

科研实践报告

摘要

了解有关 PandaX 实验项目的主要内容，参与 PandaX 试验项目的数据处理的小组。了解数据库的处理，学习比较容易上手的 MongoDB。学习并熟悉服务器上的基本操作，学习 Tmux 的使用的方法，并利用它来参看已经完成的程序，并继续完成未完成的数据。

关键字：数据库; 粒子物理; MongoDB; PandaX; Tmux;

背景:

PandaX, 全称 Particle and Astrophysical Xenon Detector, 主要是利用氙来进行暗物质的探测。不同的 PandaX 的项目所用的氙的类型可能不同, 可能是利用液态的氙或者是气态的氙。

我们在锦屏山中心建造了研究暗物质的实验室, 因为地球无时无刻都不受到宇宙射线的影响, 所以会对检测暗物质产生很大的影响, 为了避免这种影响, 我们要将研究暗物质的实验室建立在尽可能不受宇宙射线的影响的地方, 所以我们将实验室的地址选择在锦屏山的实验室当中。

同时, 自然界中存在天然的放射性, 几乎任何天然存在的物质都是具有放射性的, 所以除了将实验室建立在可以屏蔽宇宙辐射的地下实验室, 我们还要将整个实验室用特殊的物质屏蔽起来, 一般是这些物质都是经过高度提纯的, 经过提纯的物质中包含的天然的放射性物质的含量相对天然物质来说大大降低, 因此辐射的强度也大大降低。

第一期 PandaX 中检测信号的仪器是一个装满氙气的容器。即使我们有了上述的屏蔽的措施之后, 我们还是很有可能收到天然的辐射的影响, 所以为了避免将这种天然的辐射误认为我们要检测的暗物质的信号, 我们将容器中的靠近边缘的信号忽略掉, 只考虑靠近容器中心的信号。

暗物质和氙产生的反应有两种, 其中一种是产生电子的反应, 反应产生电子之后由于容器中的电压会在容器中进行匀速的漂移, 当漂移到容器的上部分的时候, 会

反应产生光子，产生的光子会被容器最上的探测器接收到信号。容器如图 1 所示。

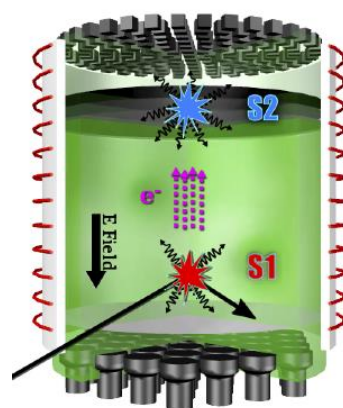


Figure 1

对于一个信号，我们一般会记录这个信号前后各 $500\mu\text{s}$ 的长度的记录。而且我们记录的信号需要是数字化的信号，所以又要将模拟的信号转化为数字化的信号，因此信号的数据量和数字化的精细化程度有关。通过计算可以大概估计出我们这个试验项目的数据量是非常大的，所以为了处理如此大的数据量，我们需要很强的服务器的性能和数据库的管理的能力。

不过对于这种情况的数据采集，我们可以有另外一种更加有效的方法来处理：我们可以先用不同的计算机来接受多束光电倍增管的信号，之后先将这些信号进行处理，剔除其中无效的部分，筛选出其中有效的部分。因为信号的特殊性质，所以有些信号会是错误的信号，此时信号的深度会超出正常的信号范围。之后我们将处理过后的信号一起储存在一个数据库里，然后我们用主计算机来从数据库里取得有效的信息。通过这样的处理方法，我们可以大大减少数据量的大小，同时我们可以通过多层的数据的储存和读取来优化数据的有效性。

工作内容：

去上海市粒子物理与宇宙学重点实验室参观 PandaX 的试验项目有关的内容

学习并运用 Tmux 来进行 Linux 上的命令行的操作，Tmux 相比传统的命令行有一些突出的优点，特别是在服务器上 Tmux 具有传统命令行不可替代的优势。

Tmux 具有三级分层的形式，从上至下分别是会话(Session)，窗口(Window)，窗格(Pane)。

对于会话，我们可以随时连接和短连，这在服务器上就很有效果，因为通常情况下我们会在服务器上运行一些命令，这些命令运行的时间可能会很久，通常情况下我们只能等待这些命令运行完成而不能做其他的命令的执行。但是利用 Tmux 的会话的功能，我们可以随时监控这些命令的执行情况并随时离开且不影响命令的执行的情况。

我们可以利用 new 命令来新建一个会话，并利用 attach 和 detach 来随时地对会话进行连接和断连。我们在连接到一个会话之后可以像通常的那样执行各种命令行的命令，而且即使我们利用 detach 来断开连接也不会影响到这些会话中的命令的执行的

情况。

窗口是会话下的分支，在一个会话中我们可以执行命令，但是执行命令我们需要基于一个窗口，在一个会话中我们可以创建不同的窗口，利用命令 `C-b c` 创建一个新的窗口，我们可以对不同的窗口进行命名，并在不同的窗口执行不同的命令。

窗格是窗口下面的分支，我们可以利用命令 `C-b %` 和 `C-b "` 来进行窗口的窗格的分割。在窗口中新建不同的窗格可以让我们在同一个屏幕内同时监控多个画面，而且每个窗格中我们可以同时执行不同的命令，利用这样的功能我们就可以做到类似下面的事情，比如查看文档的内容的同时执行我们需要的命令，并且同时监控 CPU、内存等资源的利用的情况¹。

熟悉有关 Linux 上有关服务器的操作命令，比如 `ssh`, `scp`，通过操作服务器上的数据，我们可以通过一些特定的命令来对服务器的性能，对特定的程序的性能进行监控和改善。此次的项目中，我们要将不同的数据通过不同的光电倍增管来传输到不同的计算机当中，又要将数据储存在数据库当中，并且要从数据库中读取数据。

但是这只是我们的构造的方案，并没有运用到实际的测试当中，所以我们要编写相应的程序来测试整个方案的可行性。方案的可行性主要限制于数据库的性能和各个计算机的处理能力。

我们通程序的设计，要模拟上述的过程来检验性能。通过程序将一定大小的数据进行切片并写入数据库，并进行读取性能的测试。

困难:

由于我们是通程序模拟整个程序的过程，所以并不清楚程序模拟的过程和实际的过程有没有什么冲突，是不是准确模拟了实验中的真正的过程。

由于对实验的真正过程的准确模拟的困难，所以刚开始模拟的时候测试的过程中出现了错误，性能并不理想，而且是离谱的糟糕。经过考虑应该是程序设计的问题，所以目前正在重写程序

后续工作计划:

继续参加 PandaX 项目的计划，对完成的数据进行阅读并学习，并且完成还没有完成的数据。

对数据库的原理进行学习，了解数据库的运行的原理，并学习如何对数据库中的数据进行处理。

对 MongoDB 进行深入的了解，并利用 MongoDB 进行数据的处理。

参考文献：

1. Nicholas Marriott. Tmux Manual Page. April 5, 2017. Retrieved at April 9, 2017, from <http://man.openbsd.org/OpenBSD-current/man1/tmux.1>