URSI - RCRS 2022, IIT (Indore), India, 1 - 4 December, 2022

Square Kilometre Array Pulsar Search Pipeline

Thiagaraj Prabu (and SKA PSS team)
(1) Raman Research Institute, Bangalore, http://www.rri.res.in

Abstract

The Square Kilometre Array (SKA) project aims to realize a state-of-the-art and most sensitive radio telescope to address various important science goals. The project is moving towards final implementation. Pulsar search is one of the major programs planned with SKA. It is a complex task; about 60 petabytes of data are to be handled daily, needing nearly 10 Peta-ops processing capabilities. Such complex processing requirements necessitate exploring high-performance computing options and using novel accelerators based on modern FPGAs. There are ongoing contributions from the team members worldwide to develop the SKA transient search pipeline and its integration and validation. This paper will outline the SKA transient search pipeline details, highlighting the salient features designed for accelerating the search algorithms.

1. Introduction

The unprecedented sensitivity of the Square Kilometre Array (SKA) telescope will aid in investigating areas as diverse as the formation of Earth-like planets, the detection of gravitational distortions of Space-Time, the origin of cosmic magnetic fields, and the formation and growth of Black Holes. SKA will have immense sensitivity to detect several highly accelerated pulsar systems and other transients (FRB). The timing studies of the pulsed radio emissions from these distant millisecond pulsars directly investigate the space-time metrics, which is expected to open a new window to explore Gravitational Wave Astronomy [1-3]. This paper presents an outline of the SKA transient search pipeline design with a highlight of novel FPGA-based high-performance computing developed to achieve real-time performances.

2. Pulsar Search with SKA

The current phase of the SKA (phase-1) will quadruple the known population of pulsars. It will be able to detect all pulsars in the Galaxy that beam toward Earth. Pulsar Search Techniques have matured over five decades, and today it is possible to plan real-time search for pulsars and other transient signals in real-time.

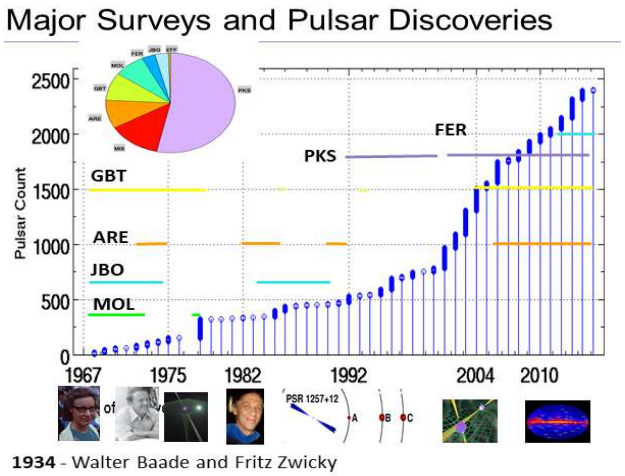

Fig. 1 Discovery Timeline of pulsars.

The computational complexity for a pulsar search comes from several factors. The search accounts for a range of distances, periods, and binary orbits and deals with RFI. The signals are multi-bit quantized, observations are made across an extensive band of frequencies, and the sky is searched simultaneously in more than 1000 directions [2].

The search effort involves the real-time processing of significant amounts of data from the SKA telescope for each observation. Real-time processing is mandated due to the mammoth volume of data generated by the SKA telescope during a search; besides, the data cannot be practically stored for offline processing. High-performance computing (HPC) solutions are explored to cope with real-time processing [4,5].

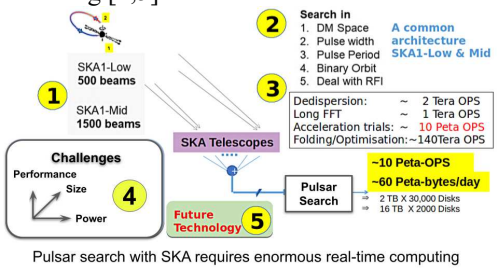

Fig. 2 High performance computing needed for the SKA Transient search [8]

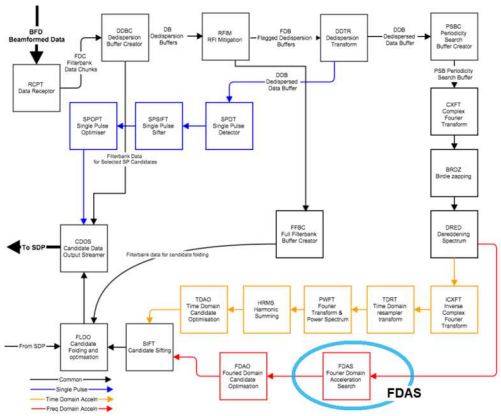

Fig. 3 SKA Transient search pipeline [2].

An outline of the SKA pulsar search pipeline is shown in Fig. 3. Beam-formed data streams from each beam arrive (~10 minutes integration) at the search pipeline. The dedispersion transform (DDTR) module corrects the dispersion-related delays caused by the interstellar free electrons over the observed frequency band. The computational complexity for dispersion across 6000 trial dispersion is estimated to be 2-teraOPS. The single pulse detection (SPDT) module searches for individual pulse occurrences (such as the giant pulses from pulsars and FRB signals). The CXFT module performs a long-length Fourier Transform to facilitate search in a Fourier domain. Computational complexity associated with the long FFT is about 1-teraOPS. The acceleration search module accounts for changes in pulse period caused by their orbital motions (Fig. 4). Computational complexity of the Fourier domain-based acceleration search [7, 9] is estimated to be about ten petaOPS. A list of candidates produced from the Fourier domain search will be passed to a folding and optimisation module (FLDO), where the significance of the candidates is obtained (SNR) and sent for further processing in the subsequent science processing stages. The computational complexity of the FLDO stage is about 140 teraOPS. Radio frequency interferences (RFI) affect the data quality. RFI is dealt with within the search pipeline at multiple locations. The science processing module compares detections across suitable neighbouring beams. It deals with RFI using heuristic algorithms. The significant candidates identified are sent along with corresponding raw data for archive and follow-up analysis.

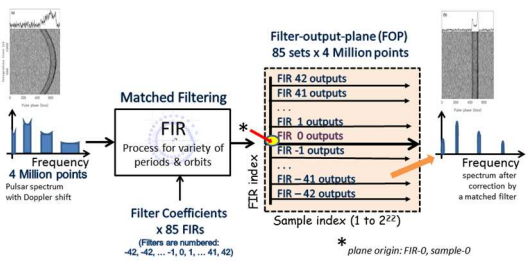

Fig. 4 FDAS Accelerated pulsar search [8]

The search pipeline employs multiple server nodes with CPU, GPU and FPGA-based high-performance computing accelerators (Fig. 5). CPU and GPU-based accelerations have been in the high-performance computing market for a longer time than the FPGA-based accelerator. FPGA-based acceleration is a new entry into high-performance computing but provides a unique handle to problems where high power performances are desirable (Fig. 6).

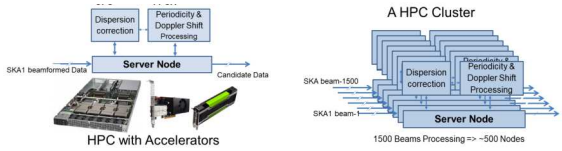

Fig. 5. HPC Cluster configuration

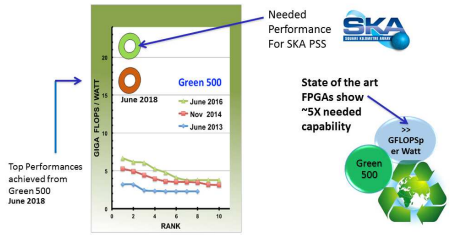

Fig. 6 Modern FPGAs offer high power performances

We have prototyped a few compute-intensive modules from the pulsar search application in the FPGA accelerators. Typically, the algorithms are developed using hardware description (VHDL) and high-level (OpenCL) languages . FPGA design architectures, and in some cases, architectures and codes investigated for

- Median filtering for RFI mitigation,
- Dedispersion,
- Single pulse detection,
- Long FFT,
- Acceleration search, and
- Candidate folding optimisation modules

As a novel approach, the machine learning techniques' efficacy has been explored to detect candidate signatures in the acceleration search module's intermediate stage outputs shown in Fig. 7 [10].

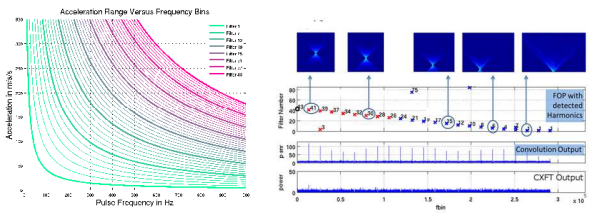

Fig. 7 Machine learning explored to detect candidates from the accelerated search intermediate stage outputs.

3. Summary

A high-performance computing solution for the SKA transient pipeline has been developed. This pipeline is being integrated and characterized during the current phase of SKA. Several student participation areas have also been identified from this work. The talk will highlight these salient aspects of the work and future timelines.

Acknowledgements

Square Kilometre Array (SKA) is one of the international mega-science projects being co-supported in India by the DAE and DST.

References

[1] Keane, E., Bhattacharyya, B., Kramer, M., Stappers, B. et al. Advancing Astrophysics with the Square Kilometre Array (AASKA14), 40, 2015

[2] L. Levin et. al., Pulsar Searches with the SKA, Pulsar Astrophysics -The Next 50 Years, Proceedings IAU Symposium No. 337, 2017

[3] Square Kilometre Array
https://www.skatelescope.org

[4] Dimoudi et al., 2018
https://iopscience.iop.org/article/10.3847/1538-4365/aabe88/pdf

[5] Wang et al., 2018 (arXiv:1805.12280)

[6] Andrew Lyne, Francis Graham-Smith, Pulsar astronomy, Cambridge Astrophysics, 2006

[7] S. Jouteux, R. Ramachandran, B. W. Stappers, P. G. Jonker, M. van der Klis, Searching for pulsars in close circular binary systems, textbfA&A 384, 532-544 (2002) DOI 10.105100046361 20020052

[8] P. Thiagaraj *et al*., "FPGA architecture to search for accelerated pulsars with SKA," *2020 XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science*, 2020, pp. 1-5, doi: 10.23919/URSIGASS49373.2020.9232304.

[9] Scott M. Ransom, Stephen S. Eikenberry, And John Middleditch, Fourier Techniques For Very Long Astrophysical Time-Series Analysis, The Astronomical

[10] S.S.Bhat et al., JoAA, special issue 2022 (accepted).