

# fasthi简要使用说明

`fasthi` 是用于分析FAST中性氢漂移扫描巡天数据的python程序，其主要功能为：

- 通过给定fits文件信息，寻找fitscube中的SDSS源
- 通过基线扣除的方法得到源的峰值流量、信噪比等信息
- 通过积分流量的方法得到源的积分流量、等值宽度、信噪比等信息
- 绘图查看源的射电信号、光学图像等
- 分析结果形成catalog以供使用

程序在有条件的计算机上支持多核并行运算。

程序部分功能（FFT降噪等）有待完善。

## 安装&使用

由于程序需依赖于 `multiprocessing` 库的共享内存功能，只支持python3.8及以上版本。

程序开发使用的环境见 `fasthi.yaml`，推荐使用 `conda` 直接导入该环境使用：

```
1 | conda env create -f fasthi.yaml
```

使用虚拟环境运行程序可以有效解决依赖冲突的问题。有关Anaconda3的安装，详见[install anaconda3.md](#)或其[PDF版本](#)。

程序在Linux系统上测试通过，Windows/Mac下理论上同样可以使用，但未经测试。运行程序前需要激活虚拟环境：

```
1 | conda activate fasthi
```

如需退出该虚拟环境，执行：

```
1 | conda deactivate
```

运行程序前先对配置文档 `config.yaml` 进行配置，详见后文配置章节。

注：大部分参数与之前版本相对应，但新增/删改了一些，请依照说明仔细确认，必要时搜索源码中对应的关键词以确认。文件中参数的默认值在作者调试程序的过程中结果符合预期。

运行程序请参考 `demo.py`，可以根据自身需求修改后使用，也可直接运行（分析配置文件中指定目录下所有符合条件的fits文件）：

```
1 | python demo.py
```

## 配置

程序会从配置文件 `config.yaml` 中读取运行需要的参数，下面依次说明。配置文件中亦有对每个参数的注释，可供参考。如有疑问，请在源代码中搜索对应的键（关键词）查看具体使用。

## DIR

与文件夹相关的参数

- `work_dir`:  
指定工作文件夹，之后程序会在该文件夹中创建一系列文件/文件夹
- `fits_dir`:  
指定存放要分析的fits文件的文件夹，当文件夹中的fits文件后缀与配置文件中的 `FITS.back_str` 一致时，程序将该文件纳入工作队列中，以待分析
- `lib_dir`:  
程序依赖的文件的存放路径，通常不需要更改

## RUN

与程序运行相关的参数

- `n_fits`:  
程序最多运行的fits文件数量
- `n_cores`:  
程序多核运行使用的核心数量，使用4核心则赋值4，单核运行则赋值1，使用全部可用核心的指定比例则赋值0并调整 `RUN.use_ratio` 参数。具体逻辑见配置文件中的注释
- `use_ratio`:  
当 `RUN.n_cores` 为0时，使用全部核心中 `RUN.use_ratio` 百分比的核心进行并行运算。取值0-100

## ANALYSE

程序中数据分析用到的参数，请根据实际需要调整

- `bsl_inner/bsl_outer`:  
基线采样的范围，单位为角分
- `bsl_hw/bsl_hh`:  
绘图时展示基线的窗口大小（半宽和半高），单位为角分
- `bsl_vn`:  
可信的最小基线采样数目，若源可选择的基线采样数目小于这个数，跳过这个源
- `bsl_save`:  
是否保存一张独立的基线图。该功能目前未作实现
- `bsl_ma`:  
对基线进行滑动平均的窗口大小，单位为像素。目前未作实现
- `bsl_rm_nan`:  
是否剔除基线采样中存在nan值的样本，True/False。若为True，基线采样中包含nan的样本不会参与到基线的计算中；若为False，直接使用 `np.nanmedian` 得到基线
- `opt_range`:  
绘图时光学图像窗口的大小（半高和半宽），单位为角分
- `opt_save`:  
是否保存一张独立的光学图像。目前未作实现
- `w50`:  
对信号的先验w50，用于确定分析信号的速度范围，单位km/s
- `view_scale`:  
配合w50确定分析信号的速度范围。实际分析信号的窗口大小为  $2 * view\_scale * w50$
- `int_vel`:  
绘制信号时的积分尺度，单位km/s
- `fft_run`:  
是否用fft降噪。目前未实现，请设为False
- `fft_range`:  
目前未实现，请忽略

- **reduce\_ifma:**  
是否在扣除基线前对信号使用滑动平均, True/False
- **reduce\_ma:**  
在扣除基线前使用滑动平均的窗口大小
- **measure\_ma:**  
在使用第一种方法测量参数前, 对扣除基线后的信号使用滑动平均的窗口大小。关于两种测量方法见 FAQ
- **intm\_range:**  
在使用第二种方法测量参数时, 拟合使用的最大速度范围, 单位为km/s
- **intm\_valid:**  
在使用第二种方法测量参数时, 拟合使用的速度范围的最小值, 单位为km/s。小于该值认为结果不可信, 返回的catalog中相关参数设为-1
- **intm\_ap:**  
在使用第二种方法测量参数时, 如果采用类似孔径测光的方式测流量, 其孔径的大小, 单位为角分。程序目前使用给定RA、Dec下对应像素的值(程序中的 `sigcent`)作为流量计算的来源, 如有需要请修改源代码
- **intn\_in/intn\_out:**  
在使用第二种方法测量参数时, 计算noise的圆环的内径和外径, 单位为角分。目前程序未使用该圆环计算noise, 事实上可结合该圆环确定噪音与上面孔径测得的流量计算信噪比snr, 程序中有实现但未将结果输出到catalog; catalog中 `intnsnr` 使用的是  $snr \sim A/C$

## SAVE

程序中保存结果的文件夹相关设定, 通常不需要改变

- **dir\_type:**  
存储结果的文件夹命名方案, 详见配置文件中的注释
- **name\_catalog:**  
生成的Catalog的命名

## SDSS

下载并保存SDSS Catalog使用的参数

- **skip\_save:**  
True或False, 为True时当程序找到本地对应Catalog时, 不再从网络上下载并覆盖
- **dec\_width:**  
指定fitscube的赤纬(Dec)方向的覆盖范围, 单位为角度, 请根据实际需要修改
- **zmin/zmax:**  
SDSS Catalog的红移区间

## FITS

- **back\_str:** 用于识别fitscube的文件名后缀

## Catalog

---

对程序生成的Catalog中每一列的说明

- **id:**  
源对应应在SDSS Catalog中数据的第几行, 从0开始计数
- **ra/dec:**  
源的赤经/赤纬, 读自SDSS Catalog中ra、dec列, 单位角度
- **vel:**  
HI源的多普勒红移换算的速度, 红移读自SDSS Catalog,  $vel=c*z$ , 单位m/s

- peakvel:  
扣基线后, 预估的 $\pm w_{50}$ 范围内信号最大值对应的速度, 单位m/s
- w50/w20:  
扣基线测量方法得到的w50和w20, 由于测量不准确, 目前没有实际实现, 值设为-1
- flux:  
扣基线后, 预估的 $\pm w_{50}$ 范围内信号对速度的求和。  
如果fitscube数据单位是Jy, 则flux单位为Jy \* m/s
- rms:  
扣除基线后, 预估的 $\pm w_{50}$ 范围之外的信号的rms, 单位同fitscube数据单位
- snr:  
扣基线后, 预估的 $\pm w_{50}$ 范围内信号最大值与rms之比
- intf:  
第二种测量方法中Aw的值, 表示积分流量。  
如果fitscube数据单位是Jy, 则flux单位为 $\sqrt{m} \text{ Jy m/s}$
- intw:  
第二种测量方法中w的值, 表示等值宽度。单位为m/s
- inta:  
第二种测量方法中A的值, 表示等效的信号流量。单位同fitscube数据单位
- intsnr:  
第二种测量方法中A/C的值, 表示信噪比。
- interr:  
第二种测量方法中拟合曲线与观测值之差的标准差, 表征拟合效果

## FAQ

1. 程序运行后, 会在 `DIR.work_dir` 下创建数个文件夹:

- optimgs:  
存放下载下来的光学图像, 用于绘图。如果文件夹中没有程序需要的图像文件, 会自动下载至该文件夹中。
- SDSS:  
存放下载下来的SDSS光学源Catalog, 用于在fitscube中搜寻对应源的中性氢信号。如果文件夹中没有程序需要的图像文件, 会自动下载至该文件夹中。
- results:  
存放分析结果的文件夹, 每个fits文件的分析结果命名由 `SAVE.dir_type` 决定。
- temp:  
存放临时文件的文件夹, 目前没有用。

2. 程序提供了两种测量的方法与结果:

1. 在SDSS光学数据推断出的射电速度 $x_0$ 附近 $x \in \{x | \text{abs}(x - x_0) < w_{50} \cdot \text{view\_scale}\}$ , 确定脏信号 $y(x)$ , 在周围同Dec处采样计算基线 $bsl(x)$  (中值法), 将脏信号扣除基线得到信号 $\text{sig}(x) = y(x) - bsl(x)$ , 表征fitscube中给定RA、Dec处, 速度在 $x \in \{x | \text{abs}(x - x_0) < w_{50} \cdot \text{view\_scale}\}$ 内的信号, 测量并给出峰值流量、信噪比等信息 (详见 `ParallelCommander.py` 340行附近)。

注: 该方法受到噪音、驻波的影响很大, 且较难去除, 如果源的信号较弱, 通常结果不可靠。同样因为这个原因, 没有对w50、w20进行测量, 生成的catalog中这两列数据为-1

2. 假设在信号对应的速度附近, 真实射电信号集中在 $x_0 - 0.5w < x < x_0 + 0.5w$ 、 $\text{sig}(x) = A$ 的矩形内, 噪音包括期望为 $C$ 的泊松噪声 $\text{noise}_1(x) = P(x, C)$ 和驻波项 $\text{noise}_2(x) = B \sin(\omega x + \varphi)$ , 则理论上观测的信号 $y_{th}(x) = \text{sig}(x) + \text{noise}_1(x) + \text{noise}_2(x)$ , 即一个分段函数:

$$y_{th}(x) = \begin{cases} A + B \sin(\omega x + \varphi) + P(x), & |x - x_0| < 0.5w \\ B \sin(\omega x + \varphi) + P(x), & |x - x_0| \geq 0.5w \end{cases}$$

若将信号对称地沿速度积分，可得到：

$$\int_{x_0-x}^{x_0+x} y_{th}(x') dx' = 2 \int_{x_0}^{x_0+x} y_{th}(x') dx' = \begin{cases} 2Ax + B' \sin(\omega x + \varphi') + 2Cx, & |x - x_0| < 0.5w \\ Aw + B' \sin(\omega x + \varphi') + 2Cx, & |x - x_0| \geq 0.5w \end{cases}$$

记  $t = 2(x - x_0)$ ，考虑  $t > 0$ ，可将积分变换为：

$$F_{th}(t) = \begin{cases} At + B'' \sin(\omega''t + \varphi'') + Ct, & 0 < t < w \\ Aw + B' \sin(\omega''t + \varphi') + Ct, & t \geq w \end{cases}$$

而对观测得到的信号  $y_{obs}(x)$ ，类似地积分（沿速度方向求和）得到  $F_{obs}(t)$ 。用

`scipy.optimize.curve_fit` 以  $F_{th}(t)$  为模型对  $F_{obs}(t)$  进行拟合，可以确定参数  $A$ 、 $w$  与  $C$ 。此时，将  $A \cdot w$  之积作为信号的积分流量， $w$  作为等值宽度， $A/C$  作为信噪比，标准差  $\text{std}(F_{th} - F_{obs})$  作为表征拟合结果可信程度的量。

注：该方法规避了泊松噪音的影响，鲁棒性更好，结果较为可信，但对于弱源（ $A \sim C$ ）参数不一定可信；同时该方法依赖于信号的模型，泛用性有待验证。

### 3. 程序生成内容解释：

- 结果保存在 `results` 目录下，每个 fits 文件每次运行单独一个目录
- 汇总图、生成的 catalog、程序运行的 log 保存在每个 fits 文件运行产生的目录中
- 测量方法2中的函数拟合图保存在 fits 运行产生目录中的 `measure` 目录下
- 图片文件的序号对应其在下载的 SDSS Catalog 中第几行，从0开始计数
- `log_数字.txt` 的不同数字对应不同进程
- 程序运行下载的 SDSS Catalog 保存在工作目录的 `sdss` 目录中
- 程序运行下载的光学图像保存在工作目录的 `optims` 目录中

### 4. 因为效果不佳，程序没有实现利用FFT降噪的内容

### 5. 其他问题大多可通过阅读源码实现，程序编写时已尽可能让代码更直观，有需要可以尝试直接修改源码实现目的

## 联系我/Contact Me

[zhengcf@mail.bnu.edu.cn](mailto:zhengcf@mail.bnu.edu.cn)