precision of about 1% using multi-band dark standard siren detections”. 刊于: *Sci. China Phys. Mech. Astron.* 65.5 (2022), p. 259811. doi: [10.1007/s11433-021-1859-9](https://doi.org/10.1007/s11433-021-1859-9). arXiv: [2110.05224 \[astro-ph.CO\]](https://arxiv.org/abs/2110.05224).
- [99] 俞允强. 电动力学简明教程. 北京: 北京大学出版社, 1997.
- [100] 方励之. 普适性一例——伽利略相对性原理. 2012. URL: <http://fang-lizhi.hxwk.org/2012/01/02/>.