Building The World's Largest Radio Telescope: The Square Kilometre Array Science Data Processor

1st Jamie S. Farnes
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
jamie.farnes@oerc.ox.ac.uk

2nd Ben Mort
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
ben.mort@oerc.ox.ac.uk

3rd Fred Dulwich
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
fred.dulwich@oerc.ox.ac.uk

4th Karel Adámek
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
karel.adamek@oerc.ox.ac.uk

5th Anna Brown
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
anna.brown@oerc.ox.ac.uk

6th Jan Novotny
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
jan.novotny@oerc.ox.ac.uk

7th Stef Salvini
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
stef.salvini@oerc.ox.ac.uk

8th Wes Armour
Oxford e-Research Centre (OeRC)
University of Oxford
Oxford, UK.
wes.armour@oerc.ox.ac.uk

9th SKA SDP Consortium

***Abstract*—The Square Kilometre Array (SKA) will be the largest radio telescope constructed to date and the largest Big Data project in the known Universe. The first phase of the project will generate 160 terabytes every second. This amounts to 5 zettabytes (5 million petabytes) of data that will be generated by the facility each year – a data rate equivalent to 5 times the estimated global internet traffic in 2015. These data need to be reduced and then continuously ingested by the SKA Science Data Processor (SDP). Within the SDP Consortium, we are contributing to various roles in the development of the telescope including building a lightweight end-to-end prototype of the major components of the SDP system – a project we call the SDP Integration Prototype (SIP). The aim is to build a mini, fully-operational SDP, for which we have been developing realistic SKA-like science pipelines that can handle these unprecedented data volumes.**

***Index Terms*—antennas and propagation, radio astronomy, data processing, radio interferometry, astrophysics**

I. Introduction

When complete, the Square Kilometre Array (SKA) will be the largest radio telescope ever constructed. The facility will span two continents, with hundreds of antennas being located in South Africa and $\sim 100,000$ dipole antennas being situated in Australia. Through the technique of interferometry, the signals from these antennas can be combined so that the array behaves like a single dish that is 150 km in diameter.

The UK, via the STFC, are contributing £100M towards construction of the first phase of the project. The SKA Organisation itself is located next to Jodrell Bank Observatory in Manchester, with the UK community being led by Manchester, Oxford, and Cambridge, together with UK industrial partners.

Science with this state-of-the-art radio telescope will address a diverse range of leading scientific questions, including: How do galaxies evolve? What is the nature of Dark energy? Was Einstein right about gravity? What generates giant magnetic fields in space? and are we alone in the Cosmos? [1].

II. eScience Challenges

The first phase of the project will generate on the order of exabytes ($= 1,000$ petabytes) of raw data from the antennas every day (up to 160 terabytes per second). After a data reduction stage, up to 1 terabyte of data per second needs to be ingested on a continuous basis at each SKA site. This data flood from the SKA will have implications for medical researchers, neuroscientists, data scientists, quantitative financial analysts, and the general public throughout the global community, in terms of how we handle and process such large datasets. This work may also impact other scientific disciplines, by providing a flexible architecture for computationally intensive processing problems and via development of novel parallelised algorithms.

The infrastructure for handling these data rates does not currently exist. To some extent, new technologies will enable us to process the data when the facility comes online in

This work is supported by Science and Technology Facilities Council (STFC) Grants (see acknowledgments).

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00101

2023. We are also developing new techniques, algorithms, and infrastructure to solve these data challenges.

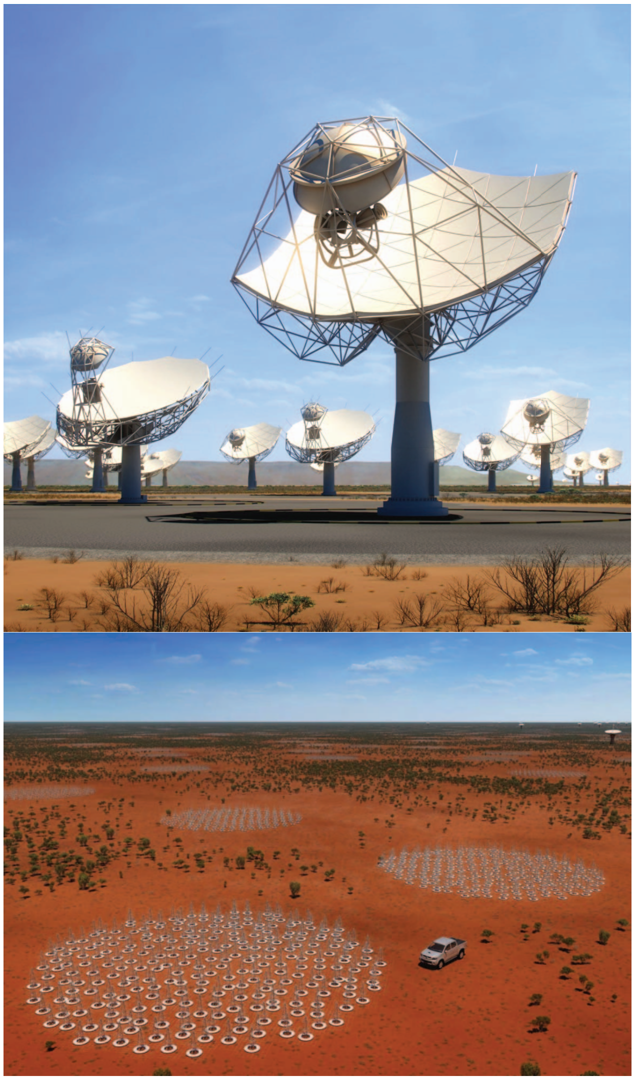

Fig. 1. Top: Dishes for SKA-Mid, in South Africa, will generate 62 exabytes of raw data per year – enough to fill 340,000 average laptops with content every day. Bottom: Aperture arrays for SKA-Low, in Australia, will generate 5 zettabytes of data per year. Unlike conventional dishes, the "beam" of these antennas can be formed electronically – allowing each antenna to look at multiple regions of the sky simultaneously. The total field of view is essentially limited by the signal processing capacity. Image Credits: SKA Organisation.

III. How are we contributing to building the SKA?

We are contributing to various roles in the development of the telescope with a particular focus on the Science Data Processor (SDP) – this is the computing brain that will be responsible for processing these huge quantities of data.

We are currently developing a lightweight end-to-end prototype of the major system components, a project we call the SDP Integration Prototype (SIP; Technical Lead: Ben Mort). We are also undertaking horizontal and vertical prototyping of the essential components that will constitute SDP. For SIP, the aim is to build a mini, fully-operational SDP that runs all relevant services and is scalable to the size of SDP. We are using the Performance Prototype Platform (P3) – a bare metal OpenStack private cloud hardware demonstrator for the SDP.

For prototyping, we require realistic SKA-like science pipelines that can handle large data volumes. We have been testing SIP with data from the LOw Frequency ARray (LOFAR) Multifrequency Snapshot Sky Survey (MSSS; [2]). As a bonus outcome of SIP, the processed data can be used to enable the tightest ever constraints on the magnetised large-scale structure of the Universe.

Acknowledgments

This work is supported by STFC Grants: SKA Preconstruction funding supplement (ST/N003713/1), SKA Preconstruction update (ST/P005446/1) and SKA Preconstruction 2017-18 (ST/R000557/1). It has also received support from the members of the Oxford Astrophysics SKA group, from Prof. Paul Alexander and the Cambridge SKA team, and from group members at both Rutherford Appleton Laboratory and STFC. We are grateful for the contributions throughout the SDP consortium, led from the University of Cambridge, without which this work would not have been possible. LOFAR, designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

References

[1] R. Braun, et al. Advancing Astrophysics with the Square Kilometre Array. Proceedings of Science, AASKA14, 2015.

[2] G. Heald, R. F. Pizzo, E. Orrú, R. P. Breton, D. Carbone, C. Ferrari, et al. "The LOFAR Multifrequency Snapshot Sky Survey (MSSS). I. Survey description and first results," Astronomy & Astrophysics, vol. 582, pp. 22, 2015.