

SP-1877

Updates to PSS Test Vector Generation process

Benjamin Shaw + Lina Levin Preston
For PSS

Team PSS

Oxford University

Aris Karastergiou (PO)

Arun Naidu

Christopher Williams

Raman Research Institute

Prabu Thiagaraj

University of Manchester

Atul Ghalame

Michelle Hussey

Mitch Mickaliger

Vincent Morello

Mike Pearson (SM)

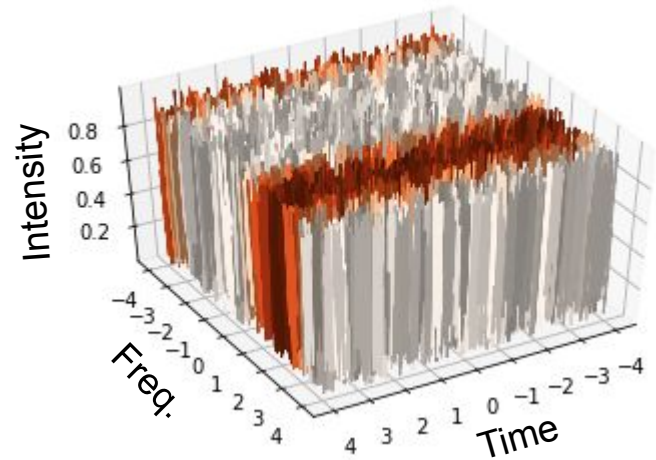
Lina Levin Preston

Benjamin Shaw

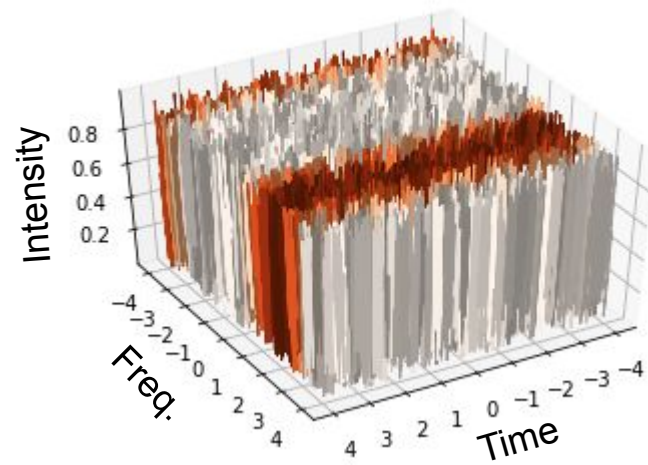
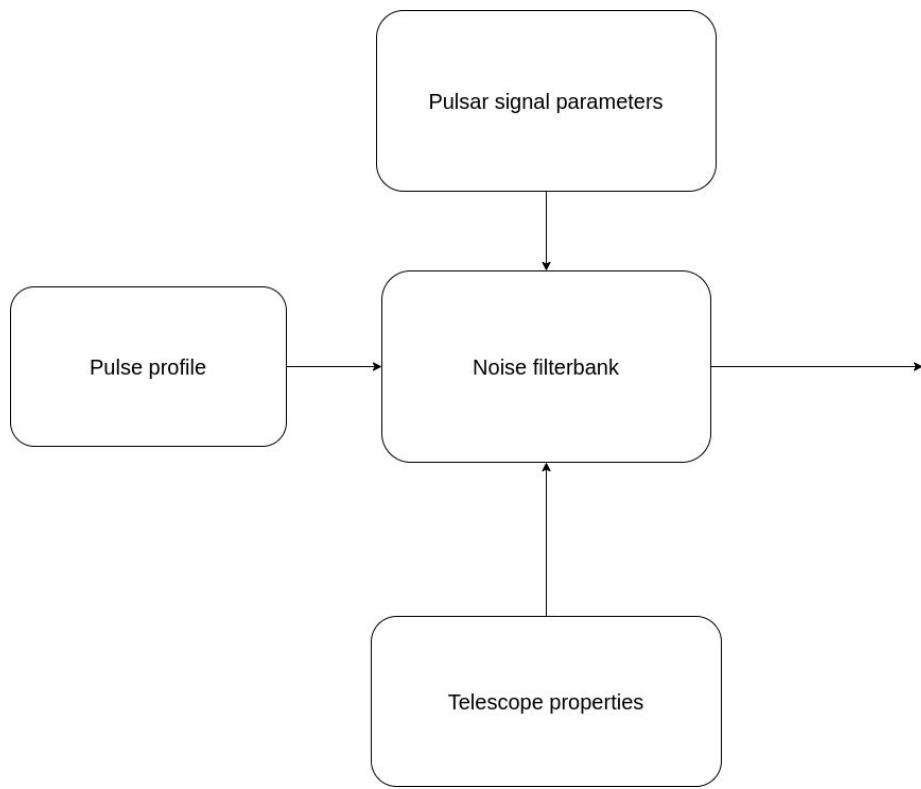
Ben Stappers (PO)

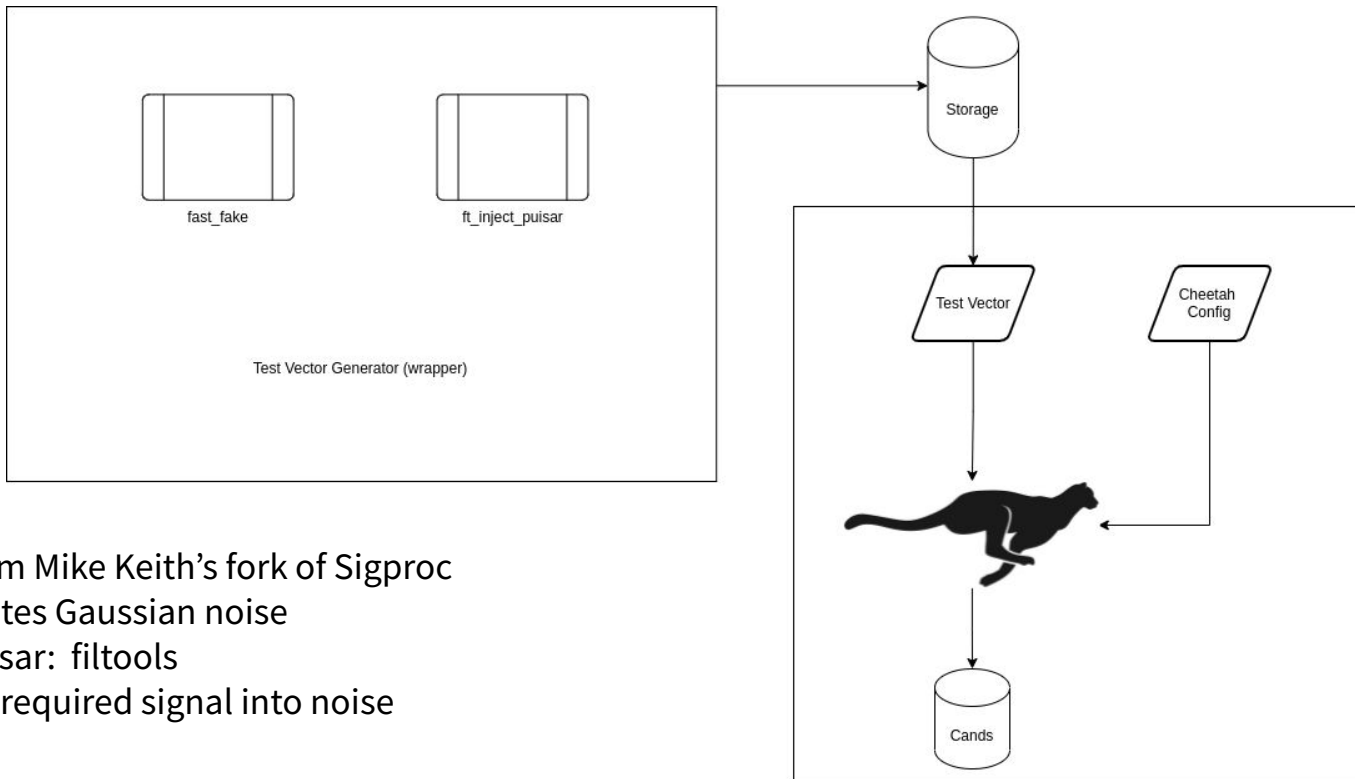
Test Vectors

- 4-D Data hyper-cube - signal intensity vs time, frequency, polarisation.
- Examples of real astrophysical signals (periodic (pulsars) and transient (FRBs, GP, RRATs)).
- SKA sized test vectors are up to 36GB based on current requirements.
- Crucial for
 - demonstrating pipeline performance
 - demonstrating compliance to system requirements
 - ad-hoc/unit/system/pipeline testing



Test Vector





fast_fake: from Mike Keith's fork of Sigproc

- Generates Gaussian noise

ft_inject_pulsar: filtools

- Injects required signal into noise


<https://github.com/SixByNine/sigproc>

<https://bitbucket.org/mkeith/filtools/src>

Shortcomings

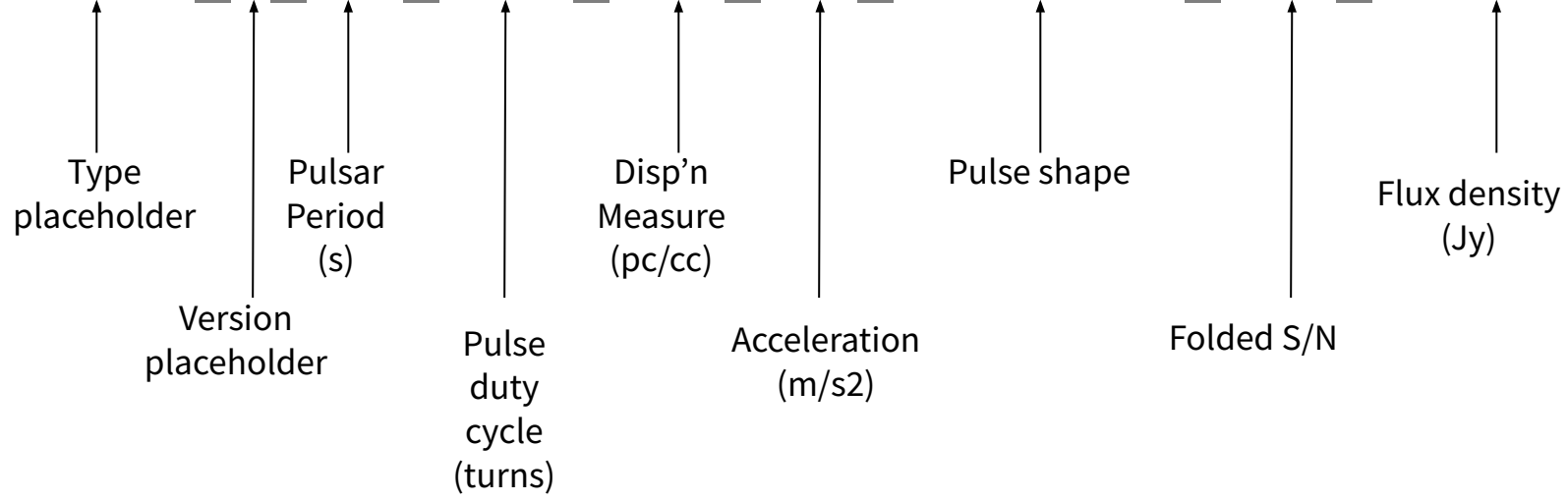
- Test vectors are name according to their properties but some traceability information is missing (e.g., which version produced them?).
- Old versions of vectors need to be retained and “latest” versions need to be easily available.
- Directory structure needs to reflect different vector types that will be available for download]
- Cheetah configs are name arbitrarily - need some standard (or some method of generating them on-the-fly)

Tickets

- AT4-559 Establish new name convention for PSS test vectors
 - AT4-560 Update PSS test vector generator to conform to new naming standard
 - AT4-561 Establish directory structure for new test vector repository on dokimi
 - AT4-562 Update PSS test vector generator to publish test vectors to repository on dokimi
 - AT4-563 Establish naming convention for cheetah config files
- 

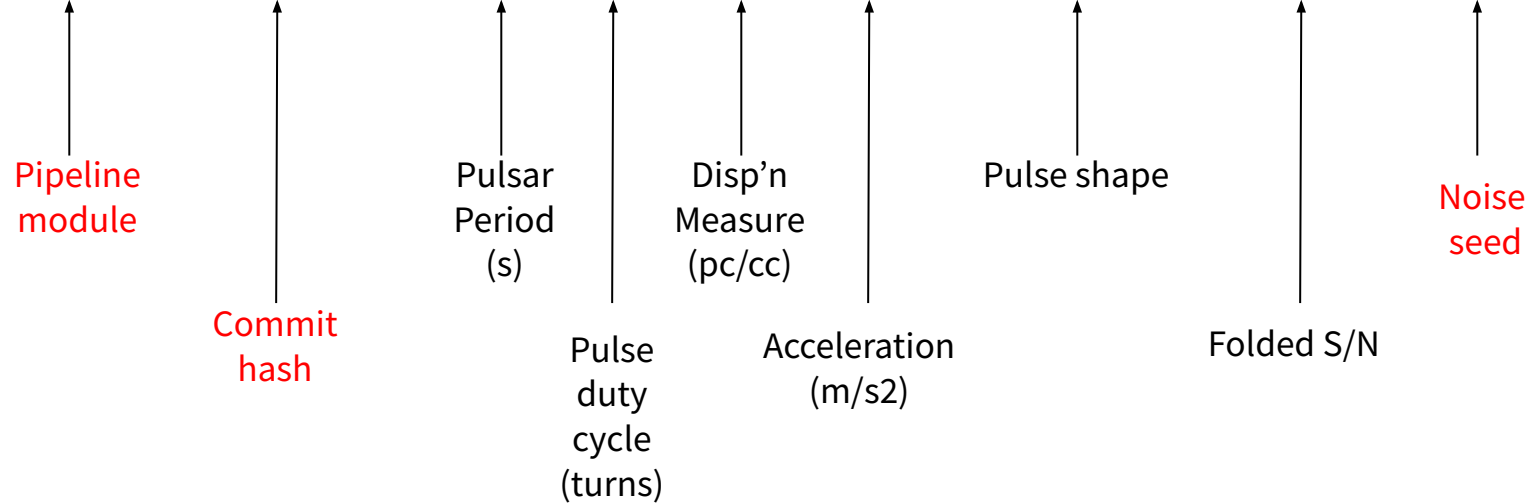
AT4-559 Establish new name convention for PSS test vectors

Default_1_0.5_0.05_100_0.0_Gaussian_100_0.0005.fil



AT4-559 Establish new name convention for PSS test vectors

TDAS_1137b9c_1.0_0.2_50.0_0_Gaussian_100.0_5050.fil



AT4-561 Establish directory structure for new test vector repository on dokimi

- Previously, vectors have been generated on an ad-hoc basis when required.
- Here we set up a formal repository for test vectors on our test-vector server *dokimi* (located at JBO)
- Requirements
 - Should include all available vectors for a range of pipeline module tests
 - Be exposed to the web for testers/gitlab pipelines to download vectors as required
 - Should be easy to maintain and modify/add to
 - “Latest” versions should be available but older versions should be retained

```
[bshaw@dokimi bshaw]$ df -h | grep -v tmpfs
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/nvme0n1p2	326G	232G	94G	72%	/
/dev/nvme0n1p5	1.5T	86G	1.5T	6%	/ssd
/dev/nvme0n1p1	497M	266M	231M	54%	/boot
/dev/sda1	55T	54T	1.3T	98%	/raid
192.168.84.2:/volume1/data1	95T	28T	68T	29%	/skatvnas1
192.168.84.2:/volume2/data2	95T	600G	94T	1%	/skatvnas2
192.168.84.2:/volume3/data3	95T	1.1T	94T	2%	/skatvnas3

AT4-563 Establish naming convention for cheetah config files

We don't have a convention at all currently for naming cheetah config files.

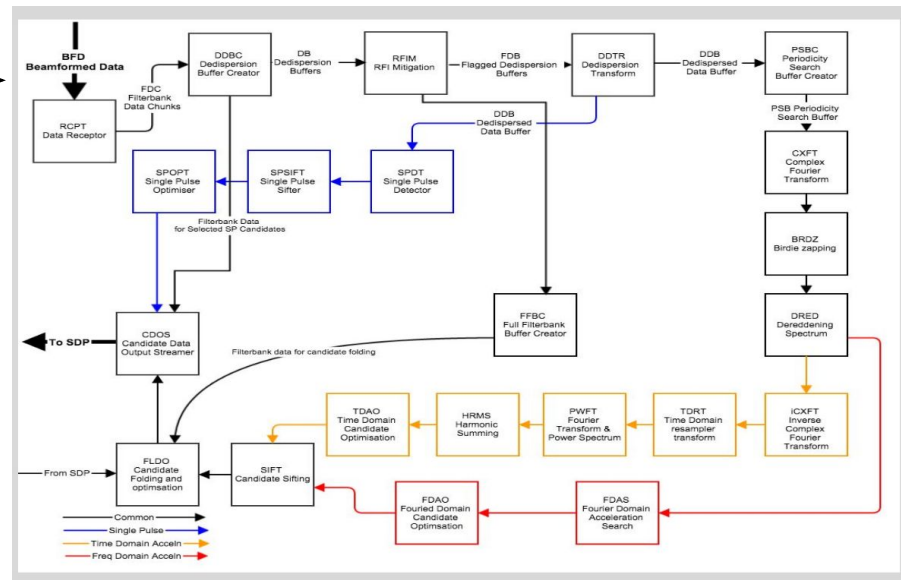
Configured by xml/json input

All (sub-)components are configured here

Huge number of parameters/possibilities.

Difficult to meaningfully name these files or understand the files contents without inspection

```
<active>false</active>
<channel_rejection_rms>1.0</channel_rejection_rms>
<spectrum_rejection_rms>1.0</spectrum_rejection_rms>
<max_history>100</max_history>
<zero_dm>false</zero_dm>
<bandpass>
  <channel_width>10000</channel_width>
  <freq_start>300000000</freq_start>
  <number_of_channels>30</number_of_channels>
  <sps>2</sps>
  <rms>12000</rms>
</bandpass>
</rfim_amp>
</rfim_iqrcpu>
<active>false</active>
<max_lag>1</max_lag>
<sigma>0</sigma>
<edge_channels>45</edge_channels>
</rfim_iqrcpu>
<rfim_cuda>
  <active>false</active>
</rfim_cuda>
<rfim_sum_threshold>
  <active>false</active>
  <sts_cutoff>0</sts_cutoff>
  <base_sensitivity>1.0</base_sensitivity>
  <window>0</window>
</rfim_sum_threshold>
</rfim>
<sps>
  <pool_id>cpu_pool</pool_id>
  <priority>0</priority>
  <threshold>0.0</threshold>
  <dedispersion_samples>131072</dedispersion_samples>
  <astroaccelerate>
    <active>false</active>
  </astroaccelerate>
  <emulator>
    <candidate_rate></candidate_rate>
  </emulator>
  <dedispersion>
    <start></start>
    <end></end>
    <step>1</step>
    <id />
  </dedispersion>
</sps>
<sps_clustering>
```



AT4-563 Establish naming convention for cheetah config files

- Pipeline tests will require large numbers of distinct configurations
- These may be similar in purpose (with subtle differences between them, e.g., to test edge cases)

Good news:

- Cheetah itself can generate config file templates.
- We will create a (set of) python class(es) that will generate and populate configuration files “on-the-fly” as part of the setup for a forthcoming test.
- This makes us immune from changes to cheetah configs

File names: cheetah_config_<ID>.xml

Where ID is a testID.

Feature planned for future PI: See SP-1914

DEMO

To come

- Many more features planned to design/build the cheetah testing framework
- See Lina's demo on our single pulse search (SPS) test vector generation and validation for testing our radio transient detection pipeline
- Dokimi isn't publicly exposed to the web - need a reverse proxy (in progress + next PI).
- Where will the test vectors *ultimately* live?
 - What backend app do we need to parse queries and supply vectors? Flask? Something else?