

## 8.2 文献译文——处理绿岸望远镜 (GBT) 的数据

**介绍** 你们在脉冲星搜寻合作实验室 (Pulsar Search Collaboratory) 的工作将是分析折叠图并确定它们是否显示了新脉冲星的证据。搜寻和识别脉冲是教你如何做到这一点的指南。然而，了解折叠图如何产生，以及如何从望远镜收集的数据中绘制出折叠图也很重要，这也是本指南的目的。有了这些信息，你就能更好地理解折叠图的信息。很多东西在“黑匣子”中进行着，你可能没有意识到，在寻找新的脉冲星时，每一步都是至关重要的。在本指南中，我将重新介绍 *Searching for and Identifying Pulsars*<sup>14</sup> 中涉及的思想，所以最好先读一下。但是，为了清楚起见，我将重申并扩展该指南中的必要概念。这个想法不是要记住数据分析的每一个细节，而是要对所做的事情有一个更广义的认识。拥有全局的轮廓会让你成为一个更好的天文学家。

寻找新的脉冲星有八个主要步骤：

1. 使用 GBT 进行观测并记录原始数据
2. 把这些原始数据转换成更有用的格式，从而得到通过望远镜收集到的功率
3. 在数据中搜寻 RFI，并确保在进一步分析时忽略它
4. 对数据进行消色散并将其转换为时间序列
5. 对时间序列进行傅里叶变换
6. 再次查找并删除 RFI
7. 搜索候选脉冲星的傅里叶谱
8. 折叠这些候选体，然后创建折叠图

让我逐一检查每一步，理解正在发生的事情。

**第一步：进行观测** 回想一下，无线电波是光的一种形式，就像你用眼睛看到的可见光一样，尽管我们不习惯这样想象无线电波，但它们就像可见光一样反射和聚焦。大多数现在可见光望远镜（更通常被称为光学望远镜）都是通过镜子来实现的。GBT 使用了一种“射电镜子”——GBT 天线的白色表面反射并聚焦无线点波。这个表面的形状是抛物面（想想  $y = x^2$ ），因为任何平行于抛物面轴线的光线（图上的 y 轴）在反射时被聚焦。在光学望远镜中，通过 CCD（于数码相机中使用的设备相同）来检测光线。无线电波是用一种不同的技术来检测的。这些电波集中在一个接收器上，在那里会产生一个微弱的电信号。对振荡信号产生的原始电压进行处理，以放大它并使用电子电路将其转化到更方便的频率。

计算机不能处理电压，因为电压是一种模拟信号。相反，计算机处理数字信号。所以望远镜的原始电压将会被数字化。这个过程将收集到功率记录为离散值，这意味着我们可以用一个 1 或者一个 2 来记录功率，而不是一个 1.5。你可以把数字化想象成一个平缓的斜坡变成一组台阶。在山上，你可以站在任何你想站的地方，所以可以站在任何高度。但在台阶上，你不能漂浮在两个台阶之间，所以你只能站在特定的高度上。

如果我们使用 8-bit 的数字化，那么我们可以记录  $2^8 = 256$  个值（类似于有 256 个台阶），16-bit 的数字化可以得到  $2^{16} = 65536$  个值（类似于有 65536 个台阶）。1-bit 数字化得到  $2^1 = 2$  个值。这就像说你不是站在山顶，就是站在山脚下，没有处于中间的值。

这种数字化的信号实际上是作为原始的望远镜数据储存的。对于我们所关注的脉冲星观测，原始数据是通过一种叫做 Spigot 的专用脉冲星仪器获得的。Spigot 可以产生大量数据——一分钟的观测就可以产生将近 1.5GB

---

<sup>14</sup>[http://pulsarsearchcollaboratory.com/wp-content/uploads/2016/01/PSC\\_search\\_guide.pdf](http://pulsarsearchcollaboratory.com/wp-content/uploads/2016/01/PSC_search_guide.pdf)

的文件。一个小时的观测将会填满许多人家用电脑的硬盘。下一步是将原始数据转换为一个更有用的格式。

**第二步：对原始数据进行 filterbank** 脉冲星天文学家感兴趣的是望远镜收集了多少功率。但是我们想知道功率是什么时候、什么频率被收集的。原始数据包含这些信息，但其格式并不那么有用。因此，数据通过一个称为 filterbank 的过程转换为另一种格式。通过 GBT 和 Spigot 收集的数据使用计算机程序进行 filterbank。最终的数据被称为 filterbank 文件，或者只叫做 fil 文件。用于创建折叠图的数据包含 GBT 以 0.00008192 秒的时间间隔、0.02MHz 的频率通道收集的功率。图42是 filterbank 文件的图形表示。仔细观察它可以帮助你更好地理解数据的形式。一旦数据进行 filterbank，我们就可以进行下一步，从我们的分析中找到并删除 RFI。

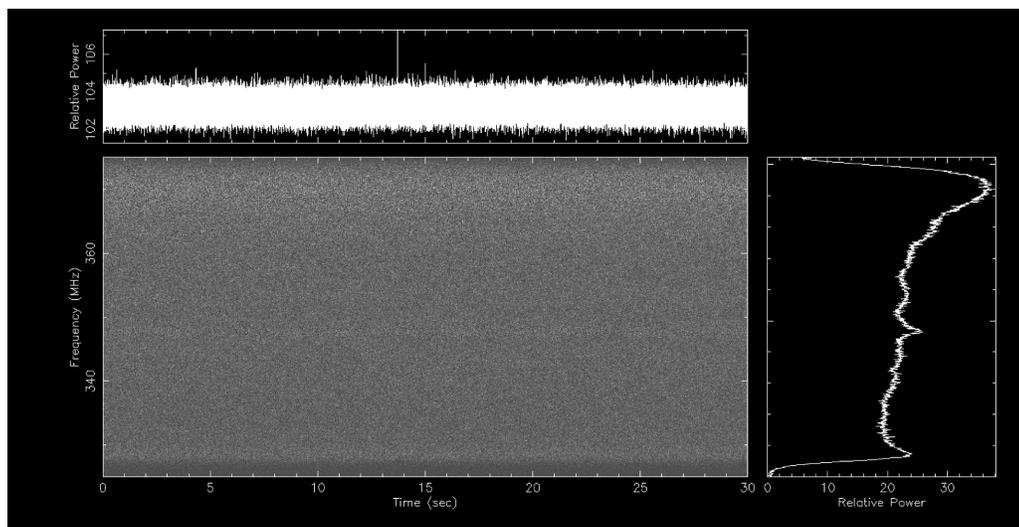


图 42: filterbank 文件的图形表示形式。中间的灰度图像是 filterbank 文件本身，深灰色代表低功率，白色代表高功率。从观测开始的时候在 x 轴上，观测频率在 y 轴上。每个灰度点表示在 0.00008192 秒的观测时间和在 0.05MHz 宽的频率通道上收集的功率。顶部的子图是时间序列，显示了收集的功率作为时间的函数。右边的子图被称为带通，它代表了作为频率函数的收集功率。时间序列和带通是分别沿着 y 轴或者 x 轴的灰度图相加得到的。换句话说，时间序列是通过将所有频率的功率相加得到的，而带通是通过将所有时间的功率相加得到的。仔细看这个灰度图，你有没有看到任何明显的横向特征？你能把这些和带通的特征匹配起来吗？根据时间序列的外观，你希望看到一些特征在灰度中垂直穿过吗？为什么或者为什么不？

**第三步：发现 RFI** 射频干扰 (RFI) 对射电天文学家来说是个大问题。就像光学天文学家无法观测到有太多光污染的地方一样，射电天文学家需要去射频天空“黑暗”的地方。因为 GBT 在国家无线电静默区，通常只有很少的 RFI。但要拜托所有的 RFI 来源是不可能的，所以我们搜索脉冲星数据之前，要尽可能多地识别出 RFI。在 *Searching for and Identifying Pulsars* 中，我们讨论了如何在折叠图中识别射频干扰，但用这种方法找出所有的射频干扰来源太费时了。取而代之的是，计算机程序尝试通过查看数据统计来自动识别 RFI。

第一步是计算数据的平均功率<sup>15</sup>。我们在假设色散量为 0 的情况下进行消色散，然后把所有单独频率通道加起来。我们将在下一节讨论消色散，但现在你需要知道的是，这确保了我们在数据中看到的任何信号都来自地球，而不是太空。消色散的结果是作为时间函数的功率值。一旦计算出平均值，我们将寻找任何功率值明显高于平均值的时间间隔。如果出现了这种间隔，我们将这些时间间隔在接下来的分析中删除。这被称为裁剪 (clipping)，因为我们基本上剪去了任何高功率的间隔。

<sup>15</sup>对于那些熟悉统计的人来说，我们计算的是中位数而不是平均值，因为中位数不会被 RFI 扭曲太多。

在之前的步骤中，我们查看了整个观测过程中数据的平均值，并将所有频率的功率相加。在接下来的步骤中，我们仅计算在时间和频率上数个小间隔数据的平均值。这个步骤最好用一个例子来说明。假设我们计算所有频率通道在观测的前两秒的平均功率。然后，我们将这个平均值与前两秒内的每个频率通道的实际功率进行比较。如果一个通道的功率比平均值高得多，这个通道被从分析中移除。

接下来，我们得到了在整个观测过程中单个频率通道的平均值。这可能是对应于 350MHz 频率的通道。然后我们在这个频率下每隔两秒观察一次。如果任何区间内的功率显著高于平均值，我们将其从分析中移除。实际上类似的技术被使用，但是我们没有计算平均功率将其作为识别 RFI 的标准，而是使用数据的标准差。但总的技术是非常相似的。

还有最后一步要做。如果单个时间间隔或者频率通道被 RFI 严重污染，我们将从分析中完全去除它，即使我们的程序没有在每个频率通道或者每个时间间隔中找到 RFI。通常情况下，我们会删除污染超过 70% 的地方。原因是这些间隔和通道可能是无用的，即使一些 RFI 只是有点太弱而无法识别。

当以上所说的都做完后，我们会得到一个关于 RFI 的 mask 文件。这个文件被其他程序用来识别哪些时间间隔和哪些频率通道可以从分析中被忽略。43 是这样一个 mask 文件的图形表示。

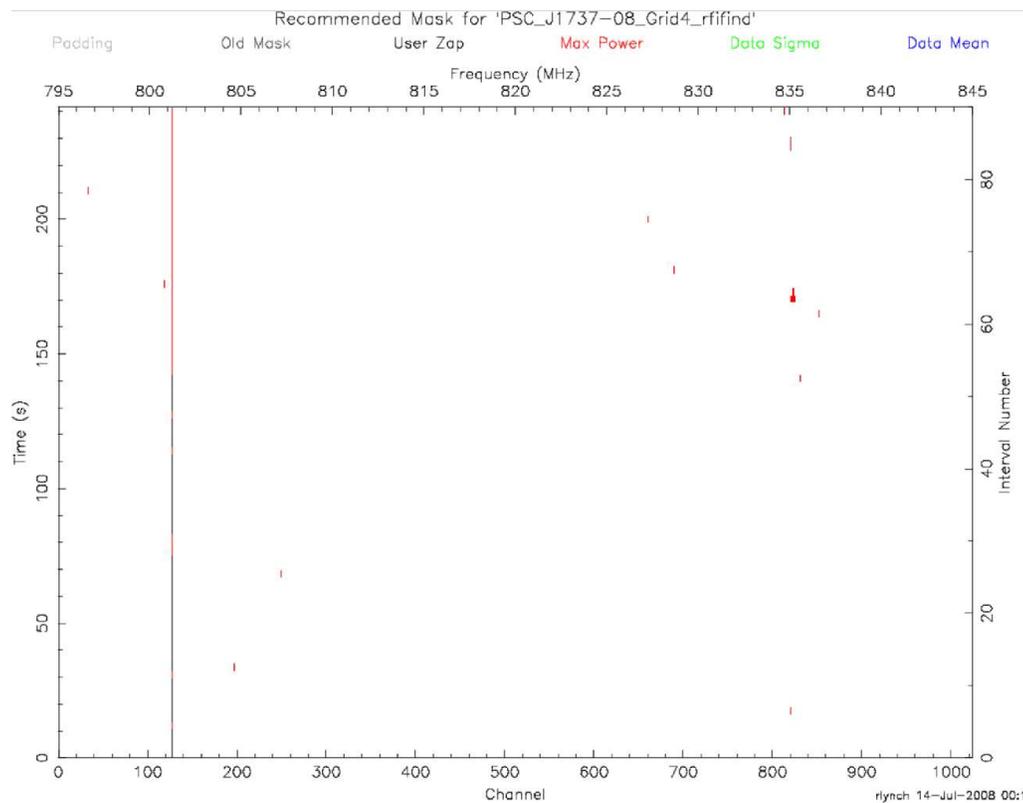


图 43: RFI 的 mask 文件的图形表示。观测频率在 x 轴上，自观测开始以来的时间在 y 轴上。我们通常将观察结果分成大约 2 秒长的块，每个间隔都有一个数字，可以在右边看到。数据中的黑色、红色、绿色或者蓝色部分已被屏蔽。颜色代表了导致数据被屏蔽的特定问题，黑色表示用户指定了数据被屏蔽，或者某个特定的间隔或者频率通道被污染得太严重。红色表示数据中的某部分总功率过高，绿色表示数据标准差异常，蓝色表示该数据的平均值异常高。

接下来的步骤便是对这些数据进行消色散。

**第四步：消色散** 为了找到新的脉冲星，我们必须在数据中寻找周期性信号。发现周期信号的关键是它们看起来很短暂。如果脉冲星在其 100% 的旋转期间向我们的方向发射无线电波，它就不会真的像脉冲。取而代之的是，它看起来就像一个次序不断的能量流，就像一个开着的灯泡。由于脉冲星的信号非常弱，这看起来不会很有趣——它只会表现为无线电波背景水平的轻微增加。幸运的是，脉冲星的脉冲只在它们旋转周期的一部分可见（通常是百分之几）。但是有一些影响可以导致脉冲持续时间变长。如果这些影响是严重的，窄脉冲可能会扩散太多而无法检测到。脉冲增宽的主要原因叫做色散，其效应叫做色散拖尾效应 (dispersion smearing)

色散拖尾效应是由空间中的电子引起的。虽然太空几乎是空的，但它并不是完全空的。我们在太空中发现的东西之一是电子，它们不受任何原子的束缚。这些所谓的自由电子会导致无线电波减速。无线电波的衰减程度取决于它们的频率，低频比高频慢得多。由于我们观测脉冲星的频率范围很宽，通常 10-100MHz，所有我们同时从脉冲星上收集高频和低频的信号。色散导致低频脉冲星信号到达地球落后于高频脉冲星。如果我们天真地把所有频率的功率加起来，脉冲就会扩散，可能扩散的太大而无法检测到。

幸运的是，我们知道如何消除色散的影响，如果我们知道色散量 (dispersion measure, DM)。色散量是一种描述有多少电子位于脉冲星和地球之间的方法。电子越多，DM 越高，因为色散拖尾效应越严重。因此，DM 为 0 告诉我们，信号在达到地球的途中没有经过任何电子。但对于真实的脉冲星来说，这是不可能的。因为脉冲星距离我们由许多光年之远，而且必须与一些电子发生相遇。唯一的解释是，信号本身来自地球；也就是说，这些信号实际上是 RFI。这很有用，因为我们看到的任何 DM 为零的信号都肯定是 RFI(见上述去除 RFI 的解释)。

去除色散拖尾效应的过程称为消色散。

但现在我们遇到了一个难题。如果不限探测脉冲星，我们绝对不可能知道脉冲星的色散量是多少，除非我们在合适的 DM 处消色散，否则我们无法探测脉冲星。我们唯一的选择是在尽可能多的合理的色散量上消色散。也就是说，我们只是猜测，看看是由有脉冲星出现。

消色散的实际过程包括在给定频率上对信号实际时移。时移量很容易计算：

$$\Delta t = 4.15 \times 10^6 \text{ms} \times \left( \frac{1}{f_{ref}^2} - \frac{1}{f_{chan}^2} \right) \times DM \quad (34)$$

此处  $f_{ref}$  是某个参考频率 (例如我们观察的最高频率)， $f_{chan}$  是我们要消色散的信道的频率，而 DM 是色散量。频率以 MHz 为单位，DM 的单位是每立方米秒差距 ( $\text{pccm}^{-3}$ )。秒差距是一种距离单位，大约等于 3 光年。让我们看一个例子来说明如何使用这个等式。假设我们在高频为 375MHz、低频为 325MHz 观测脉冲星。另外，让我们假设我们知道色散量是  $20 \text{ pccm}^{-3}$ 。为了确保任何脉冲同时达到，我们必须移动在最低信道的信号：

$$\Delta t = 4.15 \times 10^6 \text{ms} \times \left( \frac{1}{375^2} - \frac{1}{325^2} \right) \times 20 = -195 \text{ms} \quad (35)$$

如果我们不去消色散，那么拖尾效应的干扰将使得我们不可能探测到我们最感兴趣的毫秒脉冲星 (MSP)，它的周期在几到几十毫秒之间。

一旦我们以试验的 DM 对数据进行了消色散，我们就把所有频率信道的功率加起来。留给我们的是望远镜收集的功率随时间的函数。这就是所谓的时间序列。

最后一件事——即使使用这种技术，也不可能完全消除色散拖尾效应的影响。这是因为在一个频率通道内仍然会有拖尾效应。有一些方法可以完全消除拖尾效应的影响，但它们更复杂，我们现在不需要考虑。幸运的

是，由于单个通道的带宽相当小。色散拖尾效应的影响比上面的例子要小得多。但对于非常高的色散量，这在试图检测非常快的脉冲星时仍然会造成问题。

现在我们已经创建了时间序列，下一步是对数据进行傅里叶变换。

**第五步：快速傅里叶变换** 正如在 *Searching for and Identifying Pulsars* 中所述的，在时间序列中，单个脉冲几乎总是太弱而无法探测到。我们需要在数据中找到非常微弱的周期信号。微积分中有一个数学运算非常适合这个任务。它叫做傅里叶变换，它的应用叫做傅里叶分析。傅里叶分析是一个广泛的、有时很复杂的领域，所以这里我么不讨论任何细节。你应该知道的是傅里叶变换改变了时间序列，所以它代表了在某个傅里叶频率下收集到的功率。这个傅里叶频率不同于我们观察到的无线电波的频率。傅里叶频率描述了在数据集中某种信号重复出现的频率。如果脉冲星的周期信号存在于我们的数据中，那么在傅里叶频率为  $1/P$  时将会有大量的频率，其中  $P$  是脉冲星的自旋周期。

计算一组数据的傅里叶变换是计算中的一个众所周知的问题，因为傅里叶变换对许多领域的研究都非常重要。最常用的方法被称为“快速傅里叶变换 (FFT)”。出于这个原因，我们经常说我们正在对时间序列进行 FFT，或者从它创建一个 FFT。一旦我们从时间序列中创建了 FFT，我们通常会再次寻找 RFI。

图44是傅里叶频谱的图形表示。你在图中看到的峰实际上在你将要搜寻数据中的一个真正的毫秒脉冲星。这东西真的有用！

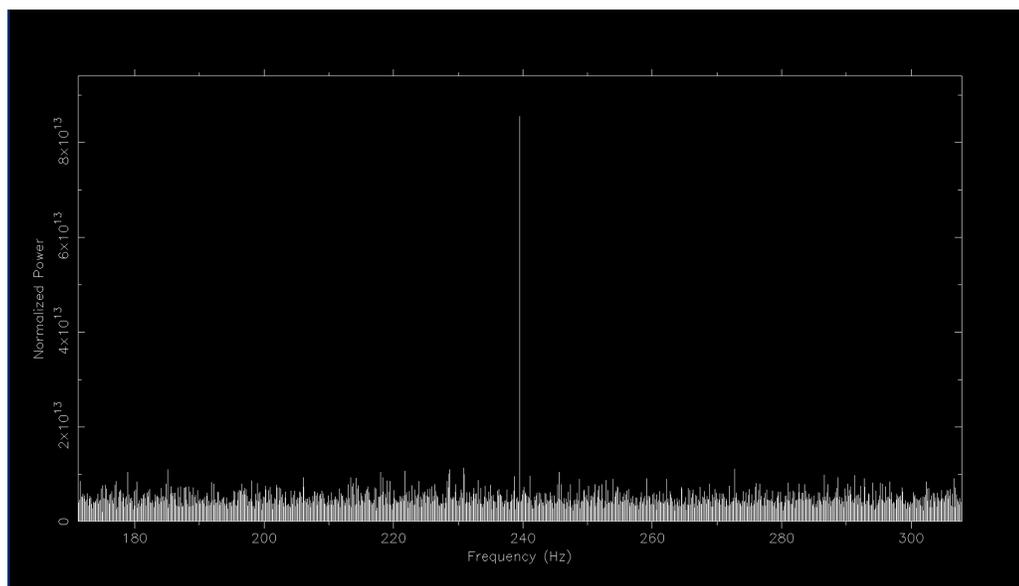


图 44: 这是脉冲星搜索合作实验室 GBT 数据的 FFT。x 轴是傅里叶频率，y 轴是功率。你看到的尖峰是一颗脉冲星。从这张图中，你能估算出这颗脉冲星的周期吗？

**第六步：再次识别 RFI** 虽然我们由一些程序可以识别大多 RFI 的来源，但有些不可避免地会漏过。处于这个原因，我们通常创建一个时间序列的 FFT，它的色散量为 0。回想一下，任何 DM 为 0 的信号一定是人造 RFI。然后我们可以检查这个 FFT，寻找强的信号。我们看到的一个非常常见的信号是傅里叶频率为 60Hz 的信号。这相当于某个无线电波源每秒重复 60 次。熟悉电子产品的人应该知道，这正是我们的家庭和企业供电的交流电的频率。因此，在 60Hz 的 RFI 是由输电线路产生的。当我们在 FFT 中看到这类信号时，我们可以在进

一步分析中明确地忽略它。这样我们就消除了更多潜在的射频干扰来源。

现在我们已经尽了最大的努力去除了射频干扰，我们已经搜寻好寻找脉冲星的数据了。

**第七步：搜寻傅里叶频率** 理论上，搜寻脉冲星是一项简单的任务。一旦我们以一些试验 DM 对我们的数据消色散并进行 FFT，我们只需要在傅里叶频率中搜索具有较强功率的信号。如果我们发现了这样的信号，我们就把它的信息记录在一个候选脉冲星列表中。当然，由于我们不能提前知道脉冲星的 DM，我们必须尝试很多不同的消色散方法。此外，我们的 FFT 通常非常大，包含数百万个需要检测的单个傅里叶频率。尽管如此，现代计算机在搜寻脉冲星方面还是很有效率的。所需要的只是时间。

不过，还有一个影响需要考虑。我们的描述到目前为止，傅里叶分析已经做出了一个不言而喻的假设，及我们试图探测的脉冲星的周期在整个观测过程中是恒定的。想象一下，有什么东西导致脉冲星改变周期。在观测的开始，它的周期可能是 2.17436ms，在观测结束时，它的周期是 2.17440ms。在此期间，这两个值之间的周期变换缓慢。当我们取时间序列的 FFT 时，功率将在  $1/(2.17436\text{ms})=459.90\text{Hz}$  和  $1/(2.17440\text{ms})=450.89$  这两个频率之间展开。当然，这是一个很小的改变，但足以使得难以探测到这样的脉冲星。原因是一个完美的周期信号，具有一个恒定的周期，通常它的功率分布在千分之几赫兹，通常是更小（我们可以解析的傅里叶频率的最小分量是  $1/T$ ，其中  $T$  是总观测时间）。即使是 0.01Hz 的位移也足以将能量分散开来，以至于很难探测到脉冲星，除非它非常明亮。这就像把一堆非常陡峭的铺在更大的区域上。很快，你甚至不知道那里有一堆东西。

那么究竟是什么导致脉冲星的周期发生变化呢？所有脉冲星的自转速度都变慢了，因为脉冲星正在失去能量。然而，这个效应太小，对于探测脉冲星来说并不重要。当脉冲星相对于观测它的望远镜加速时，还会产生另一个更重要的效应。这种加速度会导致多普勒频移，这与救护车向你驶来时的高音和驶离时的低音是一样的。假设一颗脉冲星正在加速远离我们的望远镜。它会发出脉冲，一定是在太空中传播。然而，当下一个脉冲发出时，脉冲星离我们更远了。脉冲必须走得更远一点才能到达地球，而它所需要的额外时间使它看起来比实际周期更长。如果脉冲星在加速，那么相邻到达脉冲之间的时间似乎越来越长，周期似乎变慢。相反，如果脉冲星朝着地球加速，周期看起来则加快了。

因为地球绕着它的轴旋转，也绕着太阳旋转，所以望远镜和脉冲星之间总有一定的加速度。我们可以很容易地消除这种影响，因为我们可以精确地测量地球的运动。然而，如果脉冲星围绕另一颗恒星运动，那么它也会加速。其他效应也可能引起加速。而最有趣的毫秒脉冲星也是最有可能有伴星的脉冲星，因此会加速。所以找到探测这些加速脉冲星的方法非常重要。当然，我们提前不知道加速度是多少。就像我们在对数据进行消色散时必须猜测 DM 一样，我们也必须猜测脉冲星正在经历多大的加速度。一旦我们做出猜测，我们可以使用详细的傅里叶分析来消除加速度的影响。但是我们必须仔细研究另外一个参数，这也是寻找脉冲星需要大量计算机能力的另一个原因。

然而，当所有的都说完了，我们将在几个色散量和几个加速度量上有一些脉冲星候选体。下一步是检测候选体，并确定哪些是真正的脉冲星。

**第八步：人工操作** 到目前为止我们所讨论的一切都可以自动化。计算机只需要知道要尝试的色散量和加速度，其余的就交给它了。然而，最后人们必须决定一个候选脉冲星是否值得进一步研究和确认。一旦一群候选体被收集，它们就会被折叠 (fold)。折叠在 *searching for and Identifying Pulsars* 中详细讨论过，所以我们在这里只是进行复习。一个名为 `prefold` 的命令用于折叠过程。你可以把折叠看作是把我们 `filterbank` 文件分成几部

分，每一部分都和脉冲星的周期一样长。我们对每个频率通道都这样做。然后把这些小部分折叠在一起。如果那里有一个周期信号，当我们把更多的观测片段加在一起，它会变得更强，直到最后出现一个明显的脉冲轮廓。图 4 阐述了这种技术。prefold 生成你要看的图，还可以微调候选体脉冲星的周期、加速度和色散量。

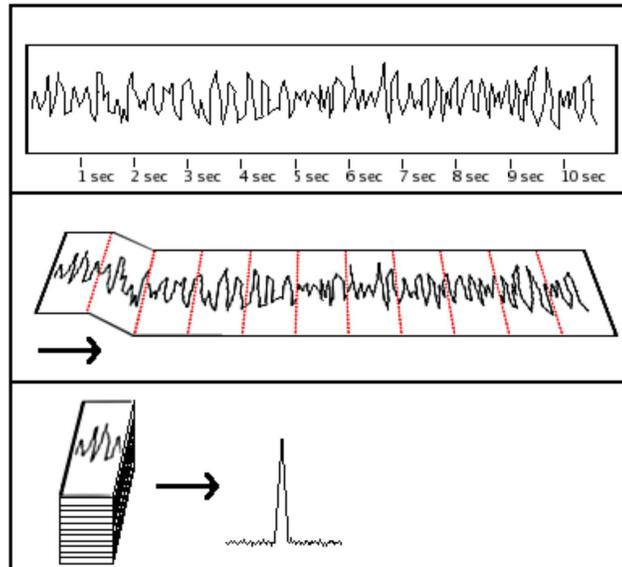


图 45: 这里我们演示折叠技术。在最上面的子图中，我们在数据的每一秒处做标记。在中间子图上，我们在这些标记处折叠。在底部的子图中，我们把折叠的每一层的信号加起来。通过这种方法，我们可以找到一个微弱的脉冲星，否则它就会消失在噪声中。然而，我们必须提前知道周期。在这个例子中，我们用的周期是 1 秒。

这就是你的切入点。你是一个天文学家，必须决定一个候选者是否是真正的脉冲星。

**结论** 我希望这篇指南能对天文学家如何寻找脉冲星有所帮助。这个过程很复杂，我已经尽力简化了一些最困难的概念。在你事业的这个阶段，了解寻找脉冲星的每个方面并不重要。专业的天文学家花费了很多年的时间来学习他们需要实现的、我们上述讨论的所有步骤。脉冲星搜索合作实验室的目的就是让你了解其中一些想法。希望你的经历将是你一生漫长旅程的开始，让你更深入地了解天文学和科学。

祝你愉快!