7 JOINT RESEARCH PROJECTS
ASTRONOMY AND BIG DATA

The Square Kilometre Array.

Photograph courtesy of SARAO

South Africa has a long history of excellence in astronomy, a sound high-tech infrastructure and clear skies. It hosts the Southern African Large Telescope (SALT), a 10 m optical telescope and the largest of its kind in the Southern hemisphere, at the South African Astronomical Observatory's Sutherland site. South Africa will also be hosting the majority of dishes of the world's largest radio telescope, the international Square Kilometre Array (SKA) project. The low-frequency array is to be constructed in Australia.

The MeerKAT radio telescope, which was officially inaugurated by the Deputy President of the Republic of South Africa, Mr David Mabuza, on 13 July 2018, is a precursor of the SKA.

Astronomy and Big Data go hand in hand. MeerKAT can, for instance, process up to 275 gigabytes per second. Signal transport and networks will be the backbone of the SKA telescope; they will interface with almost every aspect of the system and will ultimately represent the largest and most challenging network system in science. Approximately 160 Gigabits (10^9) bits per second of data will be transmitted from each radio dish to a central processor, meaning that the high-frequency dishes alone will produce ten times the current global internet traffic. The use of aperture array radio telescopes in the low- and mid-frequency ranges will further increase data rates to many Petabits (10^{15}) per second, which represent more than 10 times the current global internet traffic.

The Centre for High Performance Computing (CHPC) in Cape Town plays an empowering role for SKA partners in other African countries. It started when the CHPC unveiled the fastest computer on the continent in July 2016.

The Netherlands and South Africa have set up a data science partnership to establish national and regional data centres in order to tackle one of the most significant challenges presented by the SKA telescope: how to manage, process, and make accessible the immense amount of data the telescope will generate.

The SKA Africa's Big Data Africa Summer School aims to introduce fundamental data science tools and techniques to talented young science graduates across a range of disciplines, who have an interest to develop their skills and knowledge in working efficiently on extremely large datasets in any research environment (Square Kilometre Array, 2018).

The SKA has been identified as one of the potential large-scale infrastructures in which Switzerland could participate in the coming years, and the country is investigating the best way to participate in the project.

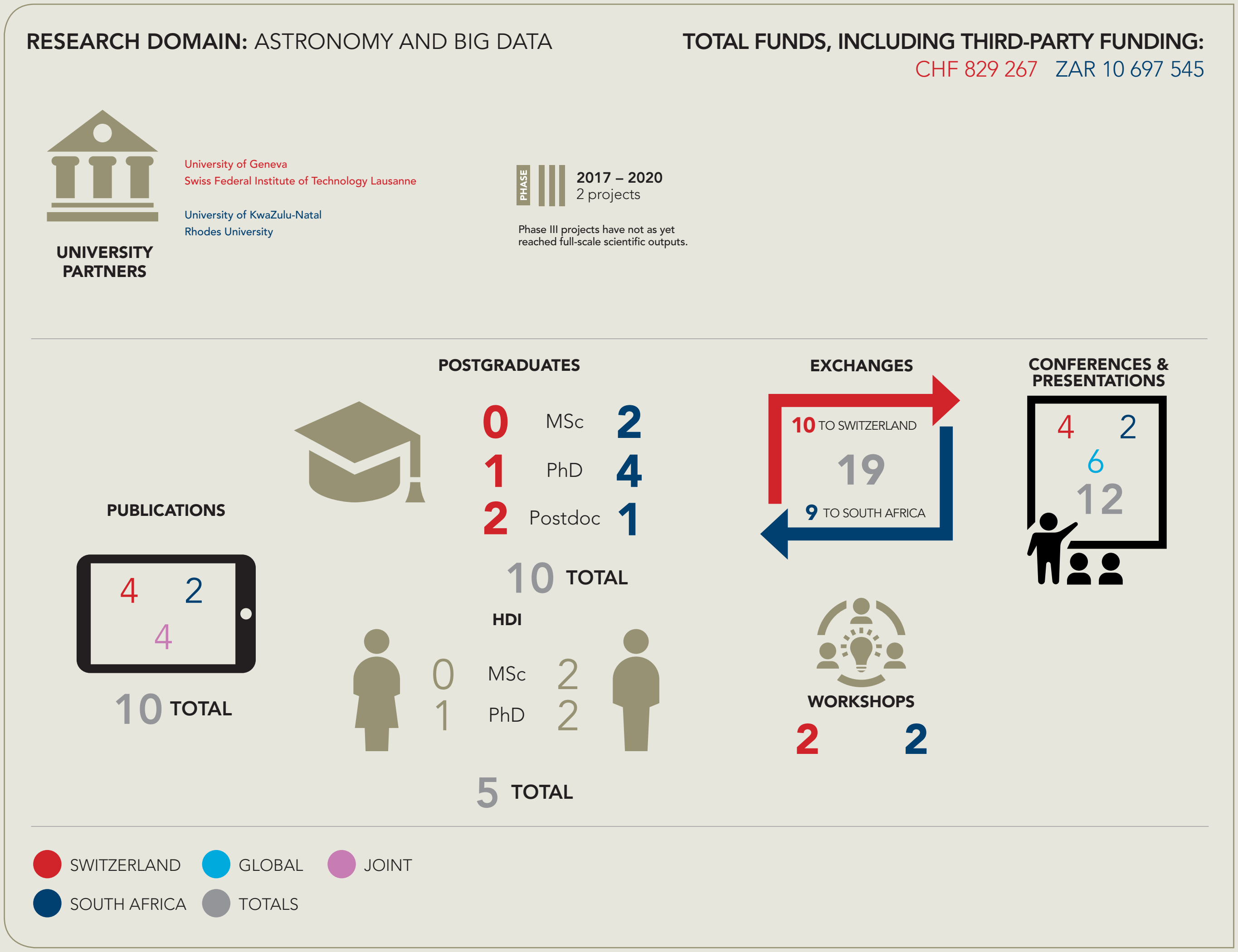
Swiss scientists are not only interested in the observations the SKA will deliver but also in its sheer complexity, which is at the forefront of modern technology. The vast amount of data produced by the SKA will need to be digested and analysed. New hardware, software and advanced algorithms have to be developed to face the expected deluge of data.

Swiss SKA activities are led by a consortium in which more than 50 scientists are participating from different research institutes throughout Switzerland. From an engineering perspective, Switzerland is interested in contributing to work packages of the antennas, central signal processor, science data processor and Big Data, and is studying other packages too.

An example of recent technology that will benefit the SKA is the fast analogic-to-digital converter developed by scientists of IBM and EPFL that will help improve the ultra-fast ethernet network (EPFL, 2016).

But Big Data has uses in many areas other than in astronomy. Big Data is larger, more complex datasets, especially from new data sources. These datasets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address problems such as help spot disease early and develop new medicines; maximise crop yields; predict and respond to natural and man-made disasters; and prevent crime (Oracle and Forbes, 2018).

Outcomes of the Astronomy and Big Data Domain (2 projects)



Addressing the Big Data challenge: developing transferable technologies and methodologies in the astronomy domain

UNIVERSITY PARTNERS

University of Geneva
Dr Martin Kunz
University of KwaZulu-Natal
Dr Kavilan Moodley

The project aims to develop technologies and methodologies to tackle the Big Data challenge in the astronomy domain by involving the investigators in a number of Big Data astronomy projects. These technologies and methodologies will be transferable to other scientific domains dealing with the Big Data challenge. The project is still at a very early stage.

Advances that the research team make towards tackling the Big Data challenges for the relevant astronomy projects will advance the platforms and architectures required to derive maximum scientific value from the data.

Swiss and South African collaborators bring novel opportunities to this project; for example, the involvement in the HIRAX, BINGO, MeerKAT, SKA and DES plus LSST projects (see glossary for descriptions).

The combined unique skills of the investigators in, for example, data analysis, computing architectures and machine learning, will bring excellent scientific value to the astronomy and broader community from this collaboration.

Capacity building has already occurred: students visited and worked at the ETH Zurich laboratory on drone technology for radio telescope calibration, and a postdoctoral fellow from UKZN visited the ETH Zurich group to work on models for the distribution of neutral gas in galaxies at high redshift, and on machine learning for radio frequency interference detection.

Photograph courtesy of Kavilan Moodley

Dr Kavilan Moodley

Photograph courtesy of Yasmin Hankel

Dr Martin Kunz

Members of the Astrophysics Research Centre at the University of KwaZulu-Natal, inspecting the PRIZM antenna at the astronomy instrumentation laboratory.

TIMEFRAME
2017 – 2020

Research Domain
Big Data
Cross-cutting Domain
Astronomy

TOTAL PROJECT BUDGET
SSAJRP
CHF 547 655
ZAR 7 064 745

POSTGRADUATES

Switzerland		South Africa
0	MSc	1
1	PhD	1
1	Postdoc	1

HDI

0	MSc	1
0	PhD	1

EXCHANGES
2 TO SWITZERLAND
3
1 TO SOUTH AFRICA

SWITZERLAND GLOBAL JOINT
SOUTH AFRICA TOTALS

Wide-band imaging in the SKA era

UNIVERSITY PARTNERS

Swiss Federal Institute of Technology Lausanne
Professor Jean-Philippe Thiran
Rhodes University
Professor Oleg Smirnov

Radio interferometry (RI) allows the observation of radio emissions of the universe with great sensitivity and angular resolution. Aperture synthesis in RI correlates electric signals from pairs of antennae to produce the so-called visibilities which, under the simplifying assumptions of non-polarised monochromatic incoherent radiation on a small field of view, provide an incomplete Fourier sampling of the underlying two-dimensional sky brightness image of interest.

New radio telescopes, such as the future flagship SKA, are intended to provide images at a totally new range of resolutions and sensitivities, and on a wide frequency band. Data rate estimates for the first phase of development of the telescope only, are around a few terabytes per second. The massive amounts of data to be acquired will represent a great challenge for the infrastructure and signal processing, and the methods solving the inverse problem associated with the image reconstruction need to be fast and to scale well with the data volumes. The celebrated CLEAN imaging algorithm and its variants that have driven RI imaging so far, will simply not scale to the SKA Big Data regime.

The project aims to tackle this RI imaging challenge in the wide-band setting – when a third imaging dimension is introduced to account for observation over a whole radio frequency band – by leveraging modern parallelised and distributed algorithmic structures in convex optimisation.

The researchers aim to define a new advanced wide-band signal model relying on low-rankness and average joint-sparsity of a matricisation of the wide-band image cube of interest. A convex optimisation problem will be formulated that encompasses multiple data fidelity terms and regularisation priors accounting for the signal model.

They will leverage a recently proposed primal-dual algorithmic structure to solve the optimisation problem, presenting the extreme advantage of full splitting of the objective function: the algorithm can split not only the data but also the image into a large number of blocks that can be processed in parallel at each iteration of the image reconstruction process. It is also shipped with a very powerful randomisation functionality, which enables a random selection of the data and image

Photograph courtesy of SARAO

blocks to be updated at each iteration for lower computational requirement.

They will adapt the same algorithmic structure for Faraday depth synthesis when wide-band polarimetric data are considered. The team will perform algorithm validation with real data from the South African SKA precursor MeerKAT, of which first dishes came online simultaneously with the start of this project, as well as with simulated SKA data. They will also develop a high-performance computing (HPC) implementation.