Collaboration and data reduction software
Theme 5

Enabling pulsar timing science with the SKA

Team members:

- Yihan Tao
- Kun Li
- Lingling Wang
- Slava Kitaeff
- George Hobbs

April 9-13 2017
Shanghai

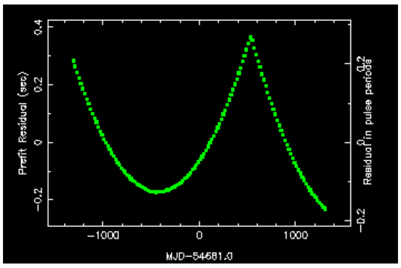
Overview and goals for the workshop

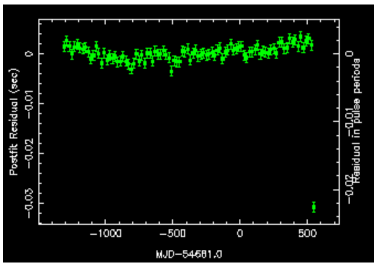
SKA-low and SKA-mid science projects include pulsar observations. In this workshop we have considered how SKA regional centres may be used to visualise and process pulsar timing observations. The team members consisted of experts in interactive collaboration, algorithm development, cloud-based processing, pulsars and in data exploration.

The work was based around simulating pulsar timing datasets using the ptaSimulate software that was installed on a virtual machine. Our primary goal was to determine:

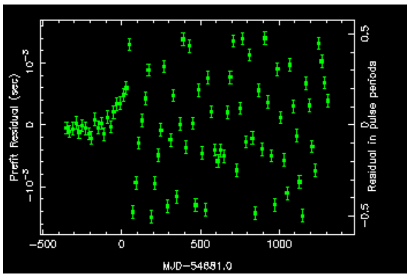
- How to identify a few pulsars out of 10,000 that have exhibited a sudden change in behaviour

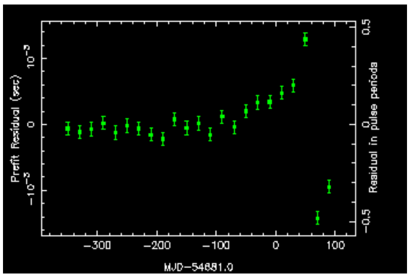
We noted that in the 2-day workshop we would not be able to complete this goal, but we made significant progress. As an example we show below a simulation of a pulsar glitch event. The figures represent the timing residuals for the pulsars. In the left-hand panel we identify a sudden change in the pulsar pulse rate around two-thirds of the way into the data span. Such large events are easy to spot by eye. In the right-hand panel we show what a single observation looks like after the glitch event and clearly the event is much harder to identify.





As another example we simulated a pulsar in which the intrinsic noise in the pulsar became so large that we lost track of the phase of the pulsar. Again, such events are hard to identify close to the time that it occurred.





We therefore defined the following goals to explore during the workshop:

1. How quickly can we identify that the pulsar signal has changed?
2. How to automate the classification of those “change of behavior” events (glitch, phase wrap, off, pulse shape change, telescope malfunction etc.), and what resources will be required at the SKA Regional Center?
3. How can we visualise a large number of pulsars in a sensible way?
4. What is the optimal way to store the pulsar data to enable a collaborative post-processing?
5. How to detect problems with SDP pulsar models, and how to correct them?

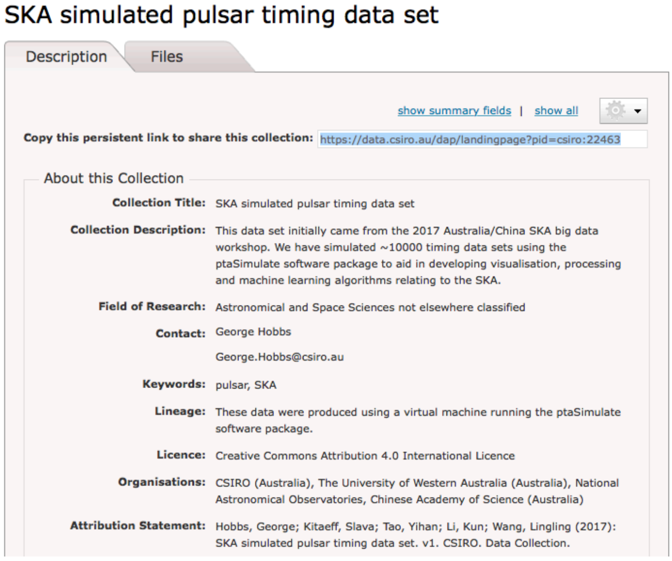
The workshop: plan and results

Our plan to tackle these questions was to:

1. Simulate a relatively small number of examples of interesting pulsars
2. Simulate 10,000 “normal” pulsars
3. Discuss and identify the various possible approaches to answer the questions listed above.

We were successful in completing this plan. In particular, we installed the pulsar simulation software on five laptops and produced 9300 pulsar simulations. 9000 of these were of a “normal” pulsar with a period of 1 second, a small amount of timing noise and arrival time uncertainties of 1ms.   We also simulated 200 millisecond pulsars in which we injected the signature of a gravitational wave background signal with amplitude $A = 10^{-15}$. These pulsars had error bars of 50ns. We simulated 50 pulsars that exhibited a glitch event. The time and amplitude of the glitch event was randomly chosen (with glitch amplitudes between 10^{-7} to 10^{-10}). Finally we simulated 50 data sets in which around half of the points were corrupted. Such corruption could model an intermittent pulsar, RFI or failures in the system.

We have setup a repository to store these simulated files (https://data.csiro.au/dap/landingpage?pid=csiro:22463). This repository is currently private, but will be made public at a later date:


SKA simulated pulsar timing data set

Description | Files

show summary fields | show all

Copy this persistent link to share this collection: https://data.csiro.au/dap/landingpage?pid=csiro:22463

About this Collection

Collection Title: SKA simulated pulsar timing data set

Collection Description: This data set initially came from the 2017 Australia/China SKA big data workshop. We have simulated ~10000 timing data sets using the ptaSimulate software package to aid in developing visualisation, processing and machine learning algorithms relating to the SKA.

Field of Research: Astronomical and Space Sciences not elsewhere classified

Contact: George Hobbs

George.Hobbs@csiro.au

Keywords: pulsar, SKA

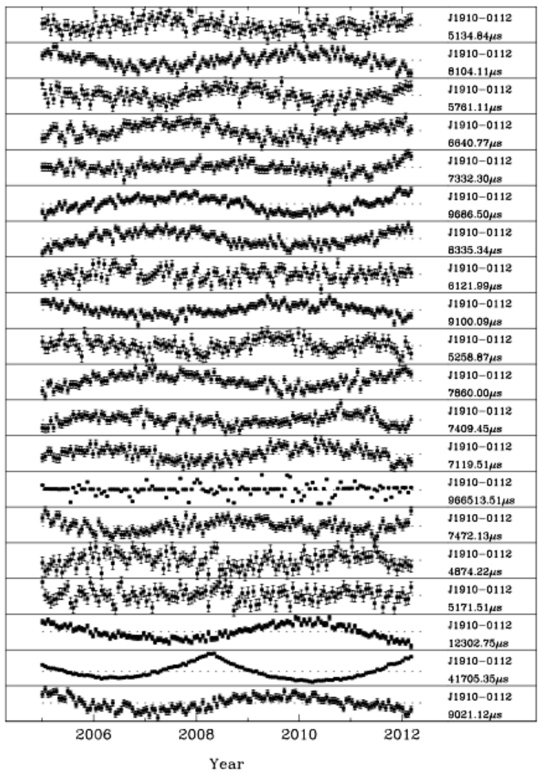
Lineage: These data were produced using a virtual machine running the ptaSimulate software package.

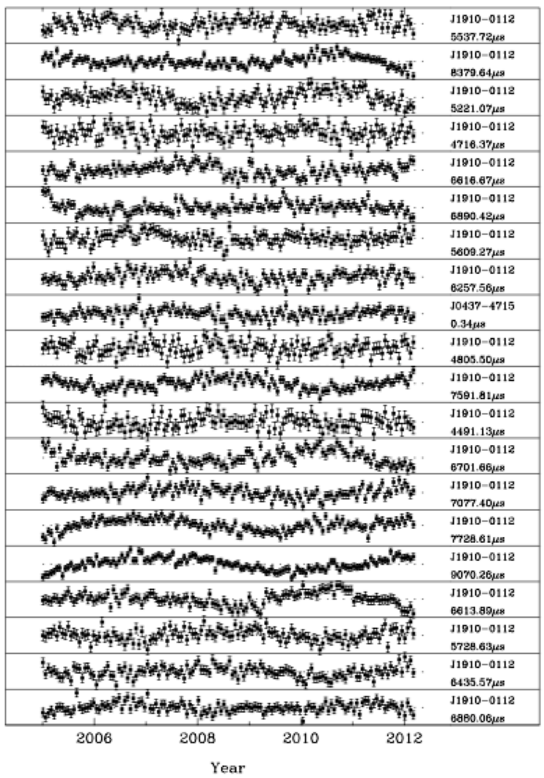
Licence: Creative Commons Attribution 4.0 International Licence

Organisations: CSIRO (Australia), The University of Western Australia (Australia), National Astronomical Observatories, Chinese Academy of Science (Australia)

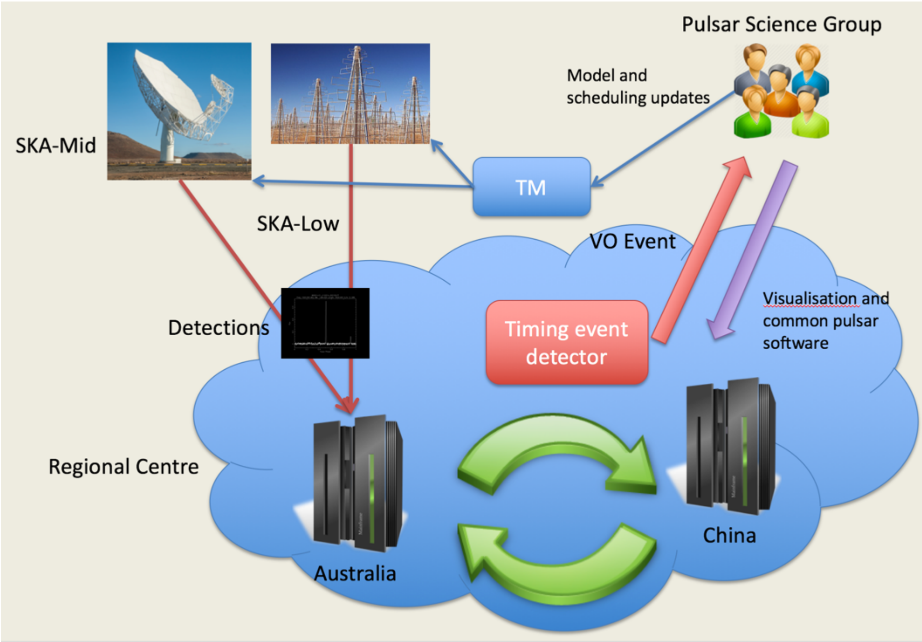
Attribution Statement: Hobbs, George; Kitaeff, Slava; Tao, Yihan; Li, Kun; Wang, Lingling (2017): SKA simulated pulsar timing data set. v1. CSIRO. Data Collection.

Examples of our simulations are shown below (using the plotMany plugin to tempo2 to make the images). Note that we have 465 such images and so this is not a viable way to visualise such large amounts of data.





Significant discussion occurred around how users would interact with a regional science centre for pulsar timing observations and what feedback may be necessary from those users/teams back to the telescope manager/SDP groups etc. Our initial thoughts can be summarised in the diagram below.



In brief, we expect that automatic algorithms will alert the pulsar teams of "interesting" events that may have occurred. The groups will then be able to collaborate together through the Regional centre to study any such events. It may be that the results of those studies needs to lead back to the systems in place for measuring pulse ToAs and for any online folding.

Long-term plan

We have group and individual goals to take this work further. The primary goals for future work are:

1. Improve the simulations (make it more realistic to what SKA will produce, explore the differences between SKA-Mid and SKA-Low)
2. Build two SKA pulsar post-processing prototypes in Australia and China (the resources are available)
3. Explore the options for databases and data analysis tools.
4. Investigate, develop and compare the algorithms to detect the changes in pulsar behavior or telescope malfunction (ML, statistical methods, digital filters, humans).
5. Develop a new remote pulsar visualisation tool.
6. Produce a report for SKA pulsar WG, SDP, SCP, FAST and anyone else that may be interested
7. Publish (simulation data challenge, algorithms, framework).

Specifically the individual team members plan to:

Yihan Tao

- Distributed pulsar databases and data analysis/visualisation tool
  - China-VO team, Aliyun (cloud)
  - Cooperate with Lingling & Slava
- Algorithms to detect changes in pulsar behavior using ML

- Collaborate with Slava's team to test several classification algorithms (random forest, neural networks, SVM etc.)

Kun Li

- Visualization
    - Simulate pulsar data, find interesting and important parts, and develop a remote visualization tools to analyze them.
- Fast Fourier Transforms
    - Develop and optimize Fast Fourier Transforms algorithm on large-scale clusters to satisfy the collaborative requirements of dealing with massive pulsar data.
    - Develop Fast Fourier Transforms algorithm which will have a good performance on heterogeneous to deal with the signal data produced by pulsar.
- Analysis
    - Communicate with the whole team and share the visualization results.
    - Make full use of visualization tools and Fast Fourier Transform algorithms to deal with various data and based on that corporate with team to do a detailed analysis.

Lingling Wang

1. Build the cluster in China on the 5 machines
2. Explore or test the options of database or storage backend
3. Investigate the algorithms to detect the changes in pulsar behavior or telescope malfunction

George Hobbs

1. Find a student who wants to work on this type of research (likely linked between SKA, Parkes and FAST)
2. Make more realistic simulations of SKA-low and mid and more realistic noise processes, pulsar sky positions and observation cadences
3. Check the simulations with SKA pulsar WG to ensure realism and that we're not missing anything obvious.
4. Work on a statistical method for determining the likelihood that the noise process in a pulsar data set has changed.
5. Apply this method to the simulations and compare with other algorithms being developed
6. Work with the team to write a publication describing the results.

Slava Kitaeff

1. I'll be working with two PhD students. We have a functioning Apache Spark/Hive cluster available that we can use. We will collaborate with Lingling Wang assisting her in building a similar system in China.
2. I'm thinking to prototype the SKA pulsar database with the simulation dataset, and I'd like to look into the options for event detection and classification algorithms. Yihan, George and I have discussed collaboration in this area.
3. We will look into the development of new remote visualisation software for pulsars.

Conclusion

The workshop has been a great success and we now have developed new collaborations. We thank the International Centre for Radio Astronomy Research and Shanghai Astronomical Observatory for organising the meeting and for the Australian Department of Industry, Innovation and Science and the Ministry of Science and Technology in China for funding the meeting.