SKA SA: Enabling Serendipity

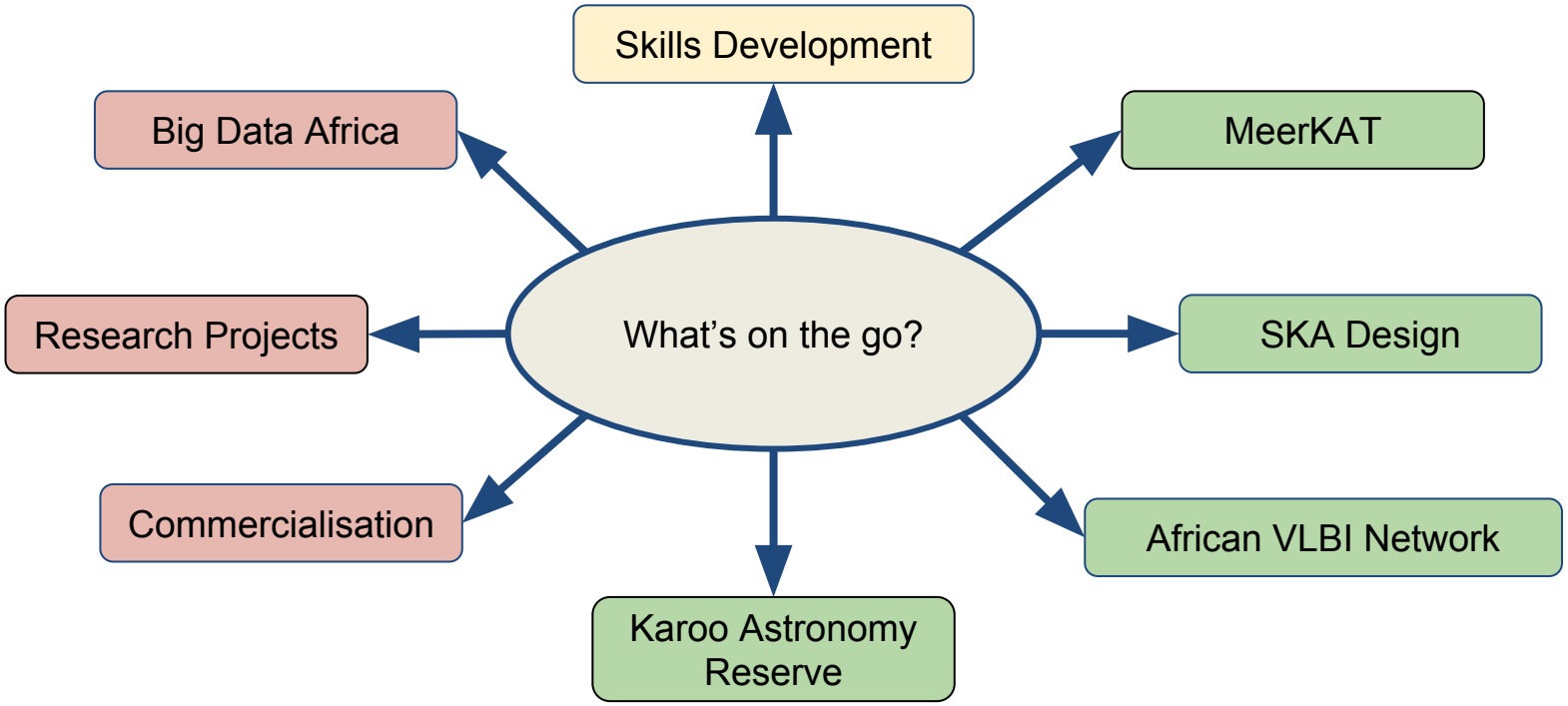
www.ska.ac.za

SKA AFRICA
SQUARE KILOMETRE ARRAY

Jasper Horrell (and the SKA SA Team)

jasper@ska.ac.za

SKA SA - Programmes



SKA AFRICA
SQUARE KILOMETRE ARRAY

KAT-7 (MeerKAT precursor - operational since 2009)

Image: Maik Wolleben

SKA AFRICA
SQUARE KILOMETRE ARRAY

MeerKAT (64-dish SKA precursor - under construction)

Image: SKA SA

SKA AFRICA
SQUARE KILOMETRE ARRAY

MeerKAT (64-dish SKA precursor - under construction)

Image: SKA SA

SKA AFRICA
SQUARE KILOMETRE ARRAY

MeerKAT (64-dish SKA precursor - under construction)

Image: SKA SA

SKA AFRICA
SQUARE KILOMETRE ARRAY

MeerKAT (64-dish SKA precursor - under construction)

Image: SKA SA

SKA2-Mid

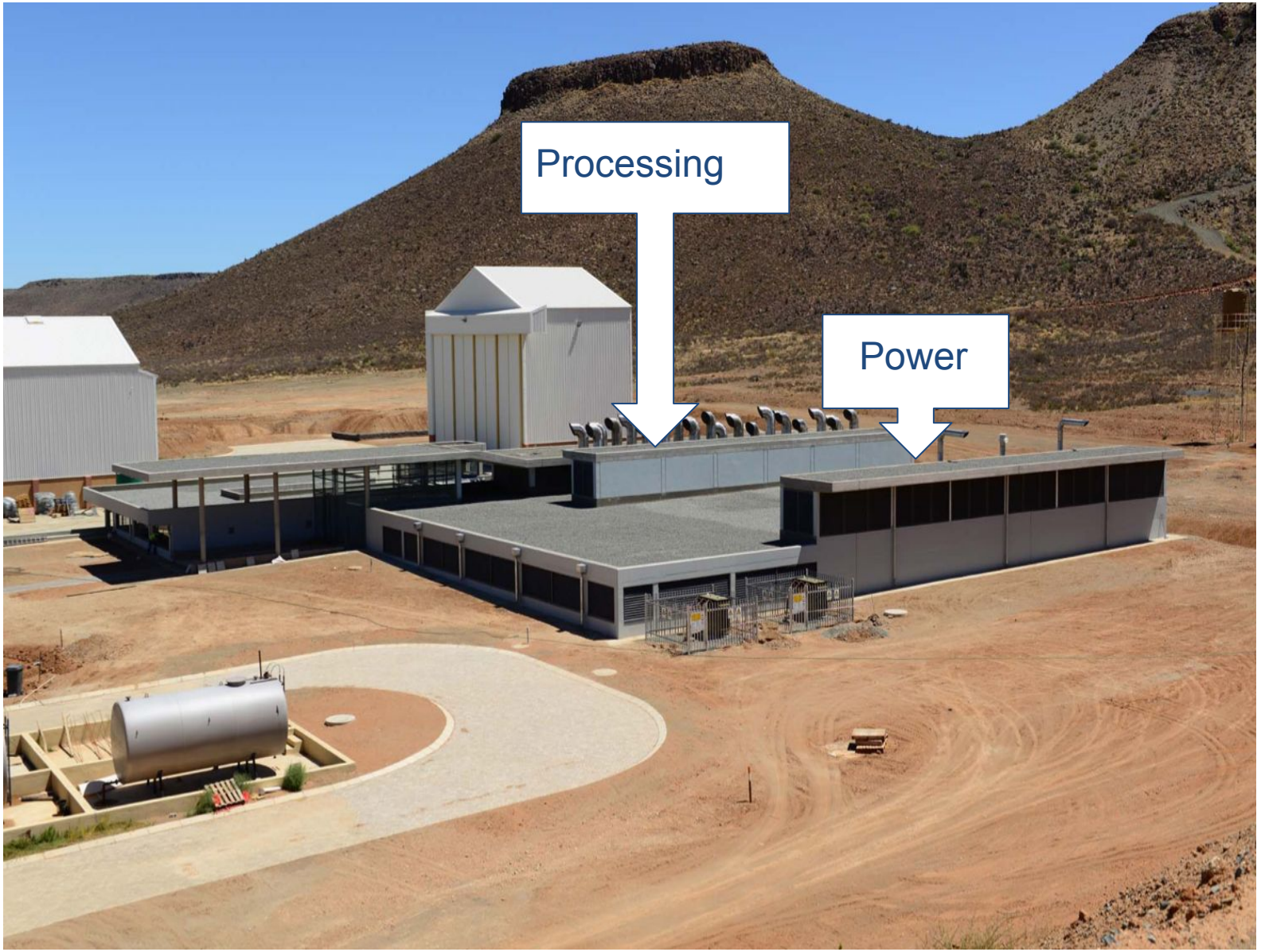
Image: SKA Organization /
Swinburne Astronomy Productions

SKA Stations in Africa - Phase II

SKA AFRICA
SQUARE KILOMETRE ARRAY

Image: SKA SA

Losberg Site Complex - Karoo Array Processor Building



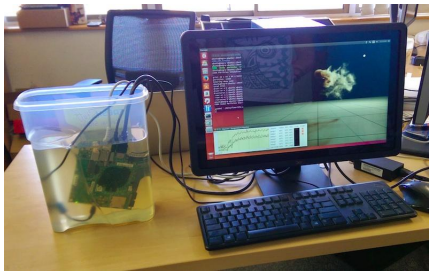
Shielded Karoo Data Centre (when still empty)

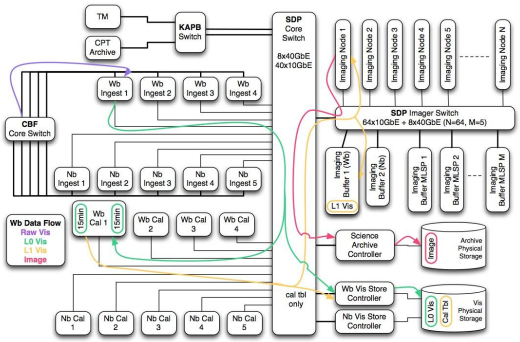
Capacity:
140 racks
few MW

Shielded Karoo Data Centre

140 racks
few MW

Africa + SKA => Technical Trigger





Advanced statistics

Model fitting

Machine learning

Automation / pipelines

Roach

Microservers

Research cloud

Telescope Manager

HPC

Data driven systems

Hardware / software

SKA AFRICA
SQUARE KILOMETRE ARRAY

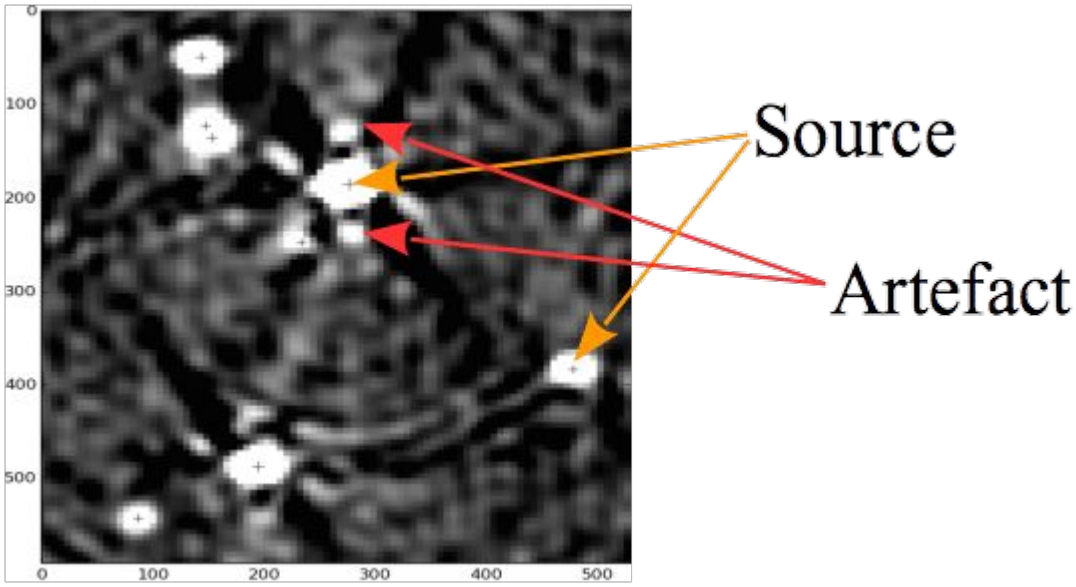
SKA SA Active Machine Learning Projects

- Source / artifact classification in radio images
- Morphological classification of radio galaxies / objects
- Pulsar candidate selection
- Radio transient detection
- RFI flagging
- Mining relations in multi-wavelength datasets
- Serendipity Machine
- + SETI ?

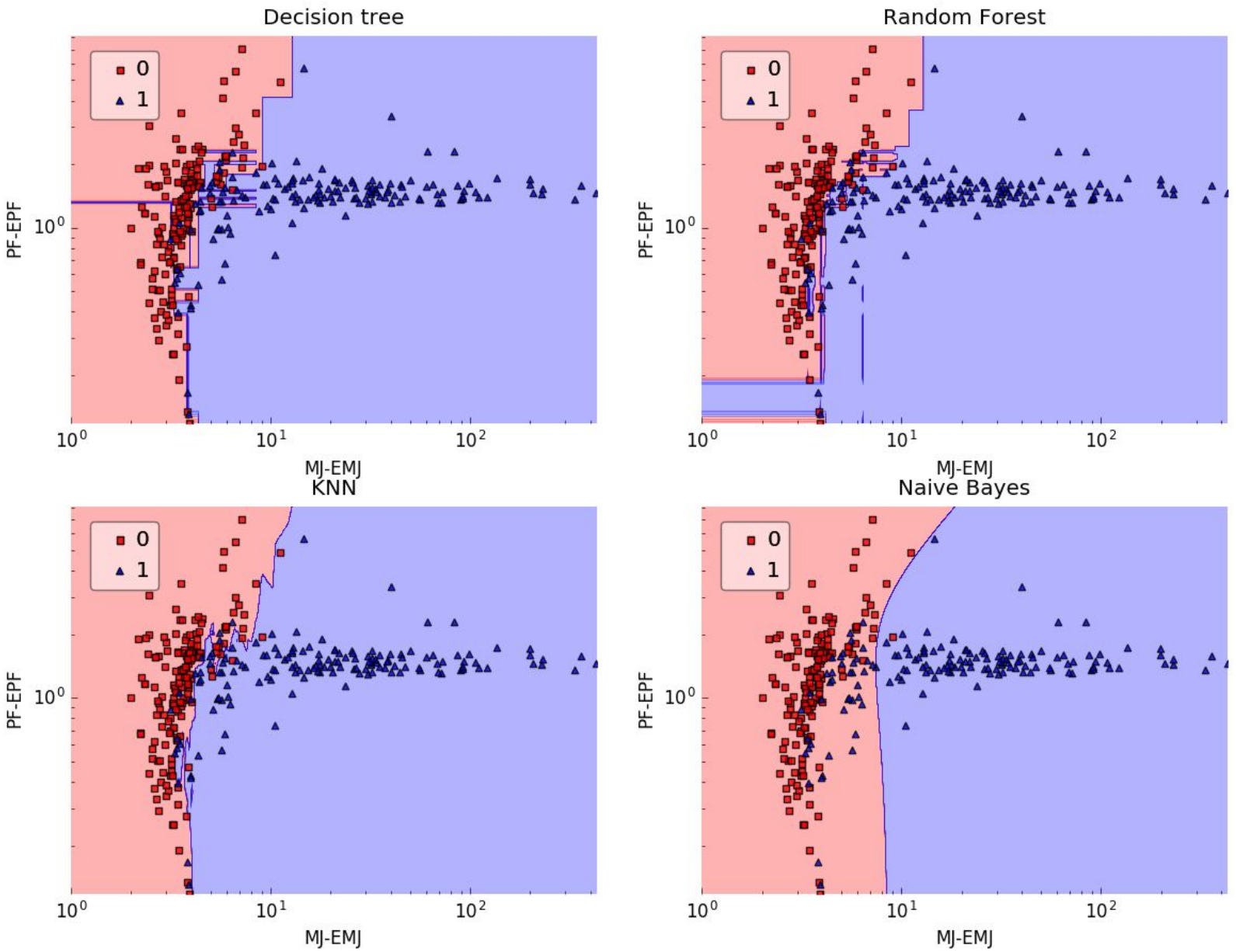
Arun Aniyan, Christine de Kock, Olorato Mosiane, Nadeem Oozeer, Jasper Horrell, etc.

Source / Artifact Classification (problem)

Arun Aniyan

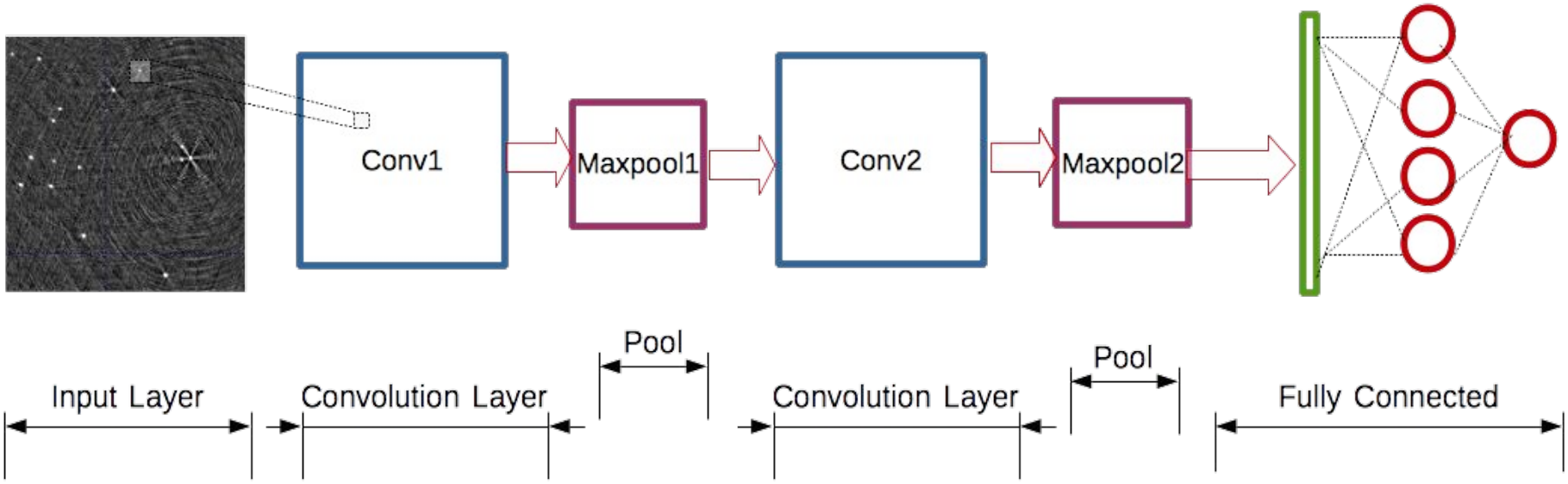


Source / Artifact Classification (features - 95.6%)

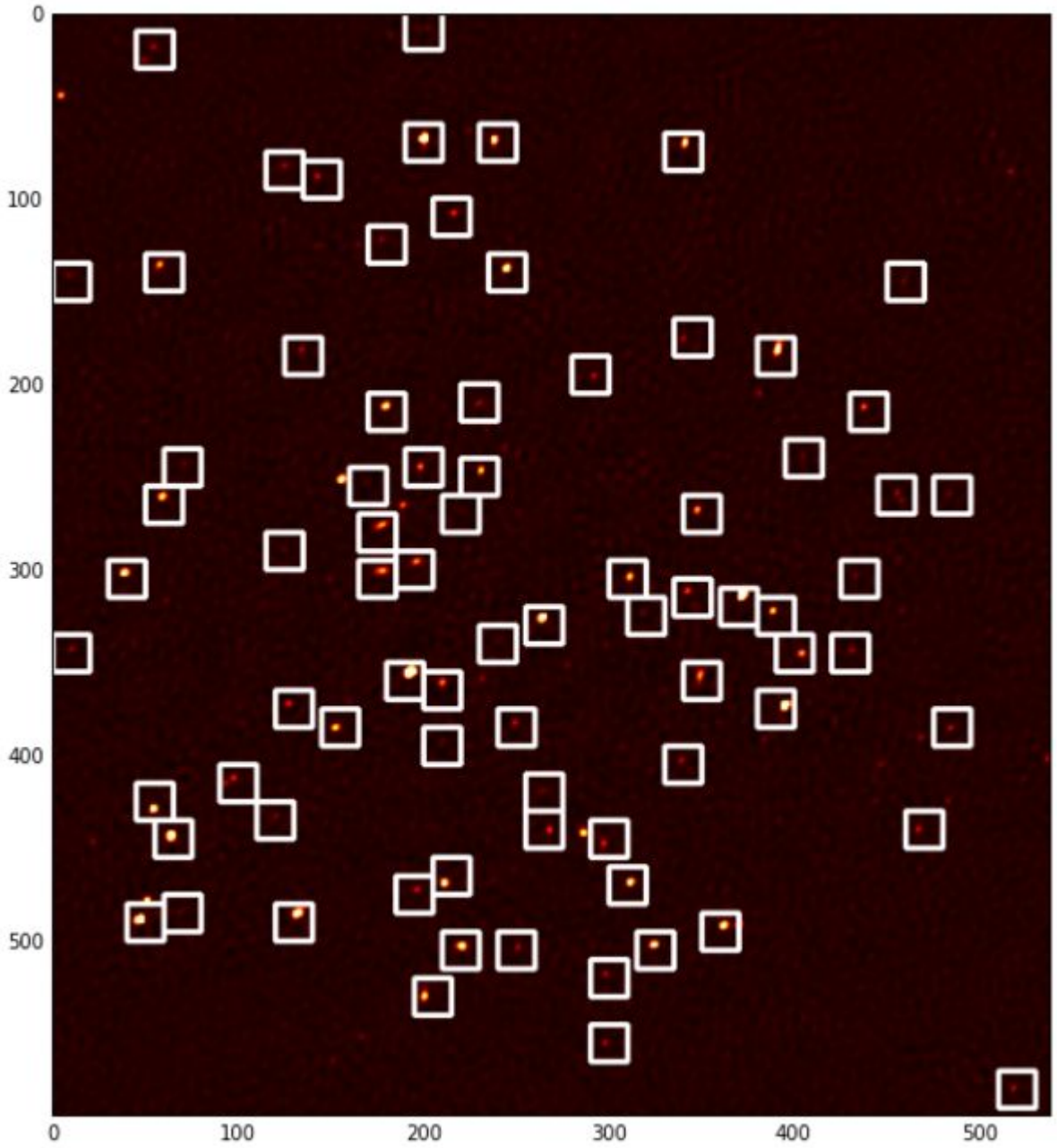


Source / Artifact Classification (CNN - 98%)

Arun Aniyan



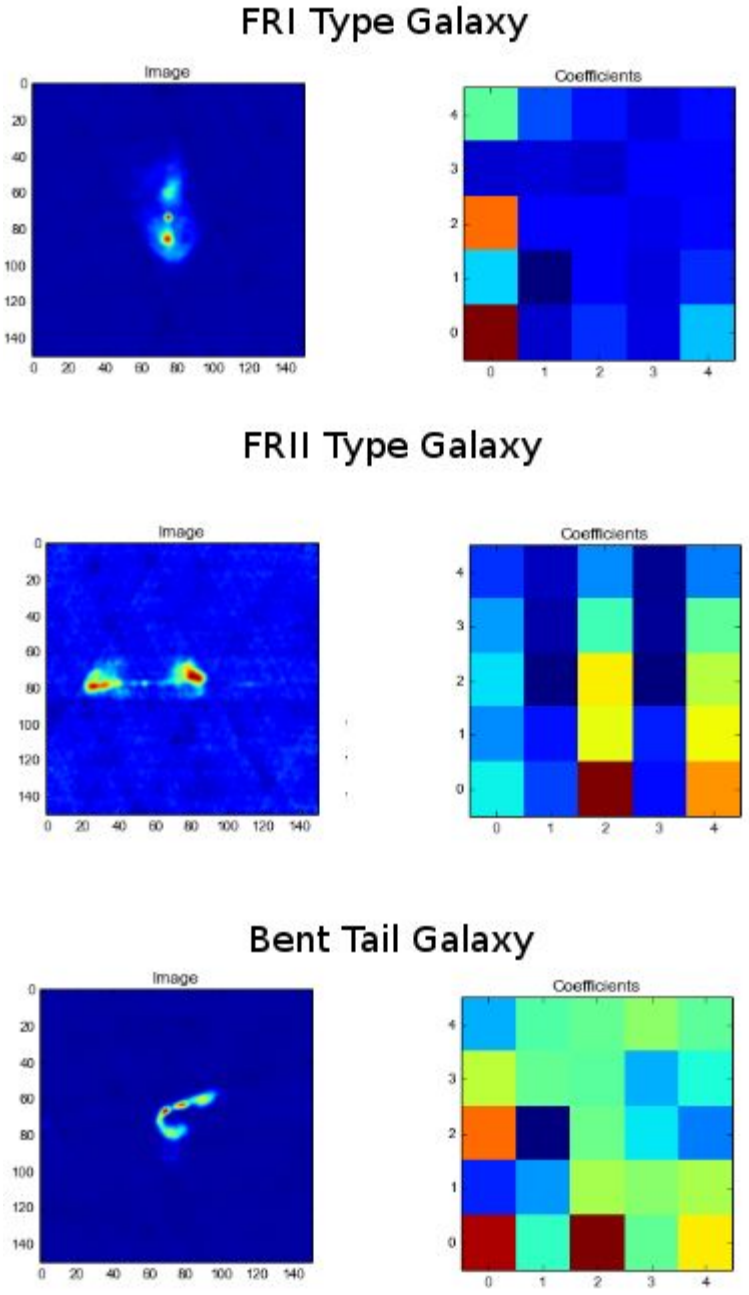
Source / Artifact Classification (CNN - MeerKAT)



Galaxy Morphology

Arun Aniyan

Shapelet and Deep Learning approaches under development



Towards Serendipity Machine

Looking into:

Local Outlier Factor
One-Class SVM
Self-Organising Maps

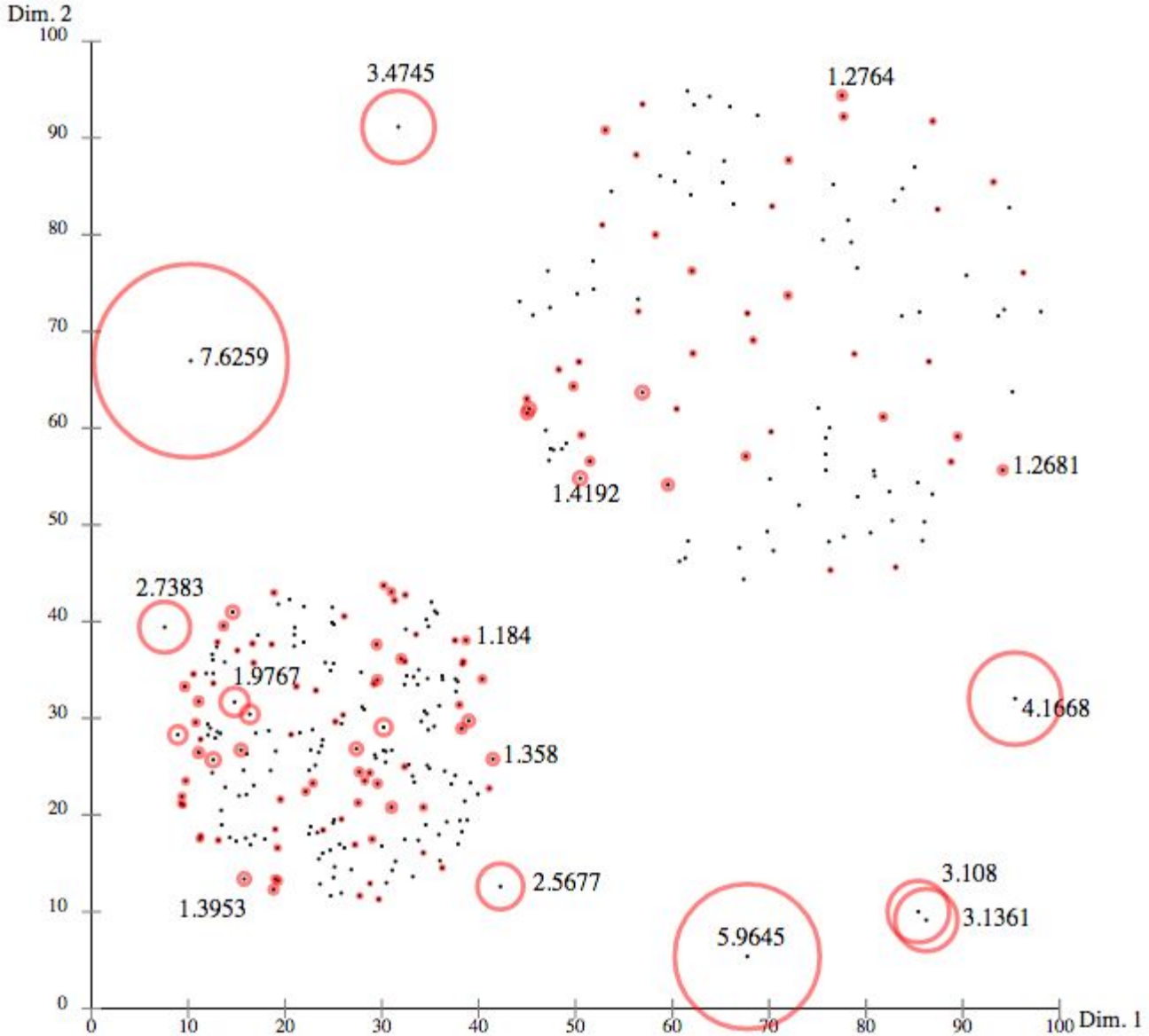
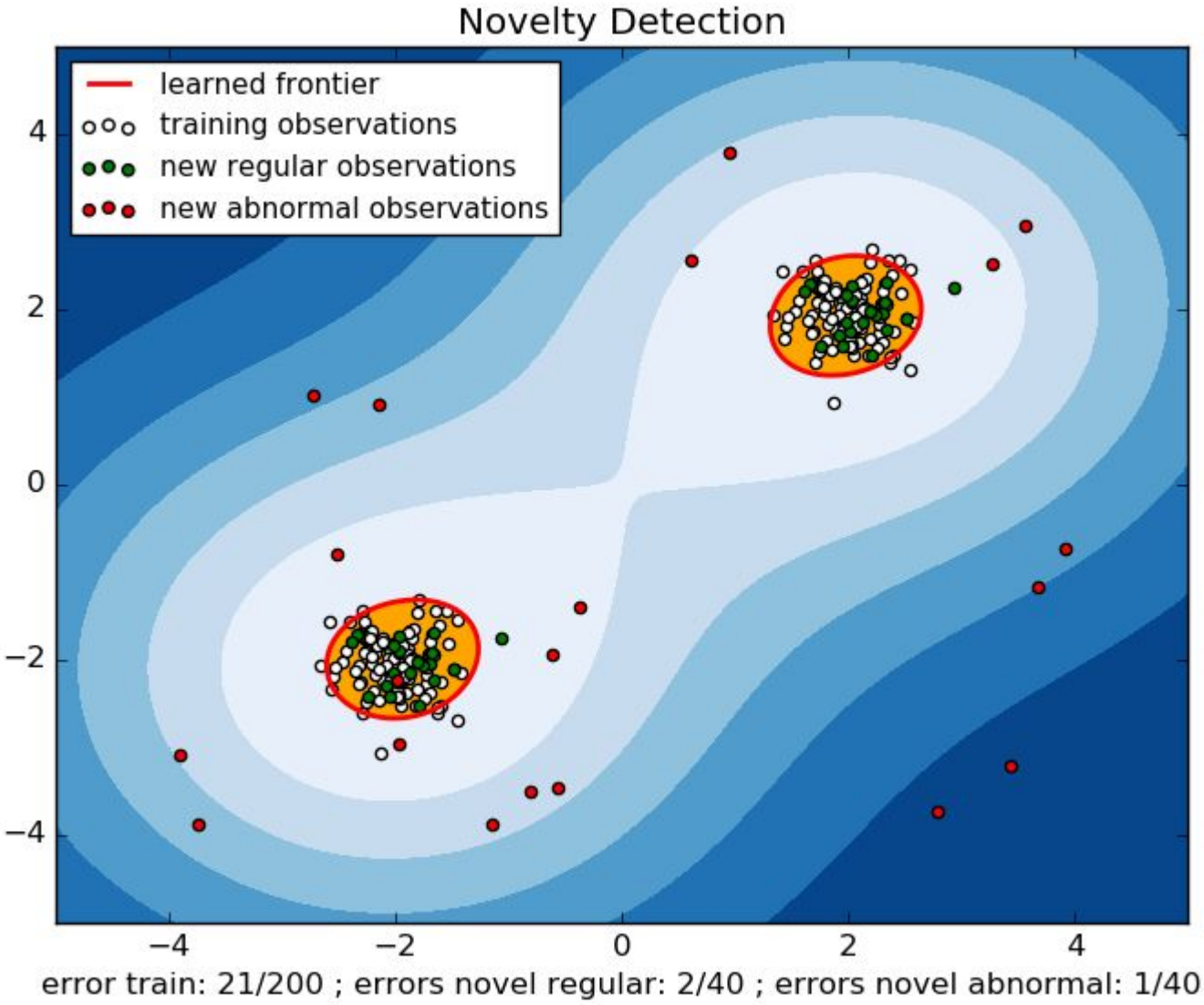


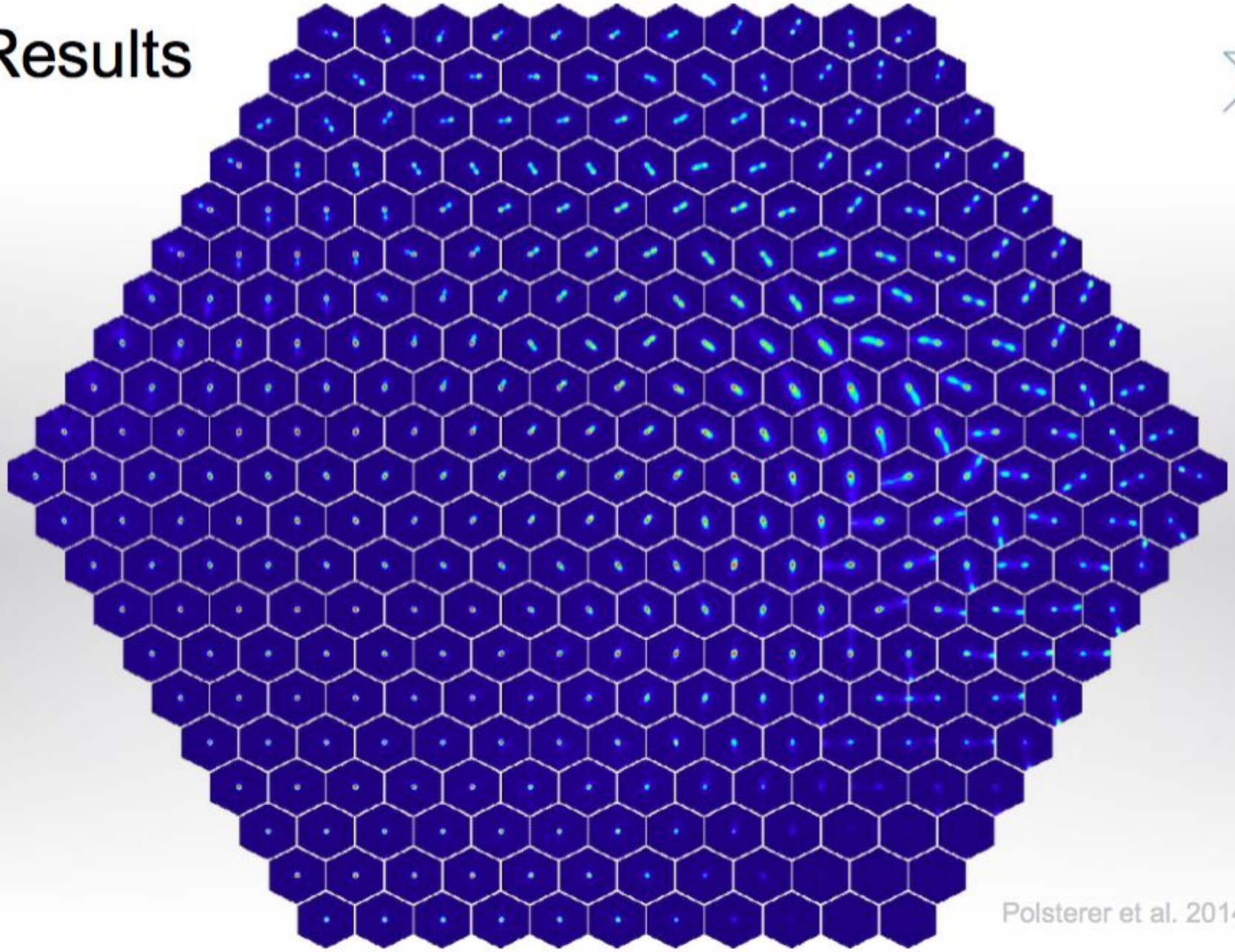
Image: https://en.wikipedia.org/wiki/Local_outlier_factor

One-Class SVM



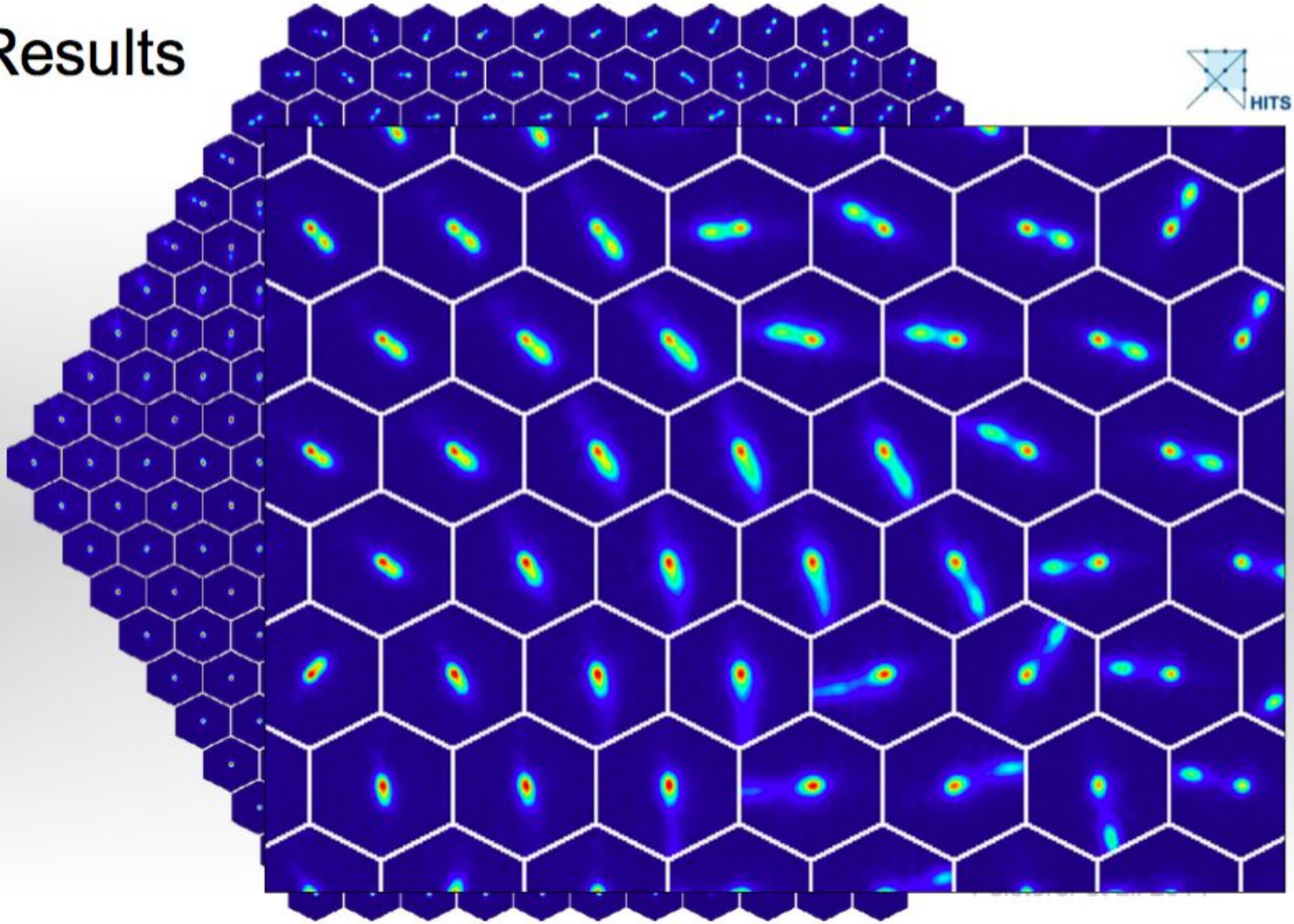
From scikit-learn examples

Results

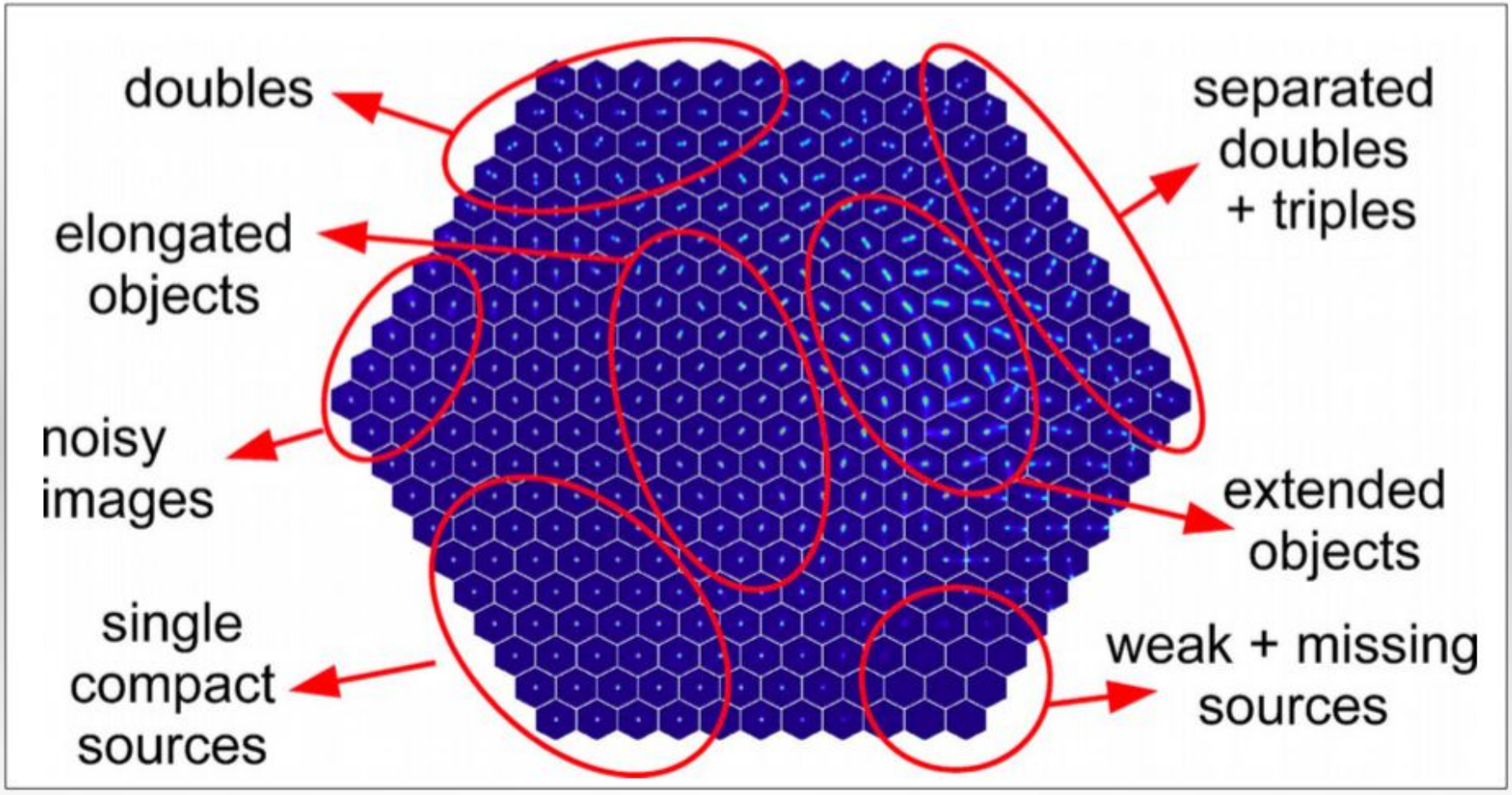


Polsterer et al. 2014

Results



Results


doubles
elongated objects
noisy images
single compact sources
separated doubles + triples
extended objects
weak + missing sources

Conclusions

- MeerKAT coming online - very sensitive data stream (L-band, UHF, S-Band)
- Expected to enable many new discoveries
- Pluggable data architecture
- Lots of interest in machine learning approaches in SKA SA
- Starting to be applied widely across different kinds of signals / analysis
- Serendipity Machine in the long run....
- Taking a step-by-step, machine assisted discovery approach

Thank you (and to the SKA SA team)

jasper@ska.ac.za

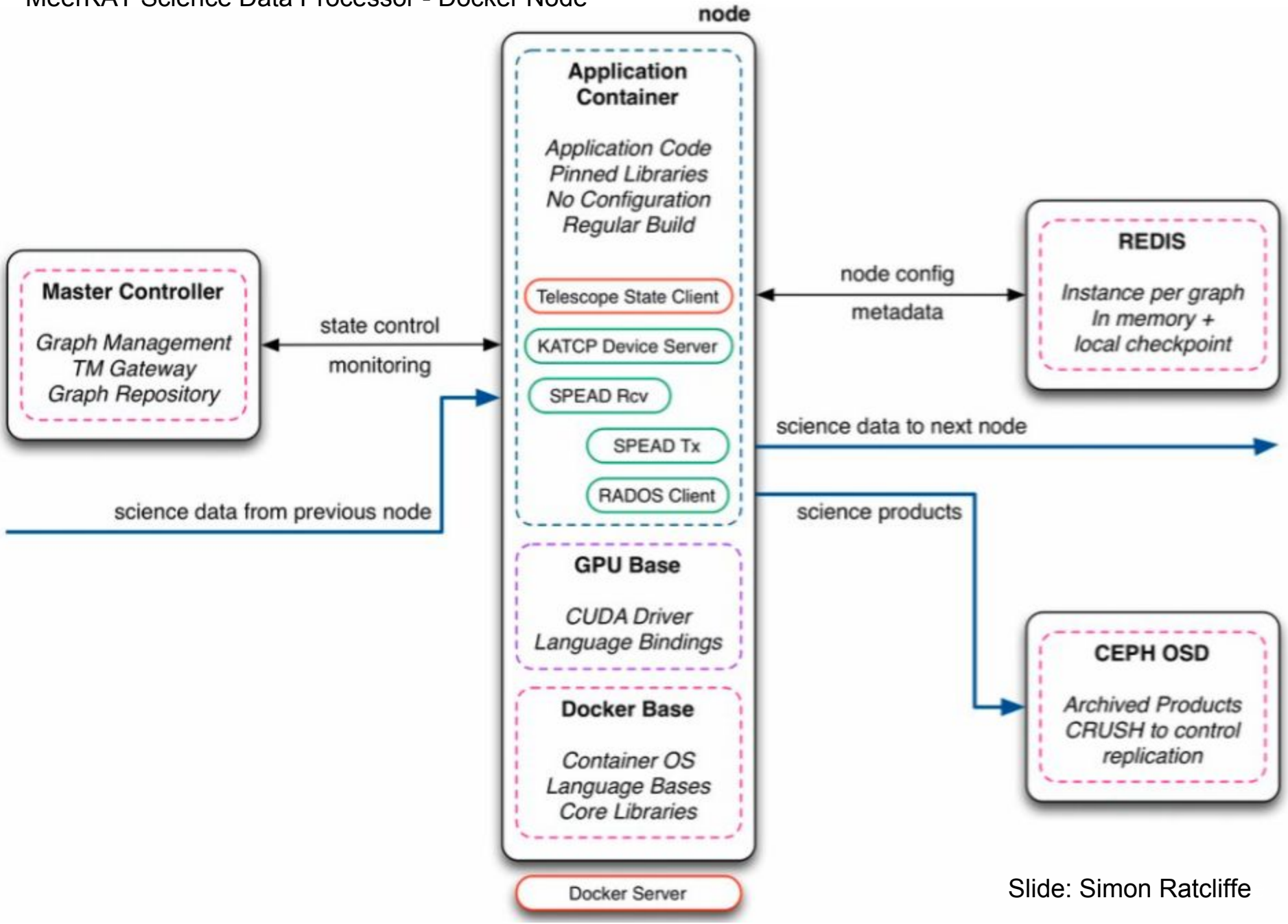
SKA AFRICA
SQUARE KILOMETRE ARRAY

Additional Slides

SKA SA

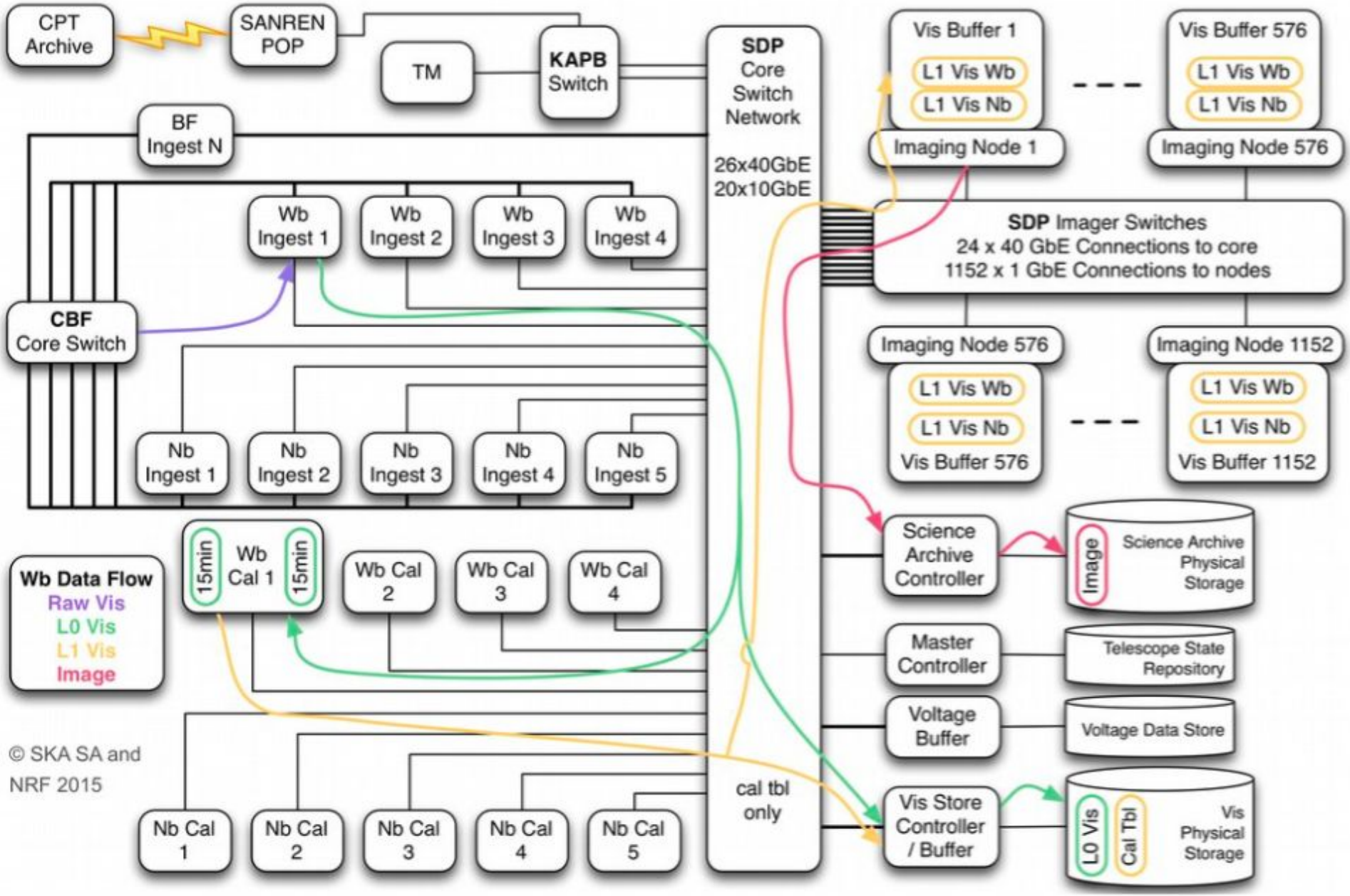
- The South African organisation responsible for hosting of the SKA in Africa (part of National Research Foundation and funded by Dept of Science and Technology)
- Building and operating the MeerKAT 64-dish SKA precursor telescope (operations from 2017)
- Owner and responsible for the SKA core site in Karoo region
- Responsible for large skills development programme
- Responsible for SA's involvement in SKA Design Phase
- Responsible for African VLBI Network project
- Driving research projects, particularly Big Data, machine learning, advanced and power efficient computing.
- www.ska.ac.za

MeerKAT Science Data Processor - Docker Node



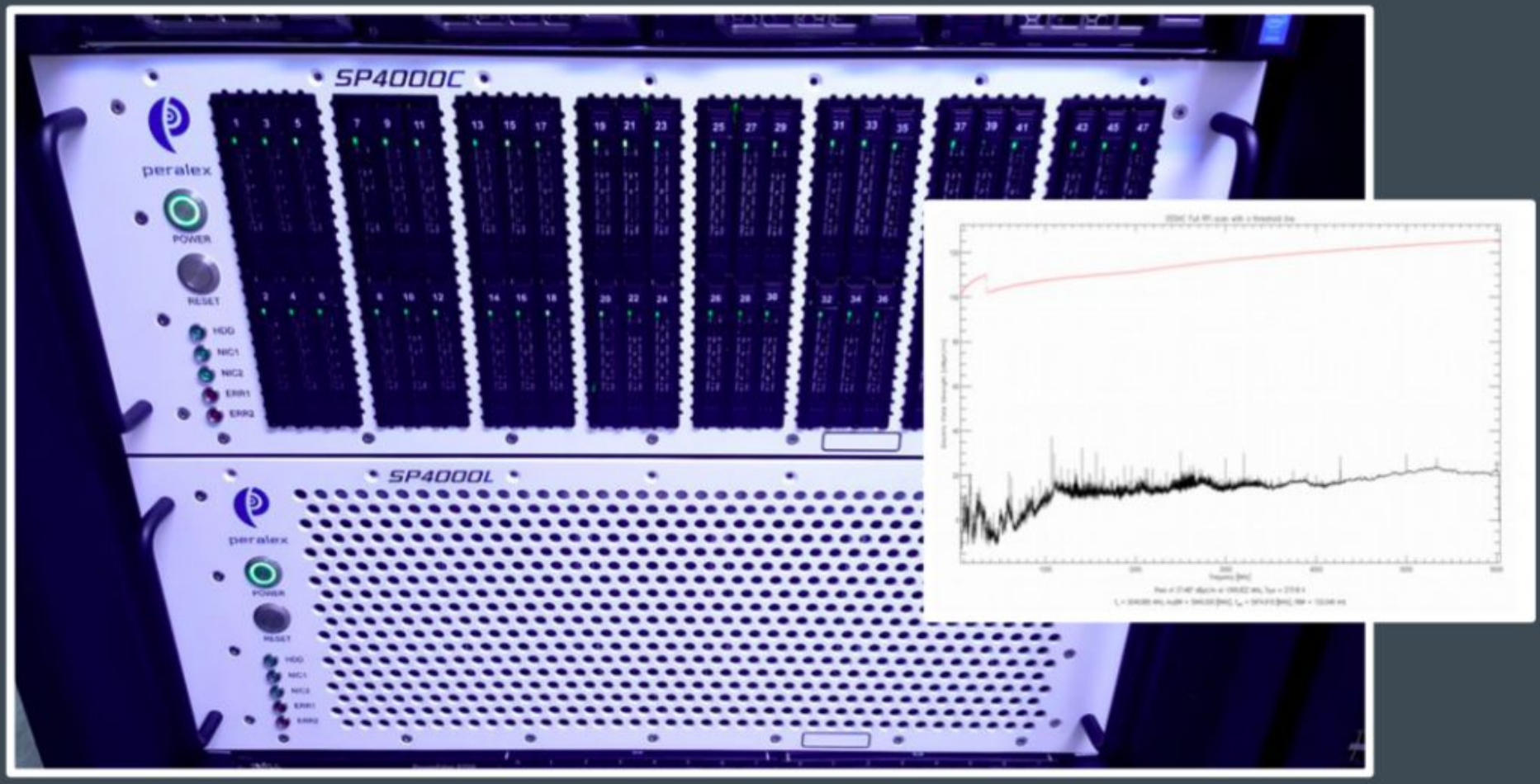
Slide: Simon Ratcliffe

MeerKAT Science Data Processor - Physical / Flows



Slide: Simon Ratcliffe

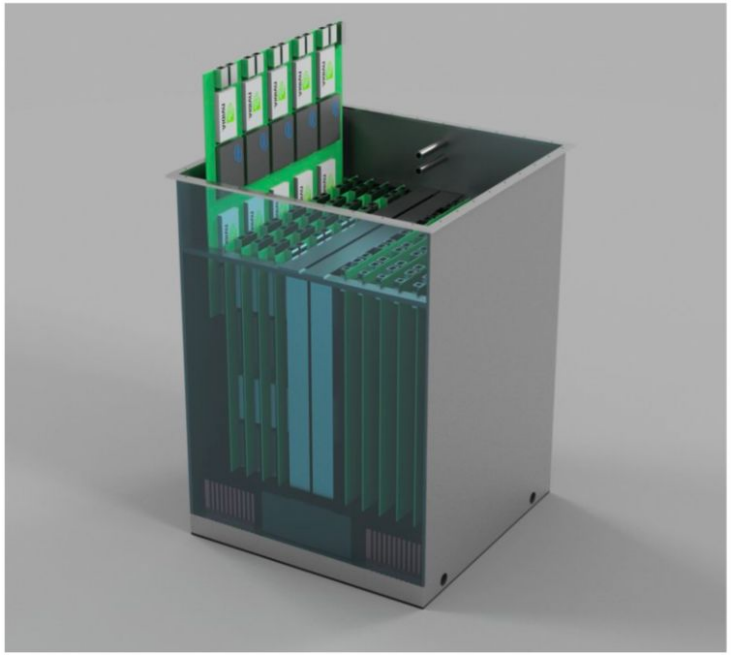
Just finished qualification....



High Speed Pod: 40 Gbps to disk / 40 TiB

Bulk Pod: 10+Gbps to disk / 360 TiB

Slide: Simon Ratcliffe



Same numbers as last year :)

TEGRA X1

1000 Nodes
Tegra X1
4 GB RAM
512 GB SSD

20 Switches
2 x 10 GbE SFP+
48 x 1 GbE

10 Pods
15M Ground Loop
100L Mineral Oil

Titan X

22 Servers
4 x Titan X
2 x E5-2660v3
12 x 2TB SATA
128 GB RAM

3 Switches
4 x 40 GbE QSFP
36 x 10 GbE SFP+

3 Racks
Just a rack

Not quite the same numbers as last year

TEGRA X1

$449 kilo
$426k Hardware
$23k Infrastructure

18.4 kW
17.9 kW Hardware
0.5 kW Cooling

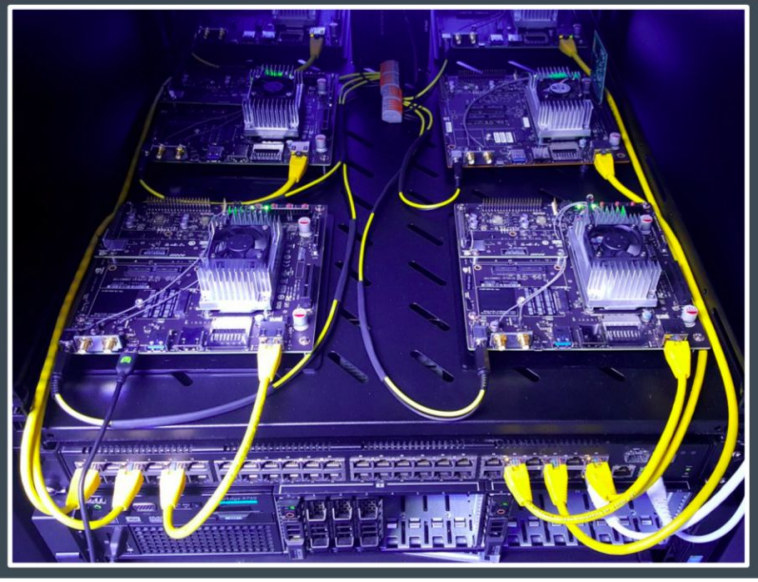
Titan X

$441 kilo
$322k Hardware
$118k Infrastructure

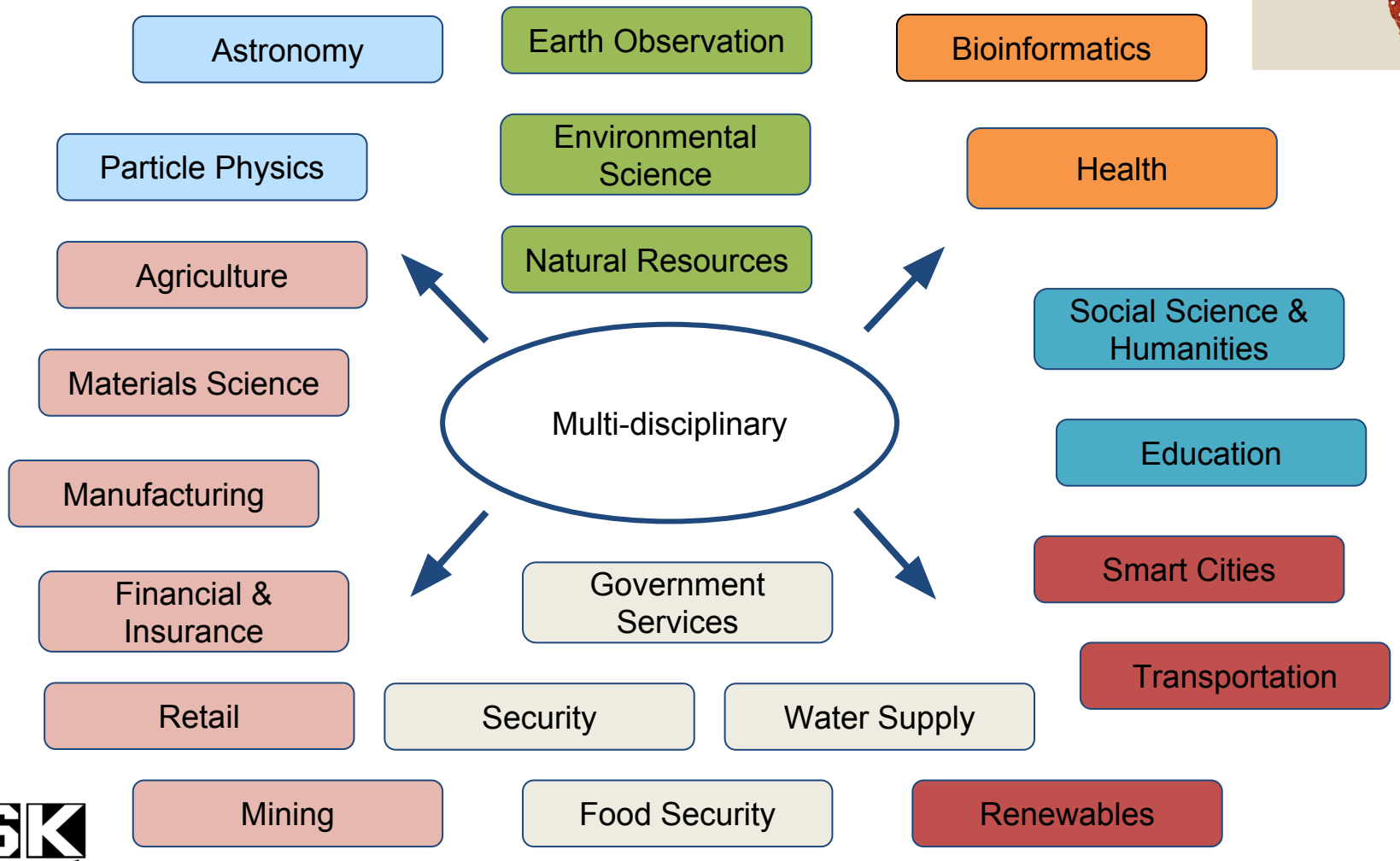
38.2 kW
29.4 kW Hardware
8.8 kW Cooling

Slides: Simon Ratcliffe

Software Stack Testbed



Big Data in Africa => Multi-disciplinary



SKA AFRICA
SQUARE KILOMETRE ARRAY