

---

## 6引力波数据分析

引力波信号非常弱，引力波天文学研究不但需要高灵敏度、宽频带的探测器,以及高超的探测技术,也需要科学的数据处理和分析方法,引力波数据处理、分析方法及相应的通用软件的研发必须与干涉仪的建造同步进行，引力波数据处理与分析已在世界各地迅速发展起来，成为引力波天文学研究的锐利武器.为引力波地发现立下汗马功劳.

引力波数据分析主要由数据文件的生成和物理结果的获得两大部分组成，前者称为数据处理，包括在通用软件建立、线数据获取、刻度和离线分析，后者称为物理分析，利用生成的数据文件获得物理结果

---

## 6.1 引力波数据分析的资源

数据分析通用工具软件和数据文件是引力波数据分析的两大主要资源

### 6.1.1 数据分析通用工具软件

做数据分析并不需要每个人都要自己编写所有的程序, 为了节省时间和精力, 也为了便于讨论和交流, 各研究单位都需要建立大量通用的软件系统。供组内成员使用。如工具软件、模拟计算软件、存放刻度常数修正参数和探测器参数的数据库、管理和控制作业进度的软件系统、事例筛选和重建系统、针对专门研究领域的分析流水线等。

在做数据分析之前, 必须对它们的功能、格式、内容、操作方法等进行深入透彻的学习和理解, 才能得心应手, 达到事半功倍的效果。

---

每个实验室都有自己的一整套完善的通用软件体系，其中每个程序包都含有百行甚至千行万行以上的源程序。这些通用工具软件是由专门的人员负责编写、管理、改进和升级的。

**LIGO 三个专门小组，**

**(1)LIGO数据分析系统LDAS (LIGO Data Analysis System)**

**(2) 模型化和模拟计算(Modeling and Simulation )**

**(3) 通用计算(General Computing)**

各大实验室的软件系统都是数十年来几代人辛勤劳动的结晶，在国际上有很大的通用性互相借鉴和移植。

---

## 6.1.2数据文件

数据文件及其生成主要包括以下几个方面

### (1)数据获取硬件和控制软件

激光干涉仪是引力波探测器的主要部分是光学仪器。数据获取首先要把光信号抽取出来，转换成电信号并进行数字化，这个获取数字信号的过程的硬件设备有光路、光电转换器件、读出电子学和在线控制插件等

光路由平面镜和透镜组成，它将需要的光信号抽取出来，引导到需要的方向和地点并汇聚在光探测器上。光探测器将光信号转换成电信号，电信号经过读出电子学放大、甄别和成形之后，进行模数转换，然后通过在线计算机控制的部件记录在存储设备上。

硬件系统还包括标定系统，定期对电子系统进行定标，以确保数据的精度。

所有数据获取硬件都置于一套软件系统的控制之下

---

## (2) 辅助系统

除了干涉仪本身的数据之外，在引力波探测中还使用了众多的各种独立地对探测器的环境和探测器状态进行监控的设备和记录仪，对诸如温度、气压、风力、大雨、冰雹、地表震动、声响、电场、磁场等环境条件进行监测，以及对引力波探测器内部的平面镜和透镜的位置等探测器自态进行监测。这些监控探测器的信号也都转化成数字信号，以数据的形式纪录和保存，供后续和在线的数据分析时使用。

---

### (3) 时间标志

无论是激光干涉仪上获取的数据，还是物理环境的监测设备所纪录的环境数据，都需要统一时间标志以便使所有探测器同步，表明数据所表示的信号的发生时间。有这个时间标志，可以对坐落在世界各地的多个探测器在同一时间段探测到的信号进行关联分析，获得更多的关于引力波的准确信息。

数据获取系统的这个全球同步的时钟(时间标志) 和时间发布设备也是数据获取的重要设备之一；整个数据获取系统由一个软件包控制，控制室的数据取样分析和显示包括在内

数据获取系统常被称为在线系统

---

## (4) 数据预处理系统

从引力波探测器获取的**原始数据**都是一些数字，不能直接拿来物理分析，要进行基本的**整理加工**，去掉无用的垃圾，用标准信号对探测器及电子学进行刻度，对特殊信息行标示等。也就是说在进行实质性的物理分析之前，要先做大量的、繁杂的、重复性的工作，在世界各个实验室都建立了专门的小组承担这项任务，**按照统一的要求和格式把数据组织起来，加入一定的标识符，形成数据文件**。该文件除了探测器获取的信息外，还包含后续分析所需要的足够丰富的资料。

**把原始数据转变为直接用来做物理分析的数据文件的过程又称为数据的离线分析**

---

## \* LIGO的数据文件的内容

第1. 提供了数据的获取时间，即时间标签。以**timeline** 的形式，标识出哪段时间里**LIGO** 探测器的数据是可以被用户利用的。同时还给出在给定时间段里可以使用的数据元的百分比；这个百分比越高，表明这段时间里探测器的状态越稳定，可靠数据的比例越大。

第2, 包含对探测器的标识信息，以表明数据来自哪个探测器。**LIGO** 数据含有来自**5** 个不同探测器的观测数据，分别标识为**H1, H2, L1, V1, 和G1**

第3为了科学研究，**LIGO** 在数据获取时，有针对性地加入一些在理论上具有已知特征的引力波，用以考察和测试开发的分析软件和算法是否能捕捉到这类事例。为了把这类人工植入的数据事件与真实采集到的信号加以区别，**LIGO** 在数据文件中对这类事件做了标识。

第4. 激光干涉仪和其他环境监测设备在运行中会有不稳定的瞬时故障，这些仪器的小故障会带来一些类似引力波产生的信号事件(**instrumental glitches**)，这类事件不是来源于天体活动，而是来源于探测器设备或周围环境的变化。它们虽然是虚假触发，但可能与天体活动产生的事件非常相似。通过适当的分析手段(称为**veto study**)，这些事例能够被判断为虚假事例，它们需要在数据中加以标识，以减少对后续数据分析过程的干扰。

第5.数据的取样率。取样率与数据获取设备的处理能力和速度相关，也是数据的重要特征信息。

第6.数据质量分类信息，在数据采集过程中对设备和系统运行状态的监控能给出当前时刻设备的状态是否稳定，以及环境干扰成分是否会影响到数据的质量这些信息。如果设备出现不稳定的情况或环境干扰过大,得到的数据就是不可靠的,采用打数据质量标签的方法注释。这些信息以不同的格式表示和存储，在访问时通过数据文件中给出，在**LIGO**数据中,用一位**2** 进制数作为一个标识。用**18** 位标识数据质量,用**6**位标识植入数据信息。



---

## 6.2 数据道和数据单元

激光干涉仪引力波探测器可以看成一个准线性装置，输入到干涉仪的信号有，引力波，地面震动，光量子噪声，热噪声，电子学噪声等，干涉仪的输出信号有引力波信号，校正信号，共模信号声，环境噪声等

### 6.2.1 数据道

探测器输出各种不同类型的信号，统称为数据。在干涉仪的不同部位获取的数据及不同的环境监视器上获取的不同的数据, 分别给予不同的名称，称之为数据道。激光干涉仪引力波探测器及辅助设备的数据通道共有数百个，分别具有不同的物理或技术含义。相同数据道的数据存储在一起，形成数据文件。

---

## 6.2.2 数据单元

探测器要连续运转相当长的时间，如数日、数周甚至数月，各个数据通道的数据源源不断地输出。为了便于记录和进行物理分析，需要把数据流切割成小段，以这个小数据段为单位进行记录，这种小数据段被称为数据元。它是数据的基本存储单位。每个数据元都有自己的编号(即名称)，相邻数据元的编号是连续的，它们在时间上形成一个连续序列。储存在数据元内的主要数据是在选定的时间间隔内的时间序列，也可以在数据元内放其他类型的数据，如：谱形、表格、矢量或阵列等。为了便于世界各地的引力波探测器联网，以进行符合测量，国际上对数据元的格式和结构有统一的规定。

### 6.2.3数据道与数据单元的关系

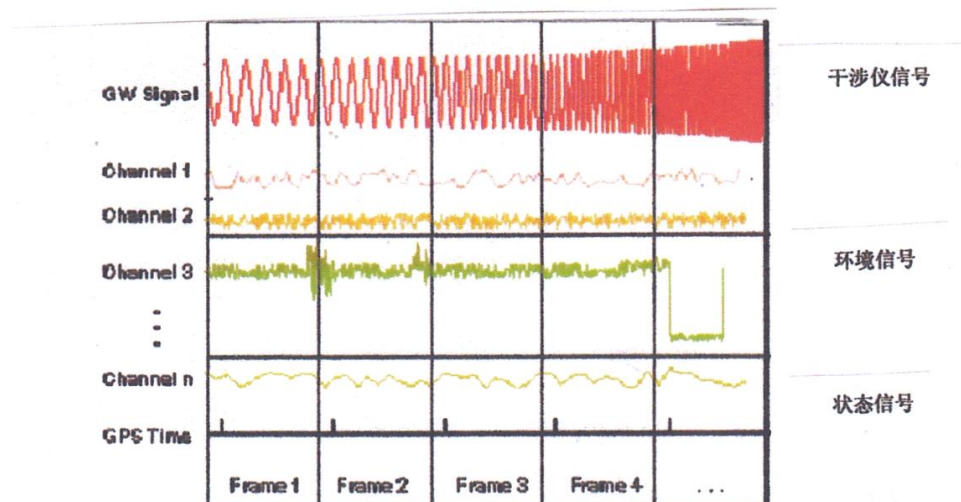


图5.1 数据道与数据元的关系

图中每一行代表一个数据道，两侧给出的是数据道的编号名称和数据道的类别。底部显示的是数据单元的划分

---

激光干涉仪引力波探测器的数据道有几百种甚至上千种, 它们分别表示不同的物理含义. 如

表示对干涉仪臂长变化敏感的数据道:

LI : LSC-AS \_\_ Q; LI : LSC-AS \_\_ I; LI : LSC - POB \_\_ Q; LI : LSC - POB \_\_ I LI :  
LSC - REFL \_\_ Q; LI : LSC - REFL \_\_ I; LI : LSC - DARM \_\_ CTRL; -----

与干涉仪状态和统调有关的数据道有:

LI : LSC - WFS1 \_\_ QY; LI : LSC - WFS1 \_\_ QP; LI : LSC - QPDX \_\_ Y; -----

表示环境监测的数据道有:

LO: PEM-LVEA\_\_SEISX; LO: PEM-LVEA\_\_SEISY; LO: PEM-PLS1\_\_MIC; -----

研究不同的课题要正确地选用不同类型的数据道

## 6.2.4数据元的图形

数据文件中，由于数据获取位置及探测器的瞬时状态不同，数据单元的具体图形是比较复杂的，满足一定选择条件的数据元的图形如下

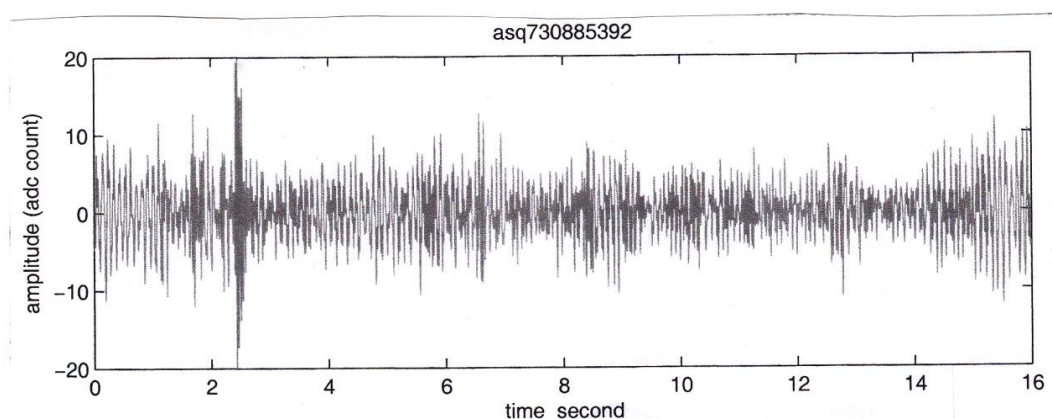


图5.3 满足选择条件的被选着出来的有用数据元

图9.7展示的是LIGO数据道LI : LSC—AS — Q中标号为asq730885392的一个数据元. 它的横坐标是时间，纵坐标是ADC的计数. 图中的尖锐突起起是我们要选出来的. (这是在做否决权研究——Veto study时我发现的一个事例，由于它在该研究中非常典型，故被组内命名为运永A事例)

数据分析的第一步就是要从海量的数据中选出有用的符合条件事例，然后进行详细分析，得出物理结果

---

## 6.3 数据分析方法

引力波探测属于类声探测，数据分析方法很多，除了以下列出的最基本的方法之外，很多人还有自己特有的。

### 6.3.1 波形分析法

激光干涉仪引力波探测器是一个宽频带幅度探测器它探测的是波形。实验物理学家在引力波探测中不用“天体物理源”的概念来思考问题，而使用“波形结构形态学”不同的天体源辐射出不同的波形。波形分析是引力波数据分析的基础。

#### \*\* 研究对象

在引力波数据分析中，波形结构形态学研究的对象主要有以下几个方面。

- (1). 具有有限时间间隔的连续性(或爆发性) 引力波
- (2) 未知的脉冲形引力波
- (3) 窄频带、连续的、周期性引力波
- (4) 随机背景辐射
- (5) 未知的引力波波型形

#### \*\*基本方法

- 根据自己确定的条件筛选事例，b.与理论模型对照
- 注意陌生波形

波形分析是快速进入引力波数据分析的入门和捷径

---

### 6.3.2 时间 - 频率分析法

在引力波探测中大量事件的波形是未知的，或者说，在足够精确度下是未知的，它们不能用匹配过滤器进行分析。时间 - 频率分析给我们提供了一种有用的方法，该方法不必预先知道波形的相关知识，而是直接去寻找感兴趣的引力波事例。

#### **\*\*基本思想**

时间 - 频率分析法的基本思路是这样的：在固定的时间间隔内把探测到的数据切割成片，对这些数据切片进行福里叶展开。利用计算机把相应的频率、时间和功率画成三维图，将自己确定的频率间隔内存在的功率与已知的噪声功率统计分布进行比较，寻找与一般噪声信号相貌不同的东西。还可以根据自己建立的条件进行判选，找出感兴趣的新奇事例。

## \*\*频率、时间、功率的三维图形

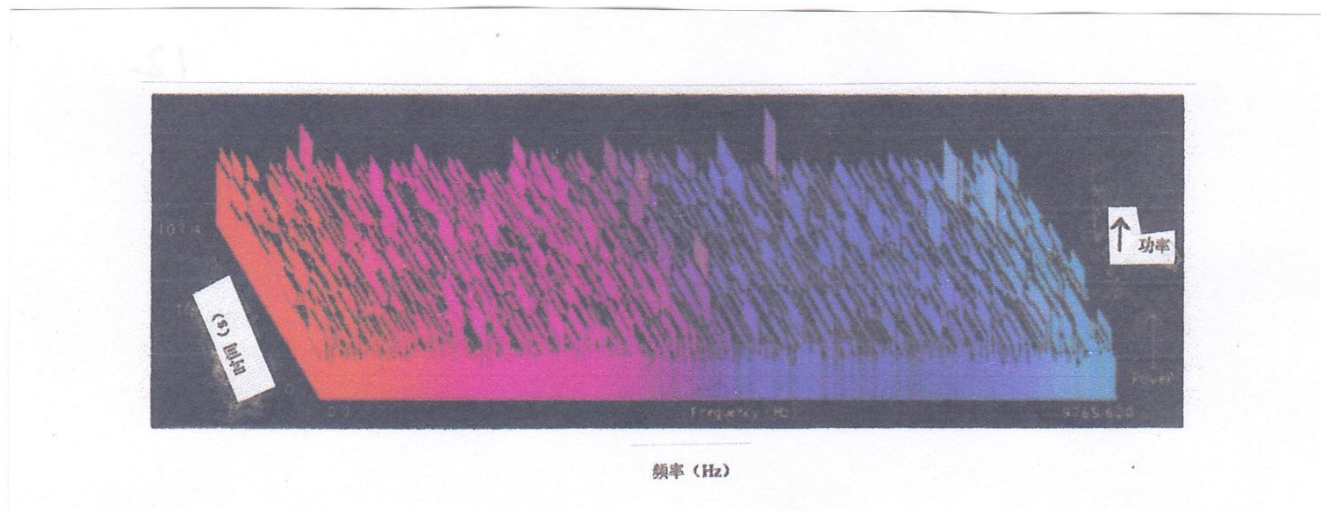


图5.3频率、时间、功率三维图

图中横轴为频率，纵轴为时间，高度为功率



---

## \*\*三维图形分析手段

### (1).超功率法:

将选定频带、选定时间段内数据中存在的功率与己的噪声功率的统计分布进行比较，确定一个阈值功率，搜寻超过阈值功率的事件，如果探测器的噪声是稳定的且具有高斯分布，那么噪声功率将遵循 $\chi^2$ 分布，其自由度数量等于时间 - 频率值的两倍。因此这种方法在探测与引力波强度有关的超过阈值功率时，它的效率取决于信号的期待持续时间和频带宽度, 也与强度有关。

---

## (2).TF Cluster方法:

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

下面我们举例说明该分析方法的实际效果例：

利用蒙特 - 卡罗计算方法，模拟在白色高斯噪声本底下产生一个双星旋绕Chirp信号，得到一批模拟数据,对这些模拟数据进行频率 - 时间分析,看看能否抽取出这个Chirp信号

## **\*\*第一步：对小单元功率设置阈值.**

- A) 在时间—频率分解数据中对小单元的功率设置阈值.
- B) 只保留超过功率阈值的小单元相互连接而构成的集团.
- C) 对集团的总功率设置阈值后获取的结果.

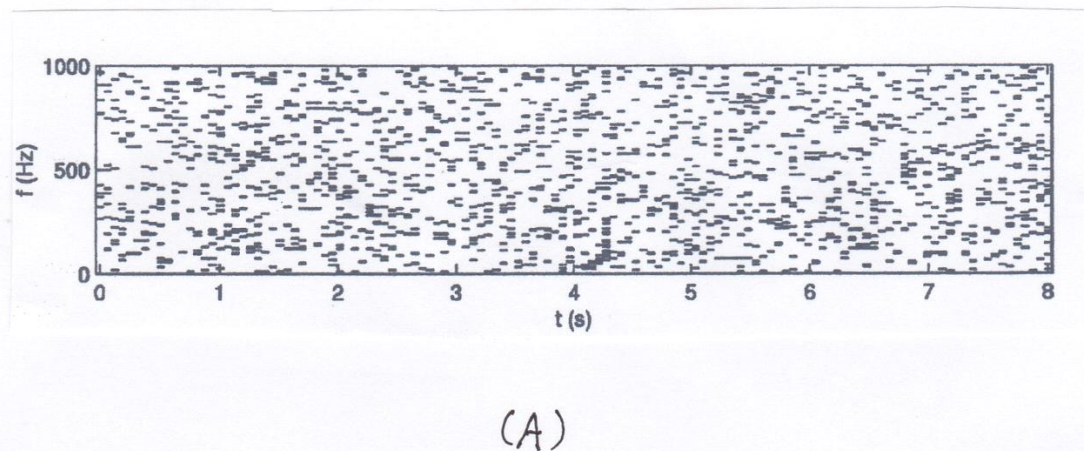
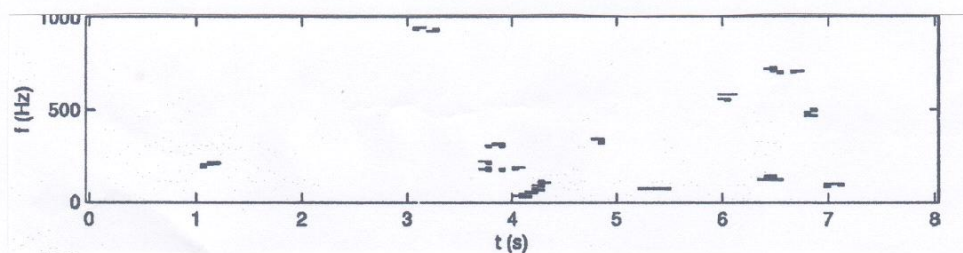


图5.4A 对小单元的功率设置阈值后的分析结果.

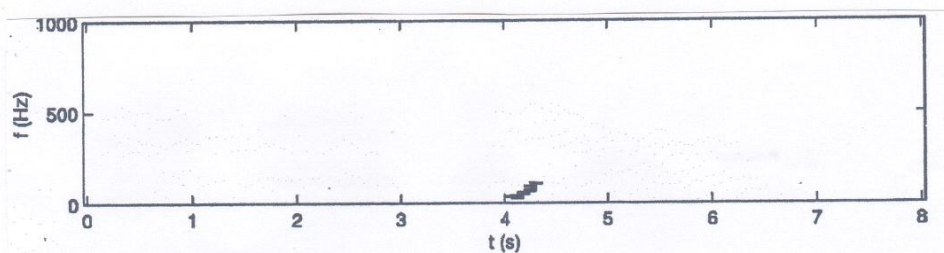
## \*第二步只保留过阈的小单元相互连接而构成的集团



(B)

图5.4B时间-频率分解数据中只保留过阈的小单元相互连接而构成的集团.

## \*第三步：对集团的总功率设置阈值



(C)

图5.4C. 时间-频率分解数据中对上述集团的总功率设置阈值后获取的结果.

**提取出Chirp信号**

---

### 6.3.3 变化点分析法

不需要预先知道波形的有关知识就能辨认引力波信号另外一种这就是变化点分析法.

#### **\*\*变化点分析法原理**

在时间域内寻找探测器输出数据的**统计学参数的变化点**

#### **\*\*具体做法**

把输出数据分成很多小组, 每个数据小组内统计学参数近似为常数。这个统计学参数用正态分布的平均值和方差来表示, 变化点定义为噪声特性(以平均值和方差表示)变化的时间点, 如果在这个时间点的随便哪一边具有不同的统计分布值, 且其统计学参数值超过给定的阈我们就可以选择它。

**这种被辨认出的变化点划定了一个数据小组的起始和终止时间点。数据小组的统计学参数以起始和终止时间内的平均值和方差来表示。**

---

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

---

## 6.4 模板分析法

模板分析法是引力波数据分析中最得力的方法，关键问题是建立描述引力波的函数，从而建立模板库。

### 6.4.1 引力波函数的建立

模板分析法的基础是建立引力波函数  $S(t)$ 。在引力波的理论研究中，有些研究对象的结构和作用机制是清楚的，物理图像是直观的，如密近双星旋绕问题。它们辐射的引力波可以用一个确定的时间函数，如  $S(t)$  来表示。有些对象我们在理论上研究的并不透彻，对其结构和物理机制了解的不太清楚，但可以根据已有的知识建立起一个我们认为恰当的物理模型，推导出所辐射的引力波的时间函数  $S(t)$  和图像，还有的研究对象我们知之甚少，只能根据自己的物理知识和经验，参考他人的论述，建立起自己认为可能的物理模型，推导出自认为可能的引力波函数  $S(t)$

---

## 6.4.2. 模板组合（模板库）

无论哪种情况，我们都认为这个描述引力波的函数具有已知的形式  $S(t)$ ，它是我们在理论上预言的形式，一旦碰到它，我们能认出来。我们称这个函数所包含的无数的引力波波形是我们要探测的引力波的“模板”

在实际应用中，通常把模板写成由有限个分立子模板构成的一个模板组合，根据实际需要，模板组合的规模可大可小，有的含有几千块子模板，有的含有成几万块甚至几十万块，上百万块子模板。

## 6.4.3模板分析法原理

在物理分析过程中，我们以模板为标准，逐一筛查获取的带有噪声的数据记录，寻找与模板类似的图样。含有这种图样的记录其强度和式样均与噪声单独存在时的图样不同。

用模板法做引力波数据分析，实际上是研究从探测器上获得的时间记录与我们的模板之间的匹配关系

\*\*\*\*\*



---

### 6.4.4 函数的互相关运算

模板分析的具体办法是在获取的数据与模板之间进行互相关运算。

#### (1) 函数间互相关运算的定义

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

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

互相关运算的涵义是：对于每一个给定的时间  $\tau$ ，先让函数  $S_2(t)$  在时间上偏移  $\tau$  后与函数  $S_1(t)$  相乘，乘积运算针对时间范围  $[-\infty, \infty]$  内所有被记录的数据进行完之后，再把得到的所有乘积加在一起。对于不同的  $\tau$  值，这些乘积之和是不同的，它随  $\tau$  值的变化而变化。 $S_1 * S_2(\tau)$  是偏移时间  $\tau$  的函数。我们称这个函数为时间函数  $S_1(t)$  和  $S_2(t)$  的互相关函数。

---

求解互相关函数  $S_1 * S_2(\tau)$  的过程称为函数  $S_1(t)$  和  $S_2(t)$  的互相关运算. 互相关性是表示两个函数  $S_1(t)$  和  $S_2(t)$  相互关联程度的量度。

## (2) 自相关运算

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

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

$S * S(\tau)$  是时间函数  $S(t)$  的自相关函数。自相关函数是测量一个时间函数的两个副本之间、在各种不同的时间偏移时函数自身相关程度的方式, 当函数自身在时间上对准时, 即当  $\tau = 0$  时, 它的自相关函数  $S * S$  有最大值,

---

### (3)模板 $S(t)$ 与数据时间记录 $V(t)$ 的互相关运算

设我们建立的模板为  $S(t)$ ，它是我们要寻找的信号。  
若从探测器上获得的数据时间记录为  $V(t)$ ，则它们之间的互相关函数为：

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

对每个获取的数据信号进行运算，得到一个互相关函数系列

。

---

## 6. 4. 5无噪声的时间序列,

如果获得的探测器输出数据序列是由真实信号自身产生的(这等效于真实信号和模板相同)而且设模板函数为 $S(t)$ , 这时模板自己引起的探测器输出信号为

$$V(t) = \alpha S(t - t_0)$$

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

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

下面我们对不同情况下此互相关函数的值进行分析

- 
- (1) 一般情况下此值大于任何偏离完全符合时的值。
- (2) 如果模板对准了无任何输出信号的时间段，即此时既无号也无噪声输出，即  $V(t) = 0$ ，则互相关函数的值也为零。

由于我们的探测器总是有噪声的，模板的互相关函数值实际上不为零，且互相关函数  $V * S(t)$  是另一个有噪声的时间序列函数，如果探测器输出的噪声具有高斯分布：

$$P(v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-v^2/2\sigma^2}$$

那么只要没有信号只有噪声存在， $V * S(t)$  的分布也是高斯型的。

---

(3)如果遇到一个足够大的有用信号，则探测器的输出就与模板匹配得很好，这时 $V * S(t)$ 有一个特别大的值，在 $V * S(t)$ 的直方图中它会明显地落在由噪声事例填充的各个道之外，任何一个这样的事例都是一个合适的候选信号，因为如果互相关值很大，就不太可能是由噪声自身单独生成的。

---

#### (4) 信号噪声比

对于给定的模板  $S(t)$ ，我们可以把噪声定义为：

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

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

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

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

信号噪声比  $SNR$  是信号存在时的  $S^2$  值与噪声单独存在时的  $N^2$  之比的平方根。

$$SNR \equiv \sqrt{S^2 / N^2}$$

---

信号噪声比  $SNR$  给出了在探测器的输出中出现含有噪声以外其他信号的可能性有多大.  $SNR$  愈大, 输出中含有其他信号的可能性愈大, 也就是大的  $SNR$  值指明, 在该输出时间序列函数中, 存在着噪声以外的东西. 它可能是要探测的引力波信号, 也可能是其他干扰信号.

模板分析法的一个重要工作是建立模板组合, 也叫模板库, 它是利用公式中各个参数的不同值, 计算出具体的波形(子模板)的组合。做数据分析时, 数据文件中的每个数据都要用模板库中的每块子模板进运算。参数值分割的越细, 子模板数量越多, 则数据分析的精确度越高。但是, 计算机机时也是大的惊人。

模板分析法是利用准备好的数据文件与模板进行互相关算, 可以在不同的时间, 不同的地点, 对不同的物理问题进行研究. 当处理的噪声是“白噪声”时模板方法是最佳的选择.



---

## 6.5 匹配过滤器

模板分析法用于对已经记录在介质（如磁带、磁盘、磁鼓等）上的数据进行物理分析。属于数据处理的离线分析范畴。匹配过滤器既能用于离线分析又能直接连在探测器的输出端，对探测器的输出进行实时处理，完成互相关运算

### 6.5.1 基本思想

匹配滤波器是引力波数据分析中功能最强大的工具之一，在这个方法里，匹配滤波器接在探测器输出端，将干涉仪输出信号直接与“模板”进行相关分析，再按照探测器的反噪声谱密度进行加权，如果一个信号事件与“模板”的相关值较高，就说明观测到的信号不完全是噪声的可能性比较高，而且有可能是与该模板所关联的引力波信号。遇到这样的事例时，控制室马上给出警示得到一个触发信号。

---

## 6.5.2 对数据进行过滤

匹配滤波器有一个非常有用的性质：由于干涉仪的输出信号可以表示为单纯的引力波信号和随机噪声的和，干涉仪输出信号与给定模板的相关性计算得到的相关值中可能找到一个最优的信噪比。它就是我们选出的值得进一步分析的有用“事例”——快速挑选

### \*\*褶积运算的定义

一个线性系统能够执行它的输入数据信号与自身脉冲响应之间的褶积积分. 设  $S_1(t)$  和  $S_2(t)$  是两个时间函数, 一个是输入的数据信号, 另一个代表自身的脉冲响应, 它们之间的褶积函数定义为:

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

---

根据褶积运算建立起来的数据分析方法称为匹配过滤器法

线性系统是我们设计的一个虚拟的模拟装置，  
它的脉

冲响应函数与真实的线性系统是相同的，脉冲  
响应函数是要寻找的引力波函数的时间反演。  
从这一点上看它类似于模板。

---

## \*\*褶积运算与互相关运算的关系

两个时间函数之间的褶积运算与它们之间的互相关运算虽然不同,但有着密切的关系。褶积运算与互相关运算的区别,也是本质上的区别,在于:在褶积运算中, $s_2(t)$ 是线性系统的脉冲响应函数,它的自变量 $t$ 的符号是负的。在互相关运算中, $s_2(t)$ 是模板,它的自变量 $t$ 的符号是正的

### 褶积运算是一种实时运算

知道了这两种运算的区别与相似之处,我们可以设一个模拟装置(线性系统),用以执行实时互相关运算,它的脉冲响应函数是要寻找的引力波信号的时间反演

---

### 6.5.3 匹配过滤器操作过程

把匹配过滤器的输入端**直接**与探测器的输出端相连, 把带噪声的探测器输出信息加在过滤器的输入端, 扫描过滤器的输出. 完全像在模板方法中寻找特别大的互相关函数值一样, **寻找大的过滤器输出值**, 它就是我们感兴趣的事例.

**匹配过滤器被广泛用于寻找双中子星、中子星 - 黑洞或双黑洞旋绕并合产生的信号。 利用后牛顿展开, 可以得到近似程度很高的描述双星旋绕引力波信号的波形, 并可用来构建匹配过滤器. 实际上匹配过滤器的输出是同时对大量的不同的子模板进行计算的. 这些子模板覆盖整个期待信号的范围.**

---

实时数据处理要求计算机系统具有非常强的处理能力，在算法层面上也需要进行诸多优化以**减少实际处理中的计算量**，线性系统中脉冲响应的**“模板库”大小**也要适中。一般说来，每个数据单元的大小为脉冲响应的**“模板库”规模的**两倍，以达到实时计算的效果，

**匹配滤波器在引力波发现的过程中发挥了非常重要的作用**

---

## 6.6 $\chi^2$ 检验

在引力波探测研究中,虽然把匹配过滤器设计成在信号与模板相匹配时才给出大的响应,但是在宽频带引力波探测器中,存在着大量的、不稳定的非高斯噪声,这些不稳定噪声可能是一个瞬发即逝的“毛刺”(glitch),也可能是在很短的时间内超过宽带噪声水平的绝热变化。及环境异常现象等.不论哪种类型的非高斯噪声,都会强烈地激发匹配过滤器,当它的输出幅度超过一定的阈值时就会产生一个虚假的触发输出.这种信号的波形与模板一点也不一样,它们与模板一点也不匹配。过滤器的这种输出并非有用信号引起的,它们是假输出。

$\chi^2$  统计检验提供了一个去除虚假事例的有效手段,通过计  $\chi^2$  统计值,检验“触发”信号与期望信号波形的符合程度。

$\chi^2$  统计值对于一个真正的信号来说比较小,而对噪声导致的触发输出  $\chi^2$  有较大的值。

---

### 6.6.1 $\chi^2$ 统计检验法的基本思想

把我们使用的单个宽频带引力波探测器想象成由 $p$ 个独立的且不同的窄频带子探测器组成的一个集合体集合中的每个子探测器工作在不同的窄频带内, 这组窄频带子探测器给出 $p$ 个数据流, 对于集合中的每个子探测器对信号构建一个最佳过滤器, 然后看看这些子过滤器的响应是否一致。

比如说, 对 $p$ 个独立的子探测器选择 $p$ 个不同的基准时间 $t_0$ , 这 $p$ 个不同的基准时间 $t_0$ 分别使每个独立的子探测器的输出最大化。最后查看这 $p$ 个基准时间是否为同一个值。

$\chi^2$  值及其出现的几率是甄别的定量判据



---

## \*\*频率区间的划分

$\chi^2$  检验的具体做法是把探测器覆盖的频率范围划分成多个(如 $p$ 个)子频率间隔, 并且把总模板也分成多个子模板, 每个子模板分别覆盖对应的子频率间隔。子频率间隔及对应的 $p$  个子模板以 $j$  来标识,  $j = 1, 2, \dots, p$ , 探测器获取的数据分别用由子模板构建的匹配过滤器进行过滤, 从而得到一组过滤器输出 $z_j(t)$ , 利用 $z_j(t)$ 可以推导出进行 $\chi^2$  统计检验所需的一系列物理参量. 为了使统计结果具有经典的 $\chi^2$  分布子频率间隔的大小必须满足如下条件: 每个子频率间隔的触发输出 $z_j(t)$ 的期待值对总信号输出 $z(t)$ 的贡献是相等的 $\chi^2$

$$\langle z_j(t) \rangle = \frac{Z(t)}{p}$$

$z_j(t)$  是第 $j$ 个子频率区间输出的信/噪比,  $\frac{Z(t)}{p}$  是共同的期待值

## \*\*例-“Chirp”信号的频率区间划分

致密双星旋绕产生的“Chirp”式引力波信号的特点是随着时间的推移，频率越来越大，幅度也逐渐增加.分析鸟鸣信号时，典型的一组子频率区间为

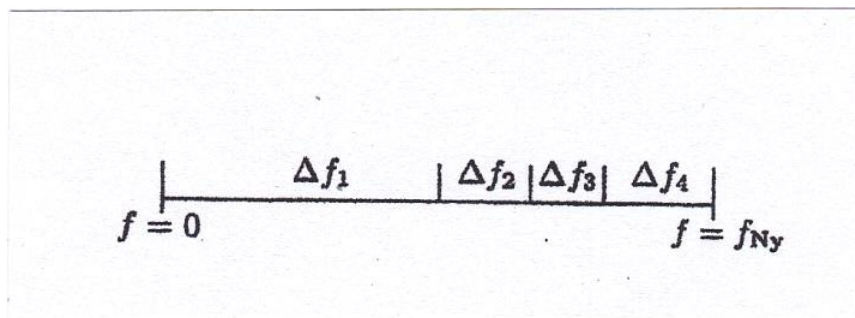


图5.6A在 $p=4$ 时鸟鸣信号子频率区间划分

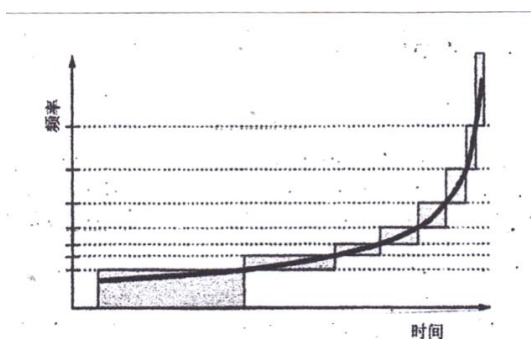


图5.6B在时间-频率平面内, Chirp信号子频率间隔的划分示意图

在探测器最灵敏的地方子频率区间最窄，而在探测器最不灵敏的地方子区间最宽。

---

### 6.6.2 $\chi^2$ 统计量的建立

我们先推导怎样建立  $\chi^2$  统计所需的物理量  $\chi^2$

#### \*\*定义

匹配过滤器输出的**信号噪声比Z**由下式给出:

$$z \equiv \int \frac{\tilde{Q}^*(f)\tilde{s}(f)}{S_n(f)} df = (\tilde{Q}, \tilde{s})$$

它是对所有频率的一个积分。在把所有频率分成很多子频率间隔后，它可以写成**从p个不同子频率间隔而来的贡献之和**:

$$z = \sum_{j=1}^p z_j \quad z_j \equiv (\tilde{Q}, \tilde{s})_j$$

**信号噪声比  $z_j$  的期待值和其平方的期待值**

$$\langle z_j \rangle = \frac{1}{p} \frac{D}{d} \quad \langle z_j^2 \rangle = \frac{1}{p} + \frac{1}{p^2} \left( \frac{D}{d} \right)^2$$

d 是探测器到波源的有效距离，D是与该信号对应的模板信号到探测器的距离

---

当信号源不存在时， $d \rightarrow \infty$

我们得到信号噪声比及其平方的期待值：

$$\langle z_j \rangle = 0 \qquad \langle z_j^2 \rangle = \frac{1}{p}$$

所有频率中测得的总信号噪声 $z$ , 摊派到各个子频率

间隔 $\Delta f_j$ 内预言的信噪比为 $\frac{z}{p}$ , 在第 $j$ 个子频率区间内

的信噪比与预言的信噪比 $\frac{z}{p}$ 之间的差 $\Delta z_j$ 为：

$$\Delta z_j \equiv z_j - \frac{z}{p}$$

这种物理量共有 $p$ 个，根据定义它们的和为零而且每子频率间隔内这个差的期待值也是零：

$$\sum_{j=1}^p \Delta z_j = 0 \quad ; \quad \langle \Delta z_j \rangle = 0$$

---

根据

$$\langle z_j z \rangle = \frac{z^2}{p}; \quad \langle z_j^2 \rangle = \frac{1}{p}$$

得到  $\Delta z_j$  平方的期待值:

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

在这里我们用了公式  $\langle z_j z \rangle = \frac{\langle z^2 \rangle}{p}$  ;  $\langle z_j^2 \rangle = \frac{1}{p}$   
根据导出的  $p$  个物理参量, 可以定义统计量  $\chi^2$  为:

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

## \*\* $\chi^2$ 时间-频率甄别器统计量的特性:

### (1) $\chi^2$ 的期待值

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

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

### (2) $\chi^2$ 的几率分布函数

在噪声既稳定又具有高斯分布的情况下,  $\chi^2$  分布函数几率是经典的自由度为  $p-1$  的  $\chi^2$  分布.  $\chi^2 < \chi_0^2$  的几率为

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

---

$\chi^2$  是不完全的伽玛函数，在噪声稳定且具有高斯分布的情况下， $\chi^2$  值期待的分布是相当窄的。它是：

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

这意味着  $\chi^2$  分布的宽度为：

$$\left( \langle (\chi^2)^2 \rangle - \langle \chi^2 \rangle^2 \right)^{\frac{1}{2}} = \sqrt{2(p-1)}$$

即我们能够在区间  $[p-1-\sqrt{2(p-1)}, p-1+\sqrt{2(p-1)}]$  内找到  $\chi^2$  的值的 possibility 有多大

---

由于该区间的相对宽度为

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

它随  $p$  的增加而减小，我们可能认为大的  $p$  值是合乎理想的，因为它好像能给出较高的甄别能力，**实践和经验表明， $p$  值并不是越大越好它的选择要适度。**因为探测器真正的噪声既**不稳定也不是高斯分布**，而且信号**并不是完美地与模板匹配**，当采用的  $p$  值非常大时会使不稳定的毛刺噪声覆盖很多子频带，削弱了它在  $\chi^2$  上的作用。



---

## **\*\* $\chi^2$ 时间-频率甄别器检验效果**

下面我们通过一个简单的例子说明  $\chi^2$  时间-频率甄别器的检验效果

该方法的基本出发点是把探测器覆盖的频率范围划分成  $p$  个子频率间隔，并建立一组相应的匹配过滤器。用它们对获取的数据进行检验，看一看这些匹配过滤器输出信号的峰值是否都在正确的时间点出现。

- (1) 用蒙特卡罗法产生“鸟鸣信号”和“噪声”两组数据，在实施  $\chi^2$  检验时，我们首先用蒙特卡罗法产生两组数据：一组是把一个模拟的鸟鸣信号（即我们期待的双星旋转发出的信号）附加在探测器的噪声上产生的；另一组数据是模拟探测器噪声中有一个瞬发即逝的“差错”，然后对它们进行  $\chi^2$  检验

## (2) 过滤器输出波形

为简单起见，我们把探测器覆盖的频率范围划分成4个子频率间隔，对每个子频率间隔都建立一个匹配过滤器，然后利用这4个匹配过滤器分别对两组数据进行过滤。分别得到四个过滤器输出“信号”  $z_1, z_2, z_3, z_4$  如下

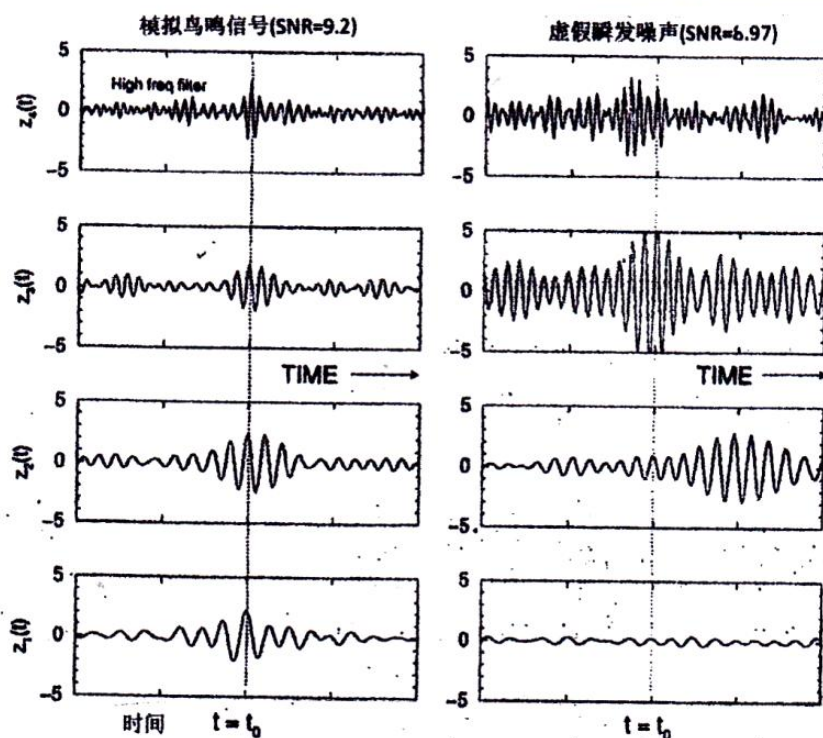


图5.7 匹配过滤器对“信号”和噪声的输出示意图

---

左边的一组图是把一个模拟的“Chirp”信号附加到探测器噪声数据流中的情况,右边一组是探测器噪声数据流中存在一个路过的爆发性信号的情况。对模拟的“鸟鸣”信号来说(左图),不同频带的匹配过滤器输出信号的峰值都出现在同一个时间偏移  $t_0$ , 在时间中的这一瞬间,所有的  $z_j$  都在同一个数值附近。

右图是注入噪声的情况,当过滤器被这个转瞬即逝的爆发性信号触发时,不同频带中过滤器输出信号的峰值出现在不同的时间点.在时间  $t_0$  处它们的值很不相同,有的很大,有的很小。

$\chi^2$  检验法统计检验法鉴别出真假信号

---

## (2) $\chi^2$ 值的比较

下面我们通过具体数值对这两种情况进行比较

### **\*\*模拟的“Chirp”信号的 $\chi^2$ 值**

从图5.7的左图可以看到，模拟的“chirp”信号的信噪比为  $z = 9.2$ ，在不同频带中匹配过滤器的输出为：

$$z_1 = 2.25 ; \quad z_2 = 2.44 ; \quad z_3 = 1.87 ; \quad z_4 = 2.64$$
$$z = z_1 + z_2 + z_3 + z_4 = 9.2$$

**$\chi^2$  统计量的值为：**

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

---

该  $\chi^2$  值出现的几率为：

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

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

$\chi^2$  期待值很小，出现的几率很大

---

## **\*\*模拟的“噪声”信号的 $\chi^2$ 值**

从图5.7的右图可以看到，转瞬即逝的噪声信号的信噪比为  $z=8.97$ ，它和“chirp”信号的信噪比差不多，在不同频带中**配过滤器的输出为：**

$$z_1 = 0.23 ; \quad z_2 = 0.84 ; \quad z_3 = 5.57 ; \quad z_4 = 2.33$$
$$z = z_1 + z_2 + z_3 + z_4 = 8.97$$

## **\*\* $\chi^2$ 统计量的值为：**

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

---

该  $\chi^2$  值出现的几率为:

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

可以看到，从信噪比的大小来看，两种情况差不多。假设我们只根据信噪比的大小来选择我们需要的“真实事例候选者”，就会把这种转瞬即逝的、爆发性的虚假信号当成“真实事例候选者”。但是，如果从  $\chi^2$  检验结果来考虑，我们就会发现，两种情况下的  $\chi^2$  值及它们可能出现的几率有天壤之别。对于附加到高斯噪声中的Chirp信号，出现这个  $\chi^2$  期待值的几率是 **73%**，而对于欺骗性的噪声信号来说，出现这个  $\chi^2$  期待值的几率是极其微小的（只有  $9.4 \times 10^{-15}$ ），通过这两种情况的比较，我们可以清楚地看到  $\chi^2$  检验的巨大威力。

---

谢谢大家