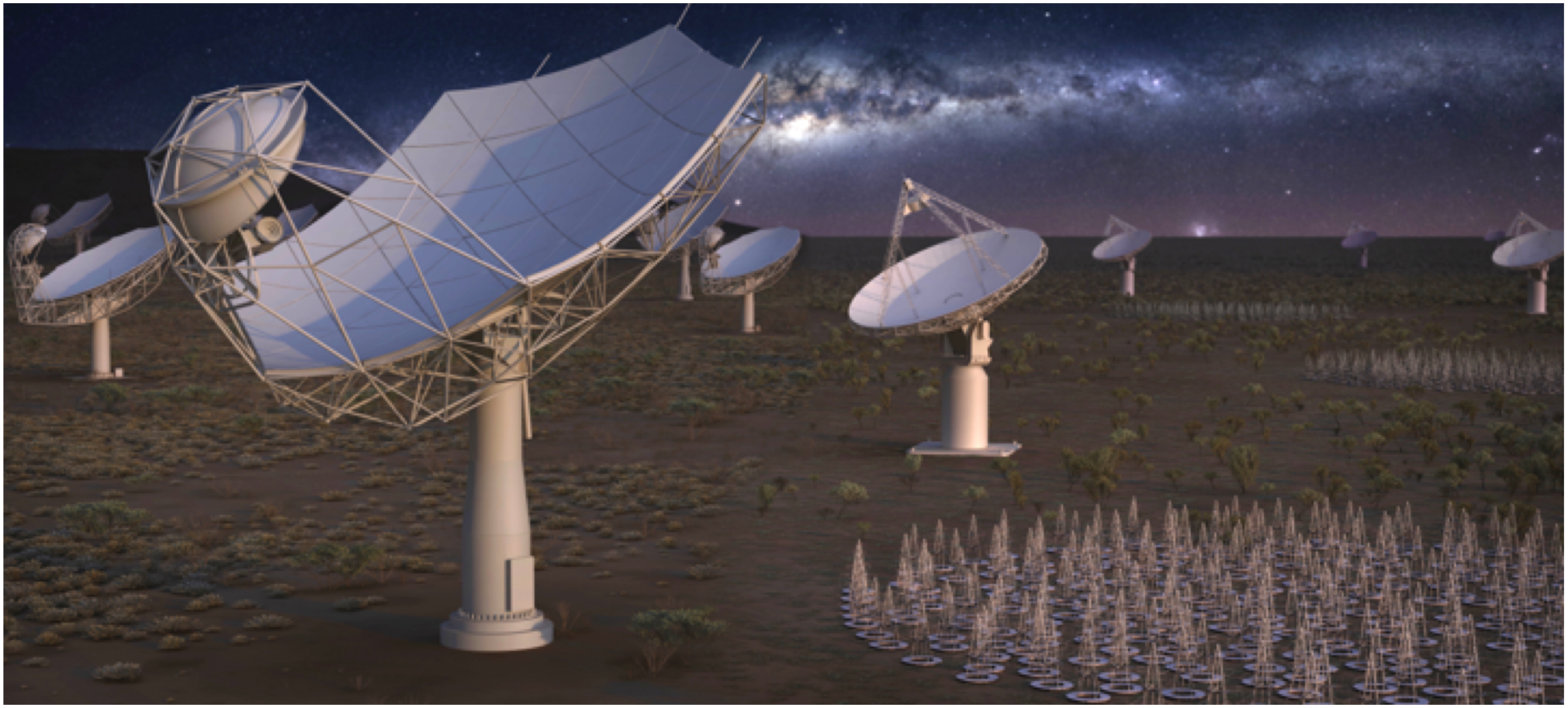
Square Kilometre Array

SKA and its data processing

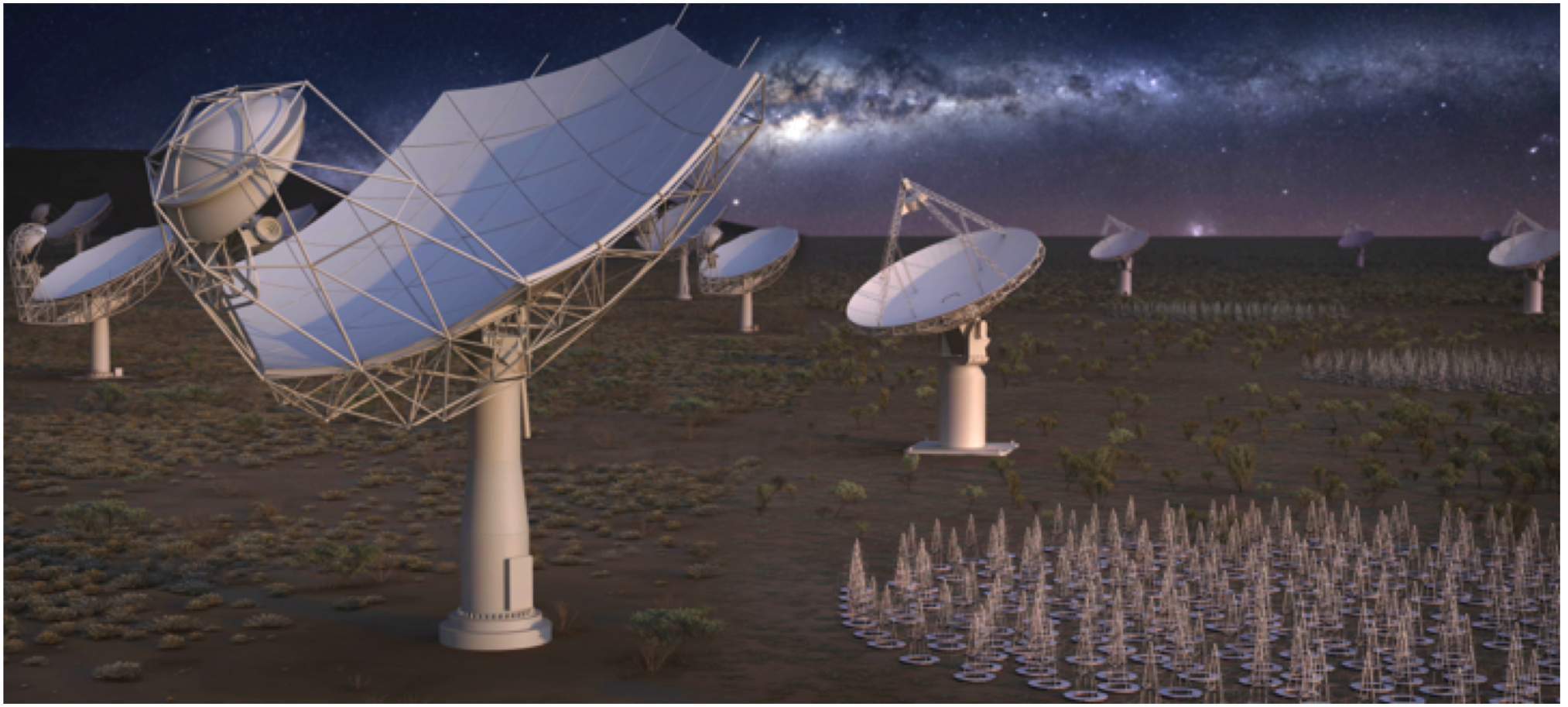


SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Square Kilometre Array

Another big telescope project with a data problem!



SQUARE KILOMETRE ARRAY

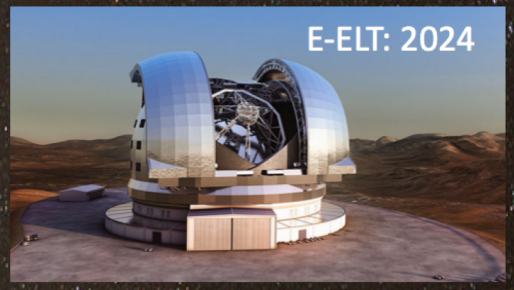
Exploring the Universe with the world's largest radio telescope

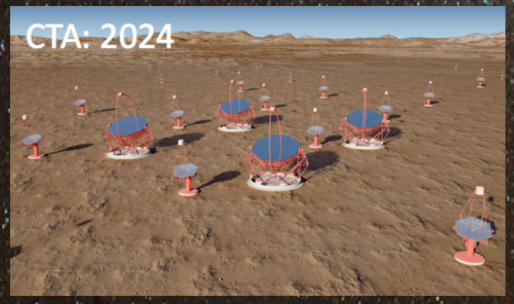
SQUARE KILOMETRE ARRAY

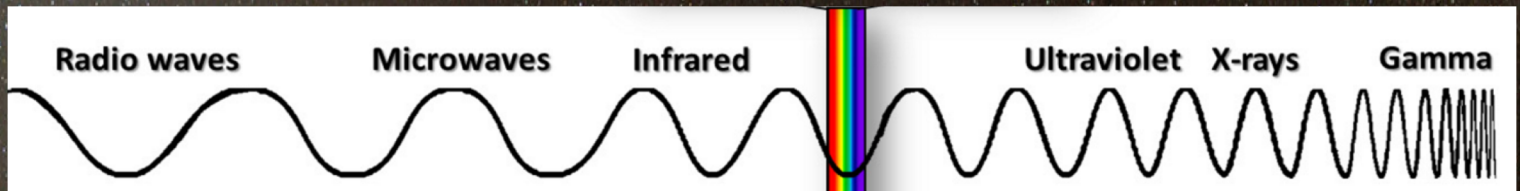
Exploring the Universe with the world's largest radio telescope

What?

Observatories and telescopes for the 21st Century



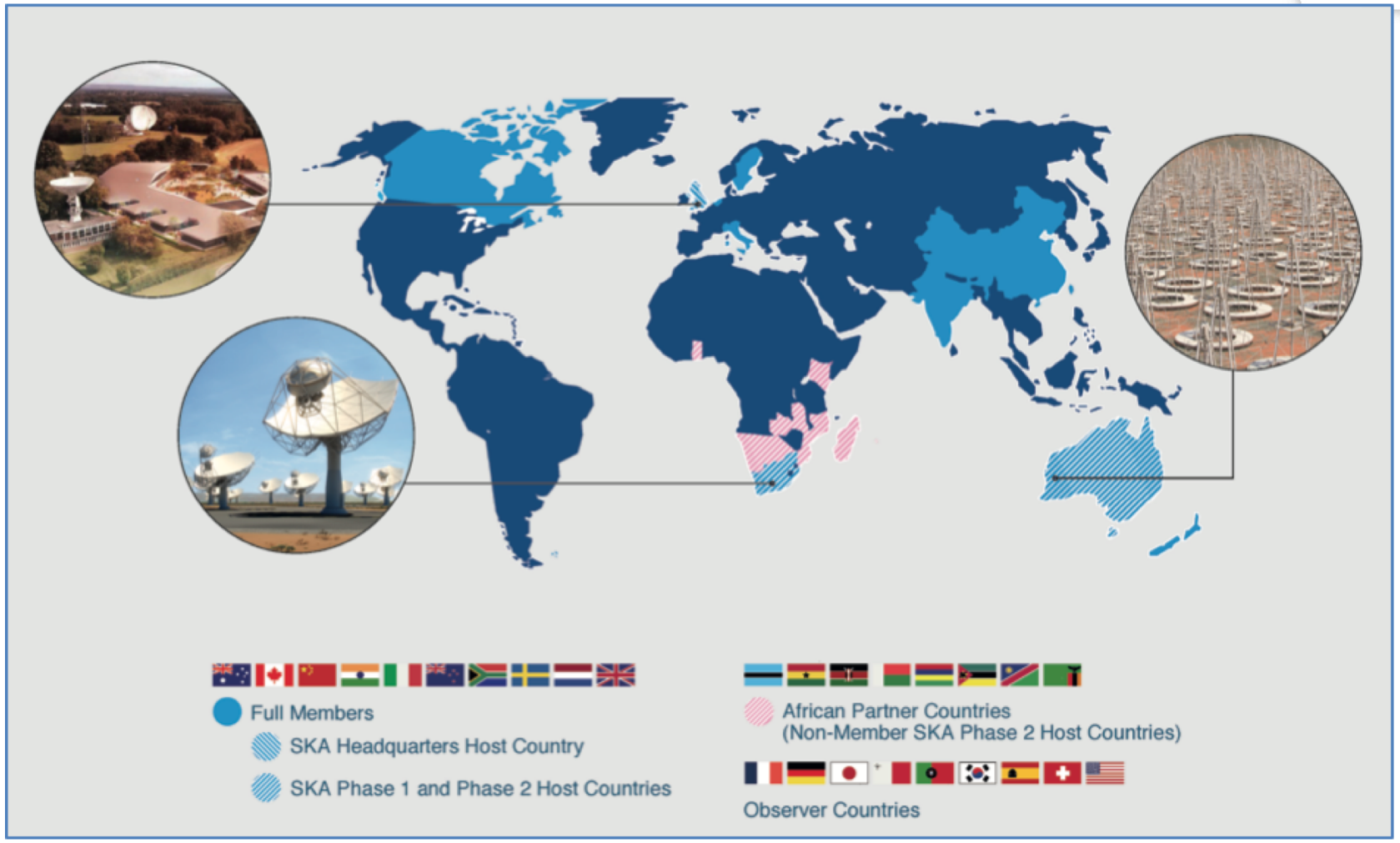




Observatories and telescopes for the 21st Century

SKA:
50 year, Billion-dollar+ project to build the World's largest Radio Observatory
Two ground breaking interferometers

SKA's Global footprint



10 member countries: AUS, CA, CH, IN, NZ, RSA, SWE, NL, UK

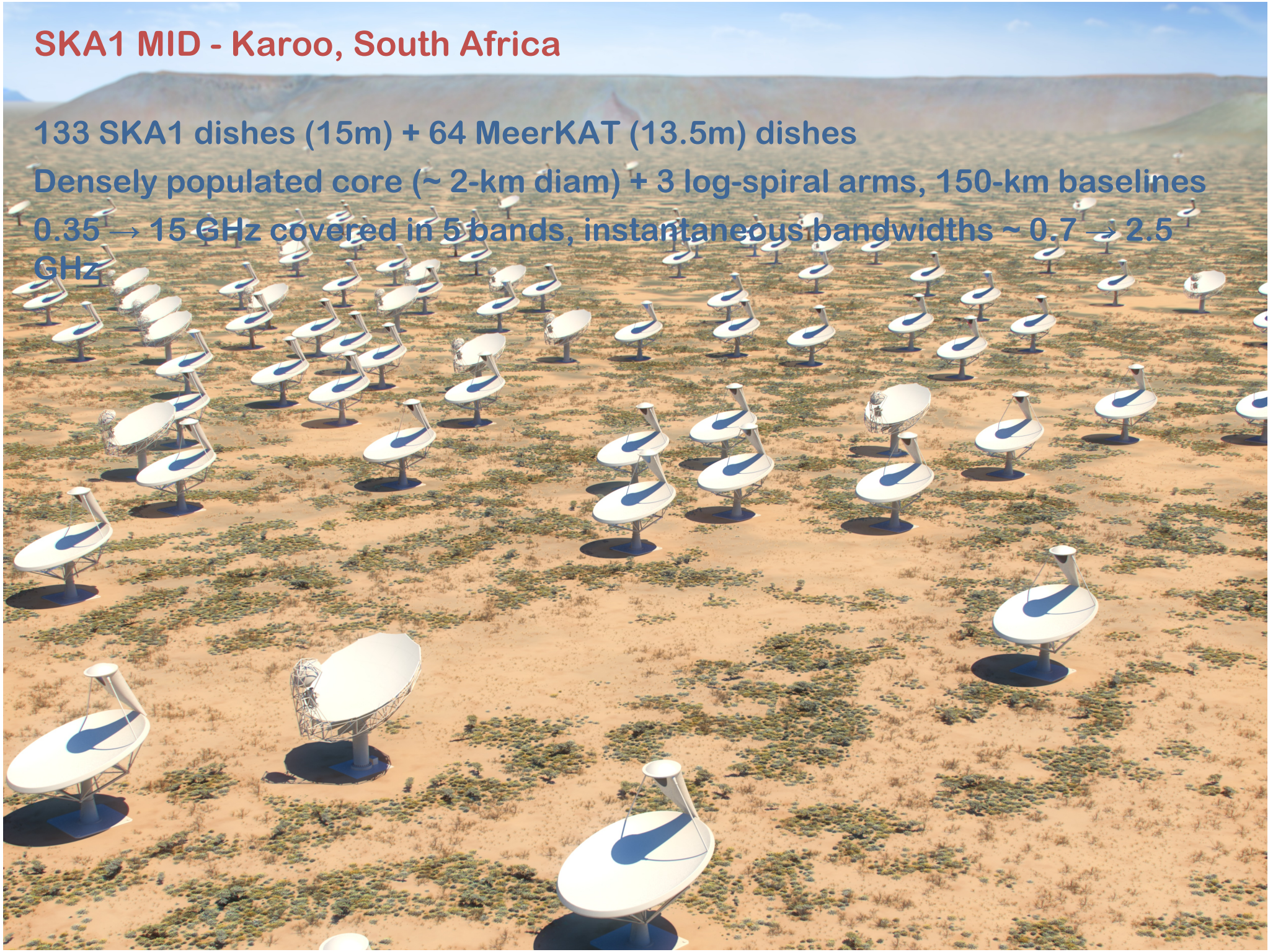
Currently in discussion with others: FRA, GER, JP, KOR, POR, SPA, SWI

SKA1 MID - Karoo, South Africa

133 SKA1 dishes (15m) + 64 MeerKAT (13.5m) dishes

Densely populated core (~ 2-km diam) + 3 log-spiral arms, 150-km baselines

0.35 → 15 GHz covered in 5 bands, instantaneous bandwidths ~ 0.7 → 2.5 GHz

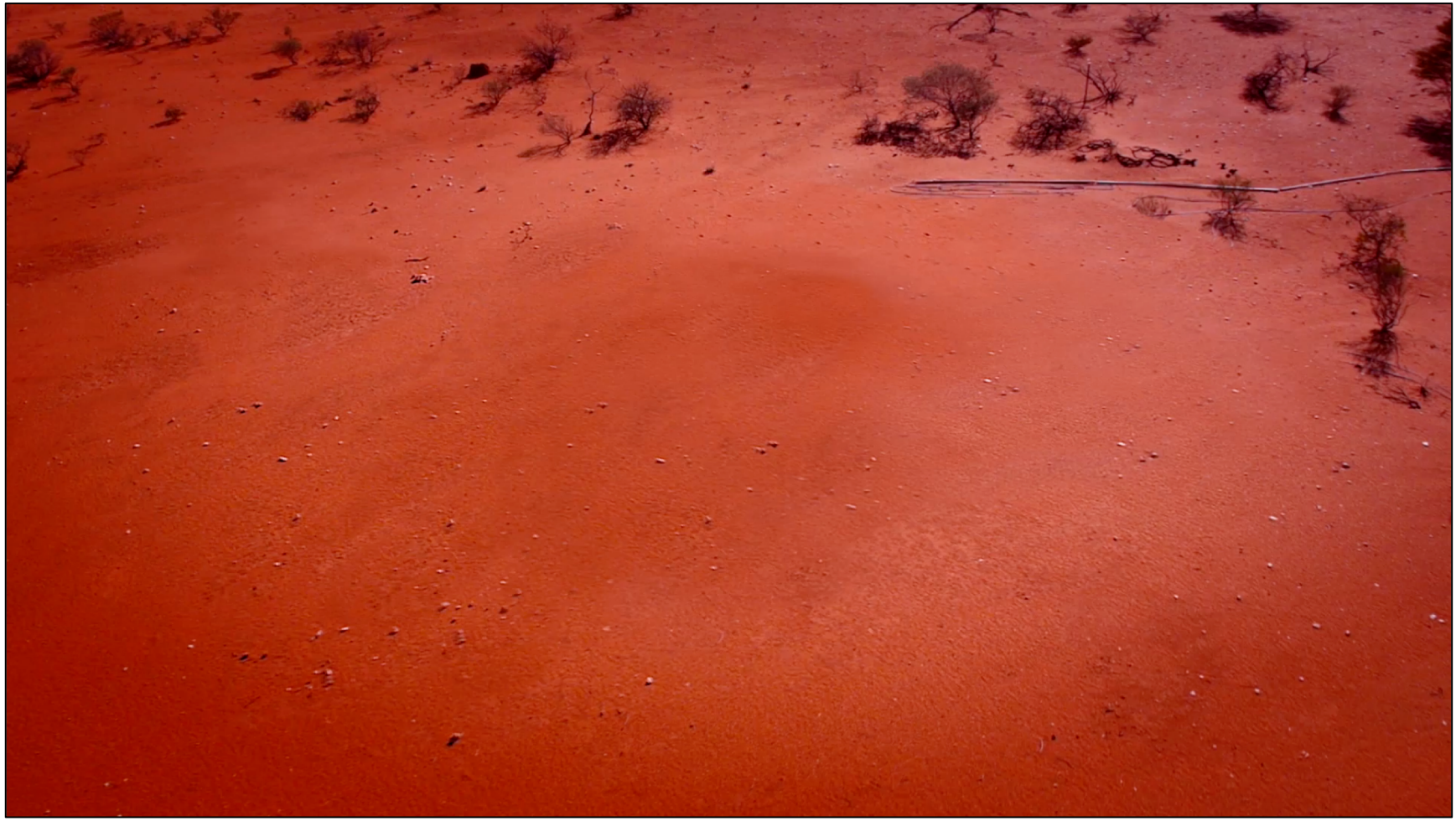

SKA1 MID - Karoo, South Africa
133 SKA1 dishes (15m) + 64 MeerKAT (13.5m) dishes
Densely populated core (~ 2-km diam) + 3 log-spiral arms, 150-km baselines
0.35 → 15 GHz covered in 5 bands, instantaneous bandwidths ~ 0.7 → 2.5 GHz

SKA1 LOW - Western Australia

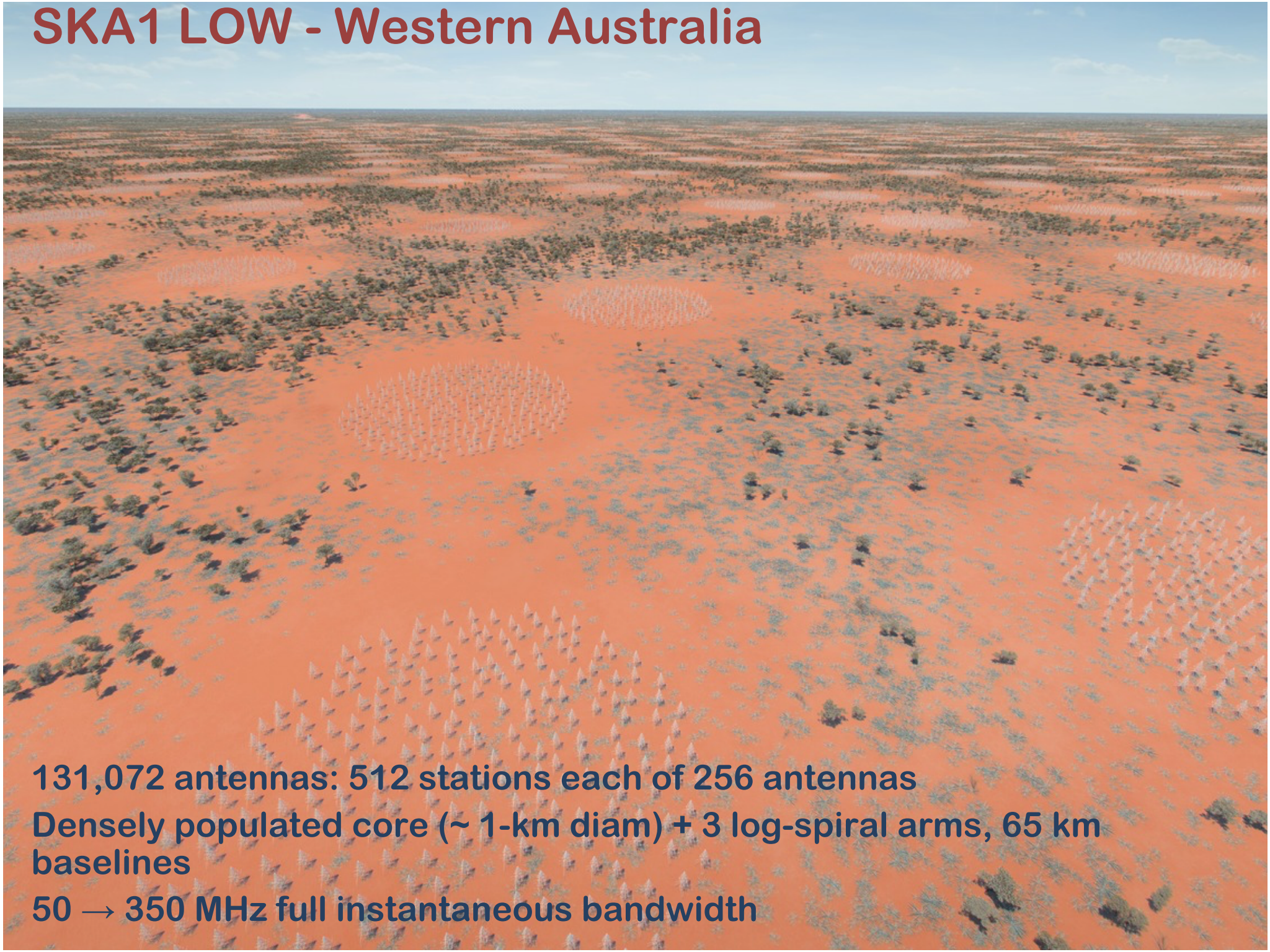
131,072 antennas: 512 stations each of 256 antennas

Densely populated core (~ 1-km diameter) + 3 log-spiral arms, 65 km baselines

50 → 350 MHz full instantaneous bandwidth



SKA1 LOW - Western Australia

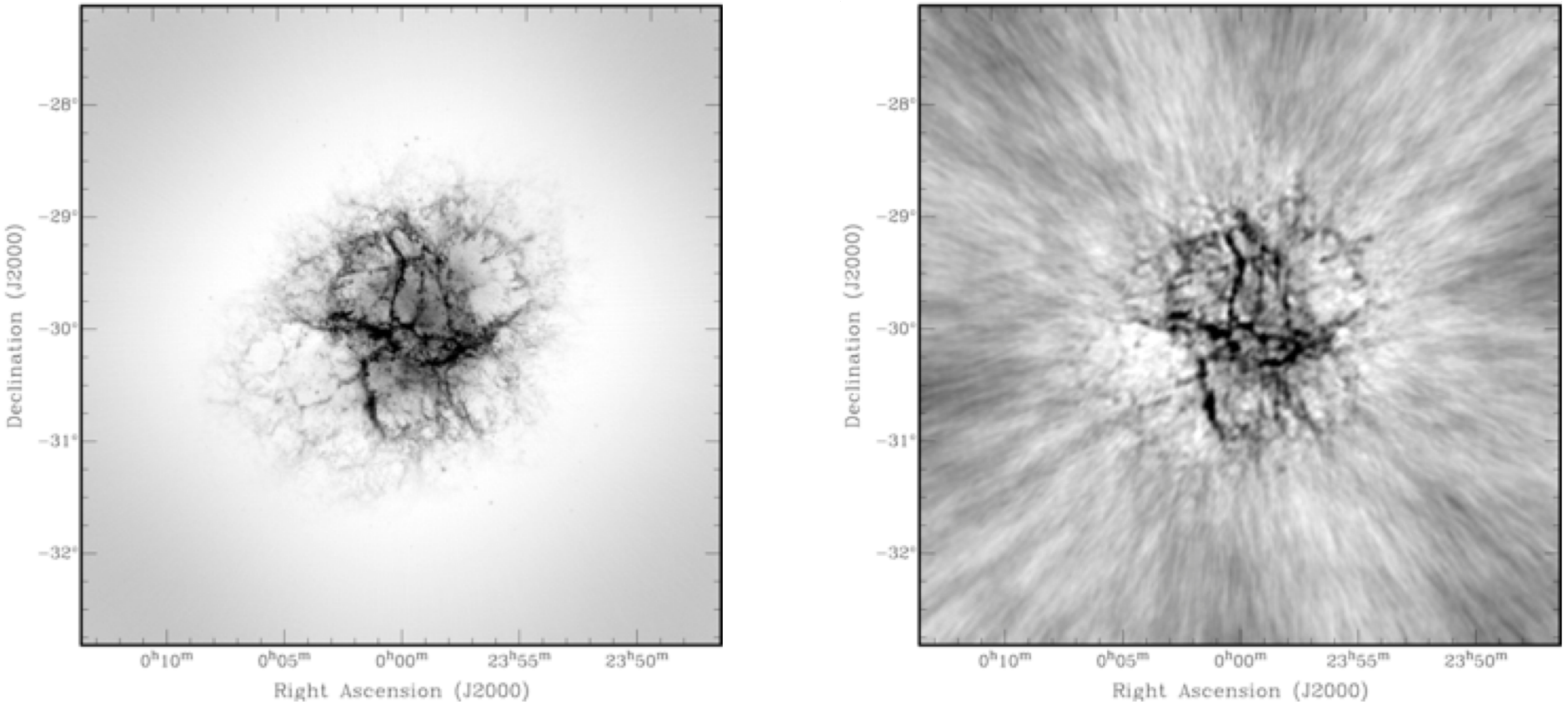


131,072 antennas: 512 stations each of 256 antennas

Densely populated core (~ 1-km diam) + 3 log-spiral arms, 65 km baselines

50 → 350 MHz full instantaneous bandwidth

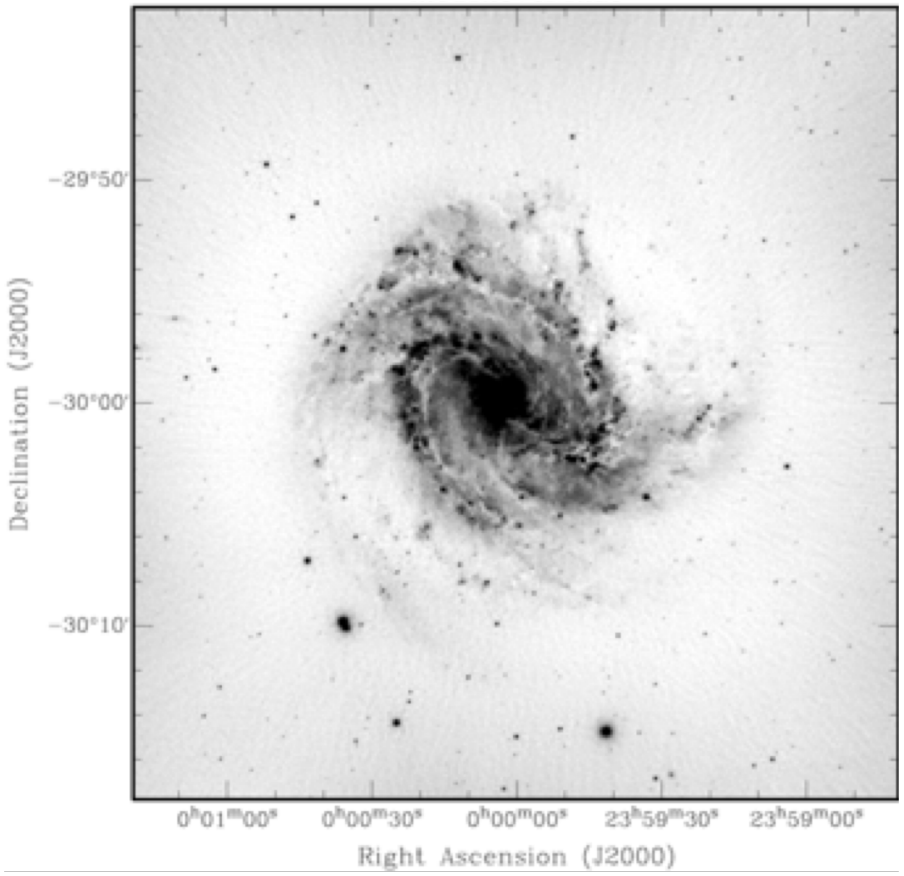
SKA science performance

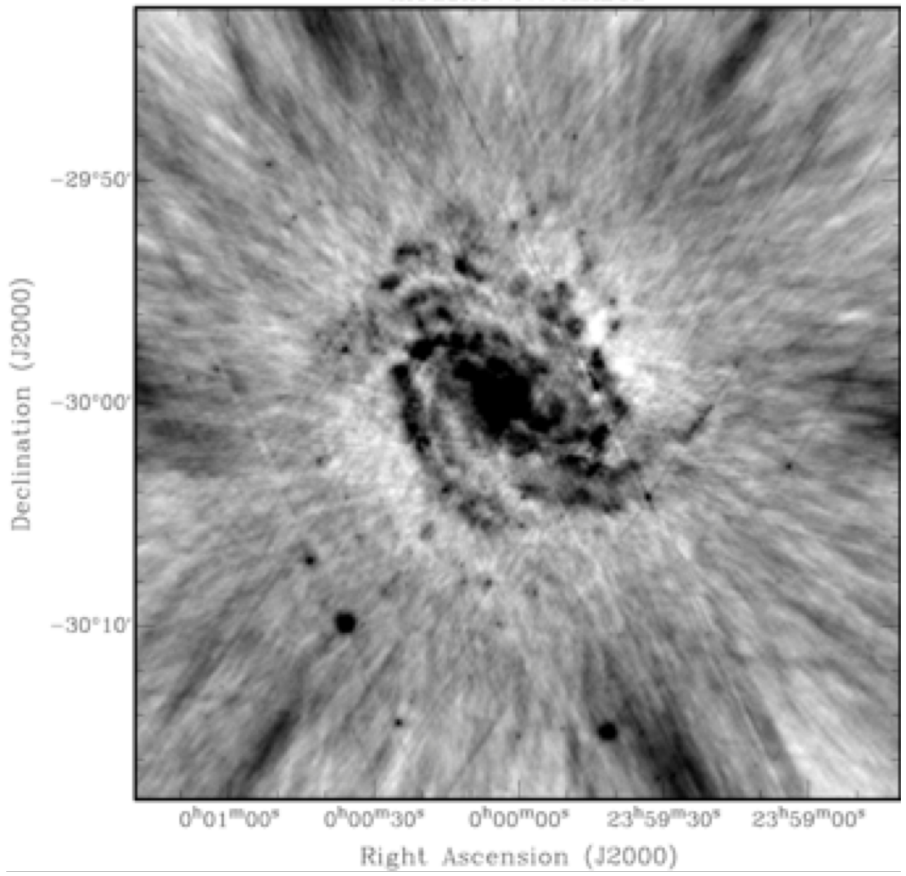


Crab Nebula: Simulated SKA1-LOW snapshot image compared to LOFAR snapshot

(noiseless images; target PSF ~10" at 140 MHz)

SKA science performance





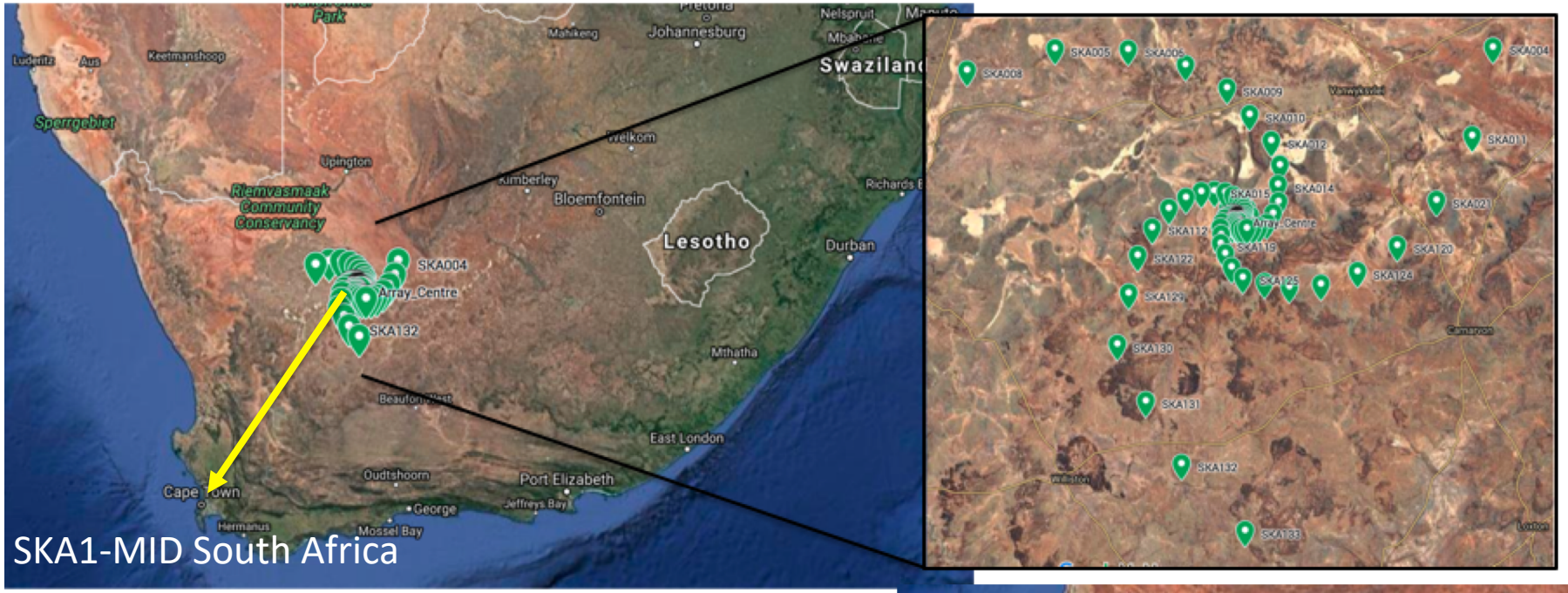
M83: Simulated SKA1-MID snapshot image compared to combination of VLA snapshot in A+B+C+D configurations

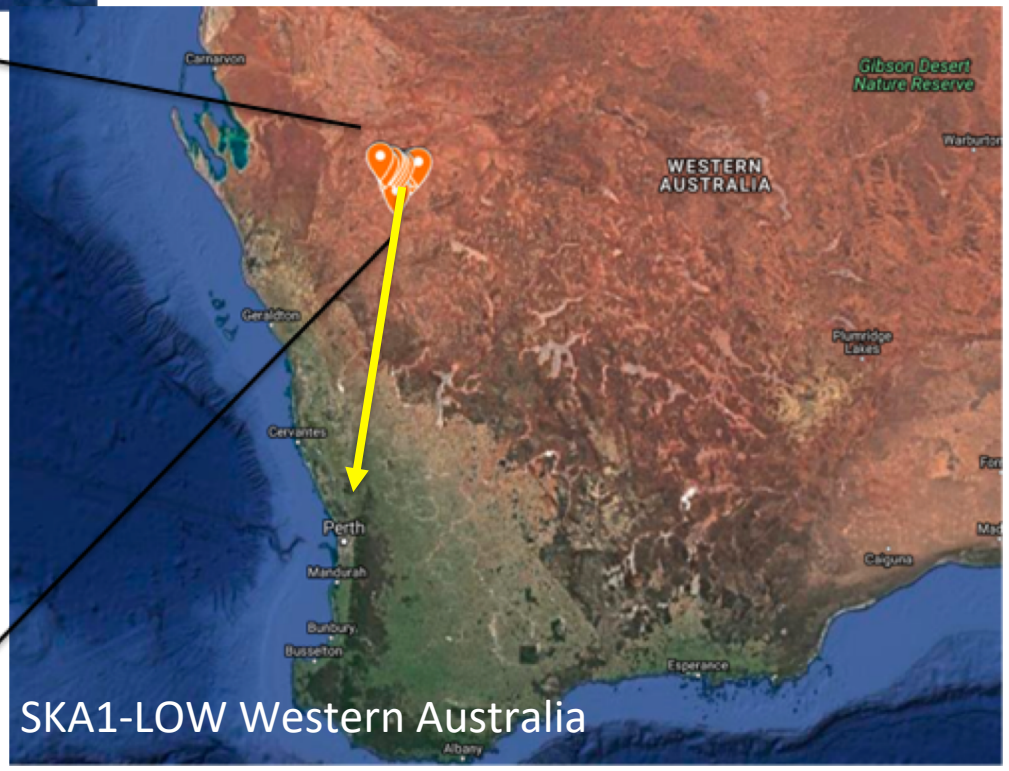
(noiseless images; target PSF ~1" at 1.4 GHz)

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Where?


SKA004
Array_Centre
SKA132
Lesotho
Durban
Johannesburg
Cape Town
Port Elizabeth
East London
Bloemfontein
Kimberley
Welkom
Upington
Sperrgebiet
Riemvasmaak Community Conservancy
Mthatha
George
Mossel Bay
SKA1-MID South Africa


WESTERN AUSTRALIA
Gibson Desert Nature Reserve
Carnarvon
Geraldton
Perth
Mandurah
Bunbury
Busselton
Esperance
Cervantes
Plumridge Lakes
SKA1-LOW Western Australia

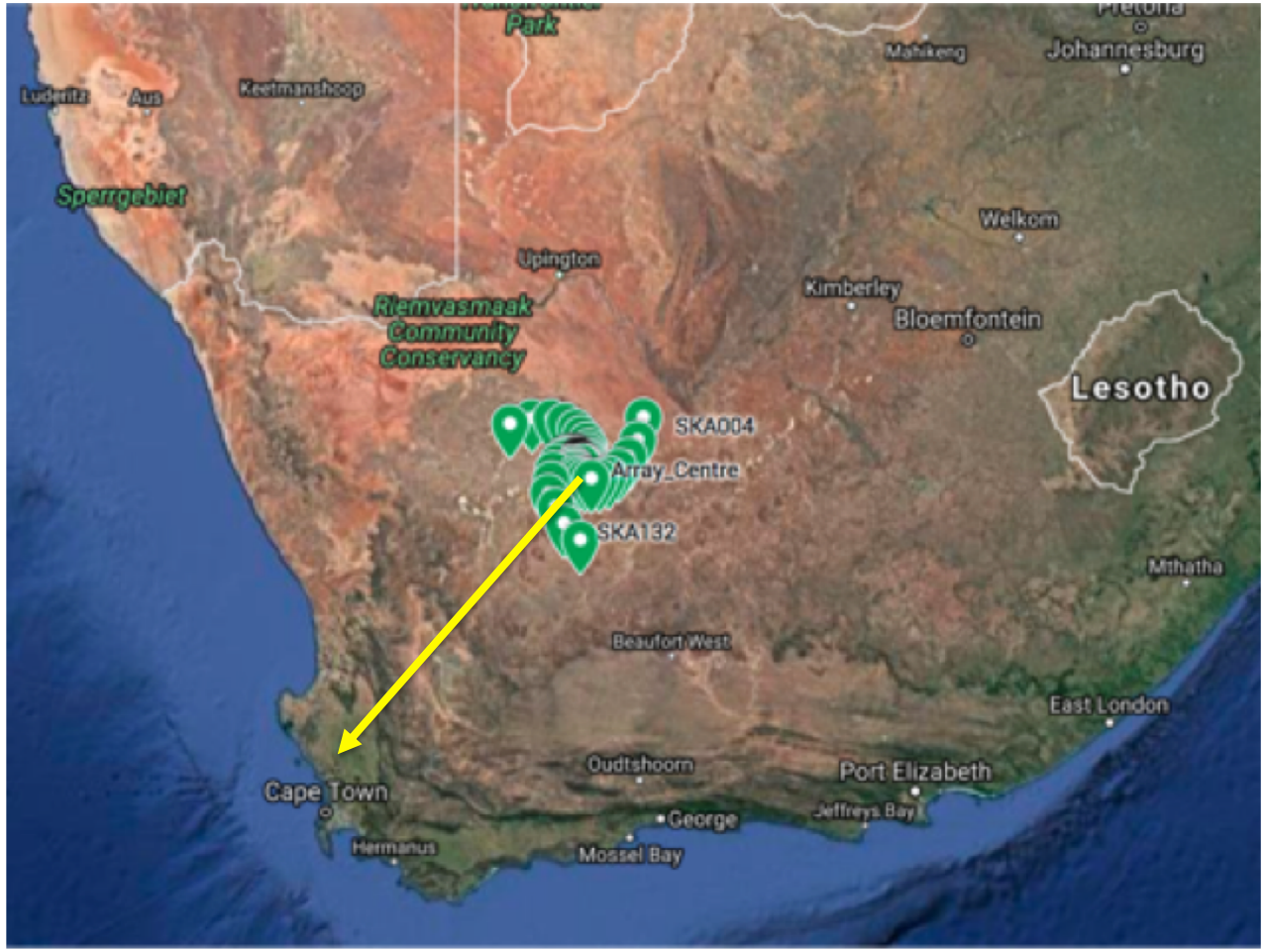
SKA Distributed Operations

Correlator

- located at the telescope site with data transferred to Cape Town to the Science Processing Facility
  - including *real-time* calibration data

Telescope Operator and Science Data processor

- located ~ 600km away from the telescope site at the Science Operations Centre in Cape Town



SKA Distributed Operations

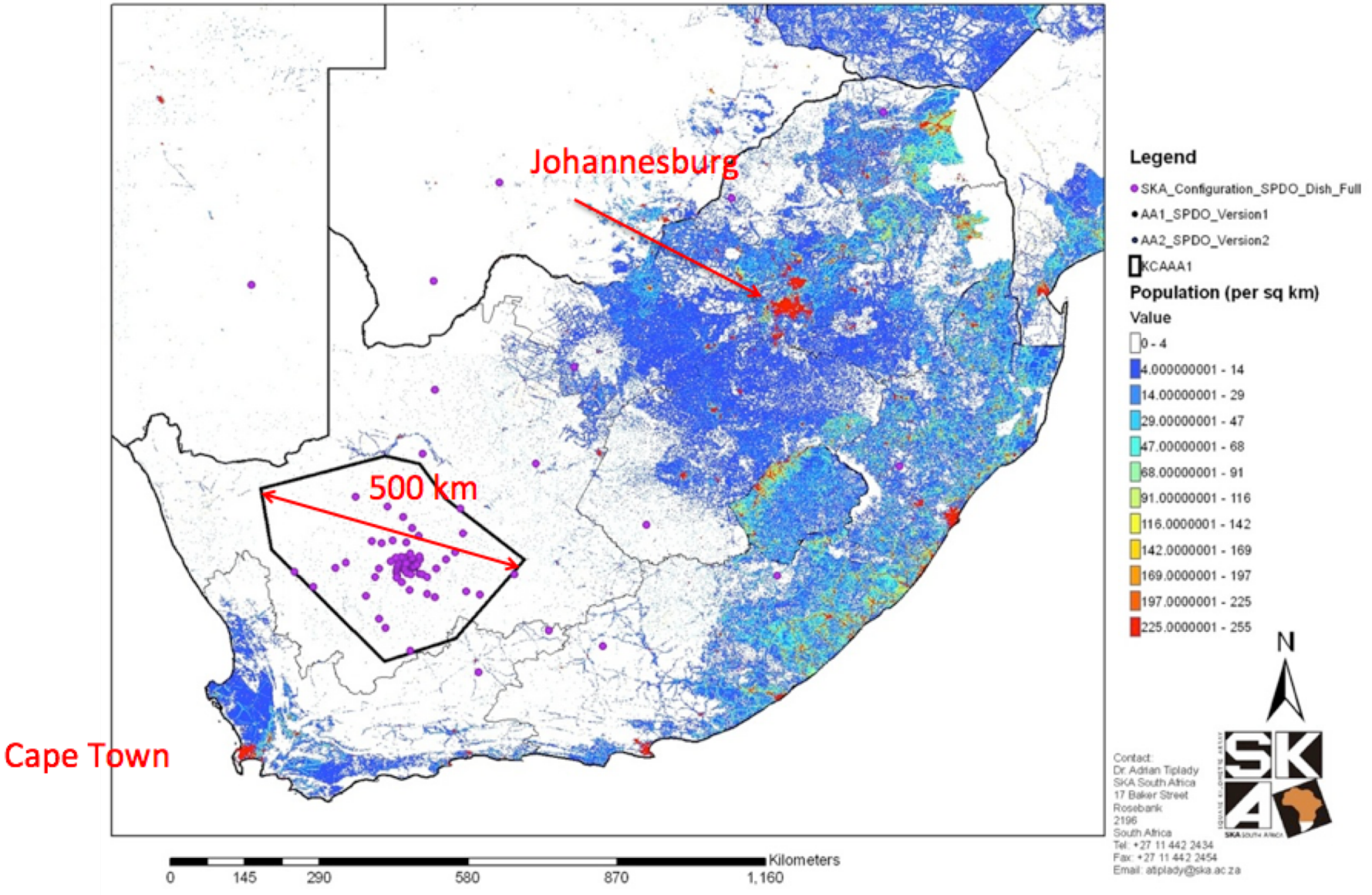
Correlator

- located at the telescope site with data transferred to Perth to the Science Processing Facility
  - including *real-time* calibration data

Telescope Operator and Science Data Processor

- located ~ 600km away from the telescope site at the Science Operations Centre in Perth

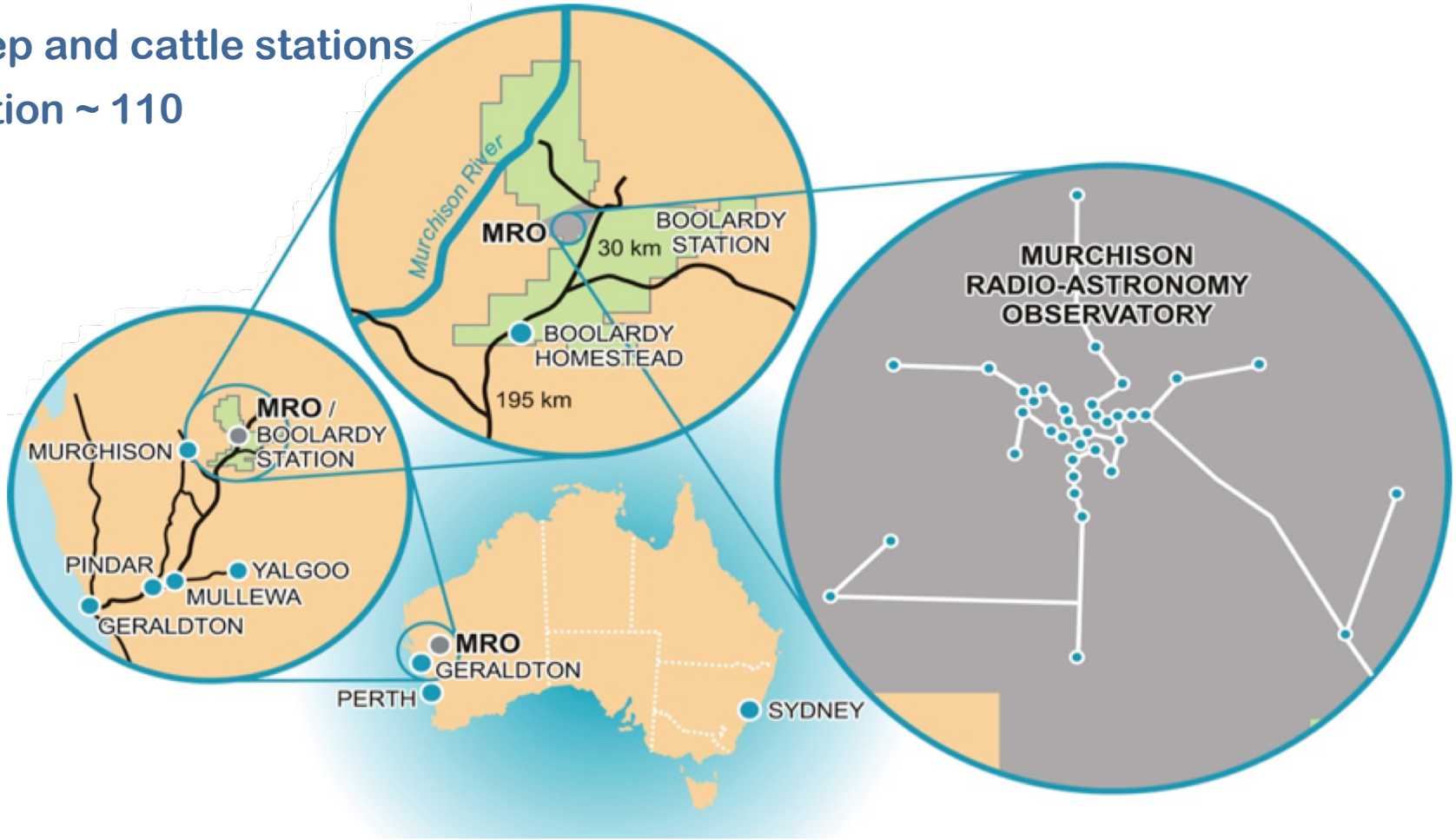
Karoo Radio Astronomy Reserve



Murchison Radio-Astronomy Observatory

Shire of Murchison

- 50,000 km^2 (Scotland ~ 80,000 km^2, Netherlands ~ 40,000 km^2, Wales ~ 20,000 km^2)
- no towns
- 29 sheep and cattle stations
- population ~ 110

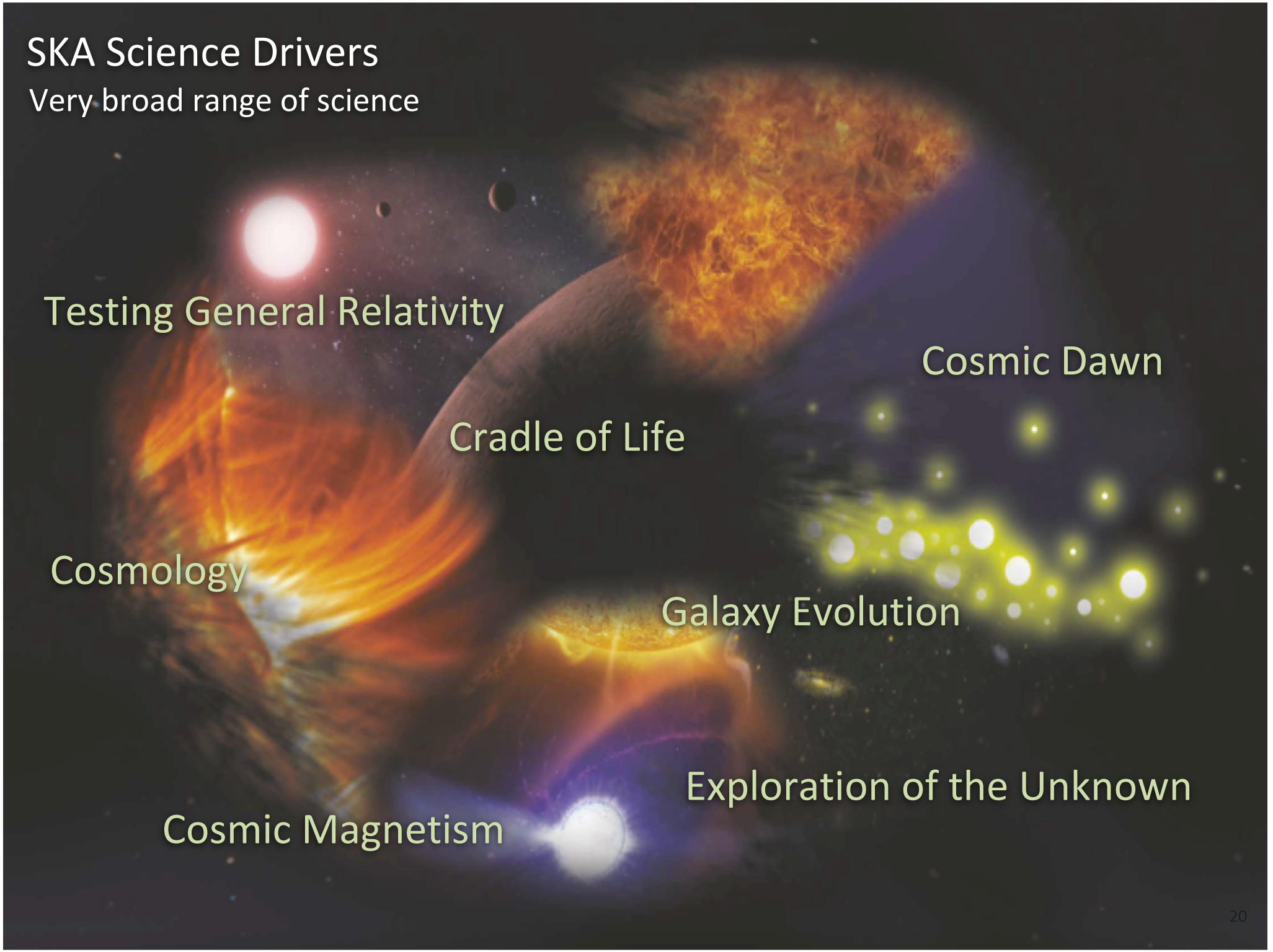


SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Why?

(If you build it, they will come…)


SKA Science Drivers
Very broad range of science
Testing General Relativity
Cosmic Dawn
Cradle of Life
Cosmology
Galaxy Evolution
Exploration of the Unknown
Cosmic Magnetism

SKA Science Case

The Cradle of Life & Astrobiology

How do planets form? Are we alone?

Strong-field Tests of Gravity with Pulsars and Black Holes

Was Einstein right with General Relativity?

The Origin and Evolution of Cosmic Magnetism

What is the role of magnetism in galaxy evolution and the structure of the cosmic web?

Galaxy Evolution probed by Neutral Hydrogen

How do normal galaxies form and grow?

The Transient Radio Sky

What are Fast Radio Bursts? What haven't we discovered?

Galaxy Evolution probed in the Radio Continuum

What is the star-formation history of normal galaxies?

Cosmology & Dark Energy

What is dark matter? What is the large-scale structure of the Universe?

Cosmic Dawn and the Epoch of Reionization

How and when did the first stars and galaxies form?

SKA – An Observatory

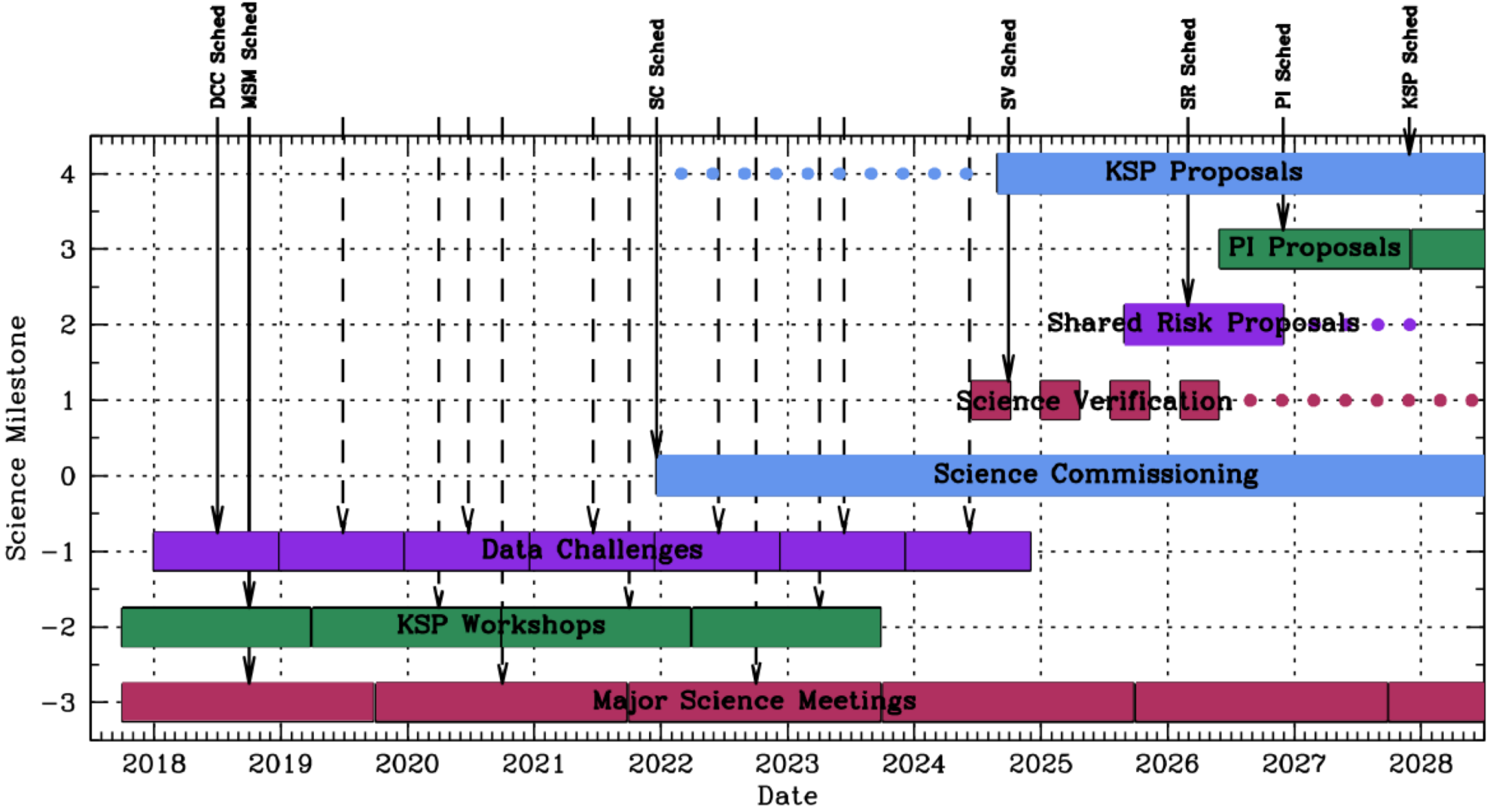
- Open calls for proposals
  - SKA observing programme will be scientist-led
  - Large fraction (perhaps 70% of the time) of the resources will be dedicated to large "Key Science Projects" – large projects that will take years to complete observations for and which will require many hundreds to thousands of hours of telescope time.
    - KSP membership models still work in progress, but will need to be fair and open, reflecting the regional contributions to SKA budgets.
  - Remaining resource will be for PI projects – shorter experiments for which observations can be completed within a single observing cycle (i.e. < 100 hours of telescope time typically).
  - Data will have a proprietary access period (TBD) before becoming public.
  - Target of Opportunity events will be published with low latency.
  - SKA will respond to TOO events
  - Raw antenna voltages stored in special buffer, can be saved if triggered

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

When?

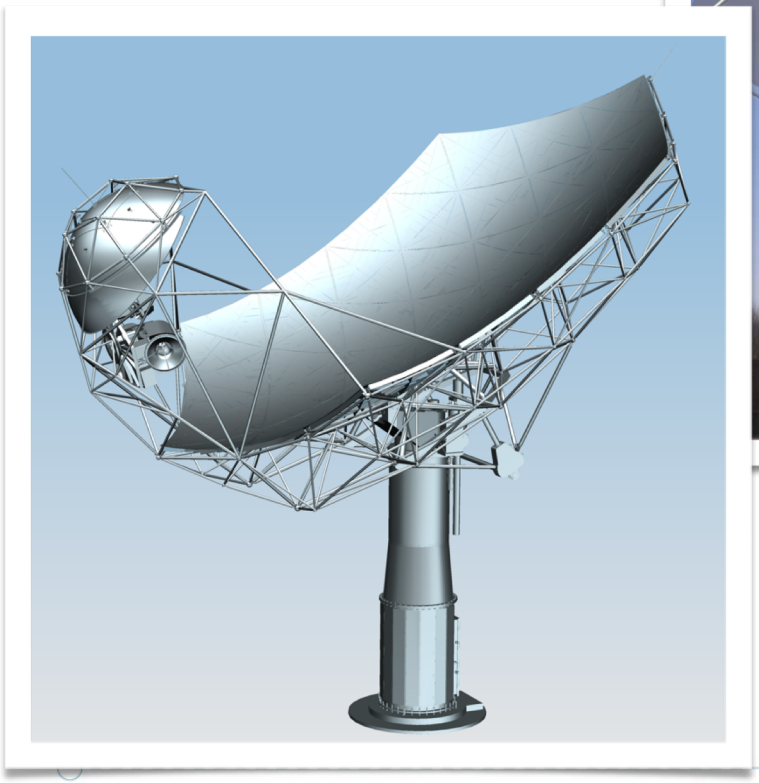
Schedule to SKA Science

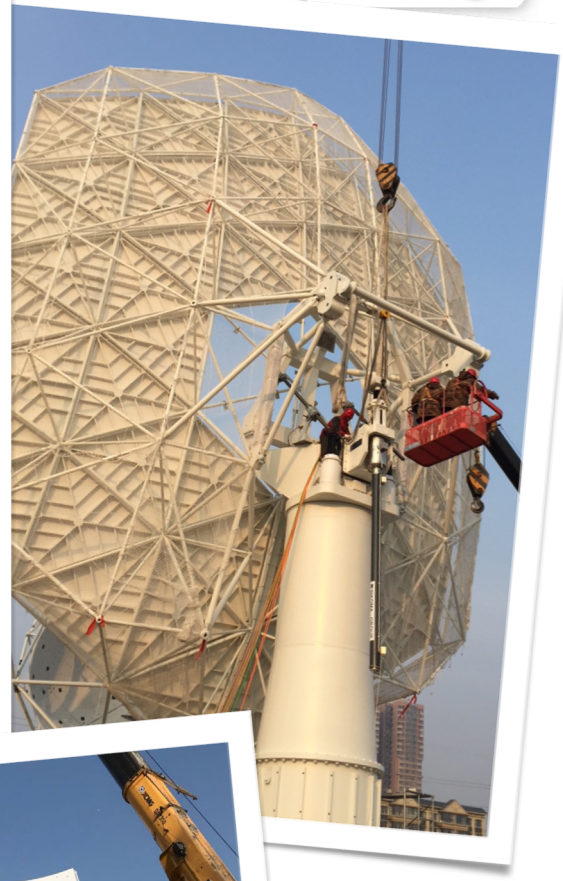


Technical Progress now!

First prototype dish constructed

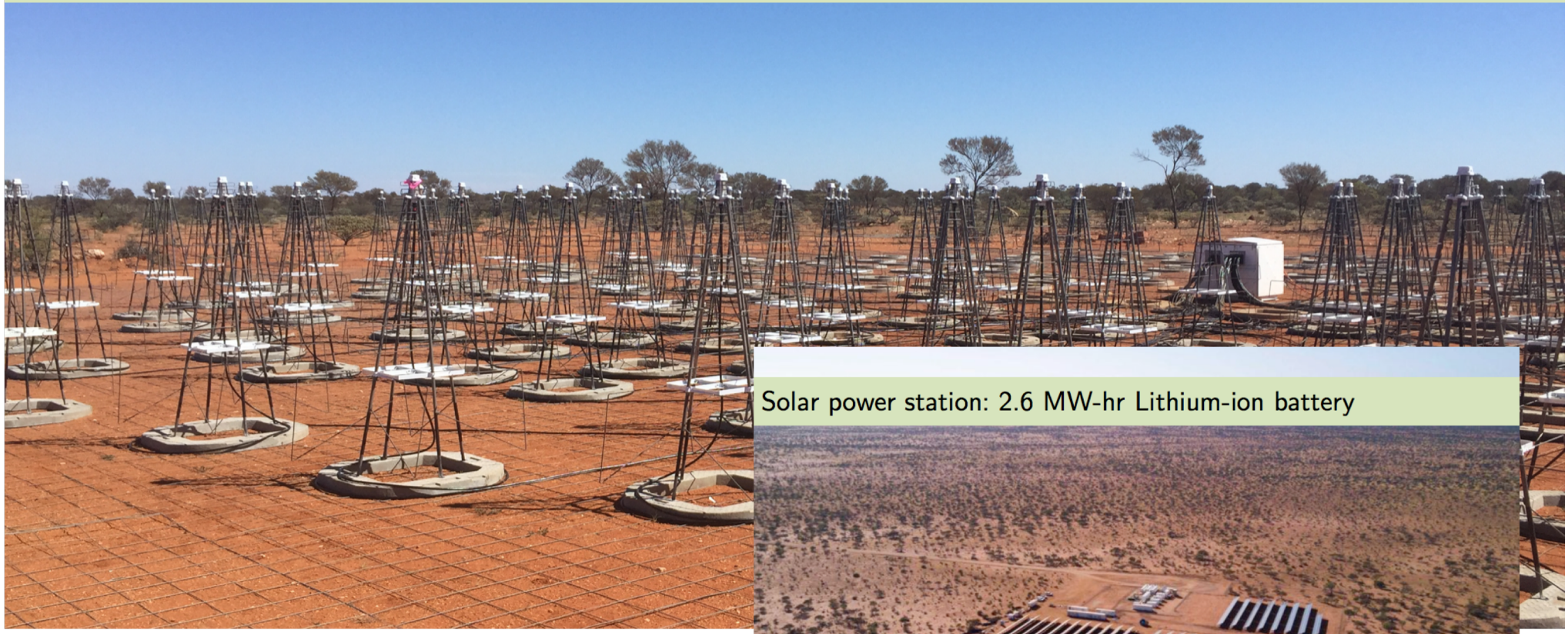
China/Germany/Italy /South Africa





Technical Progress now!

SKA-LOW prototype antenna station deployed



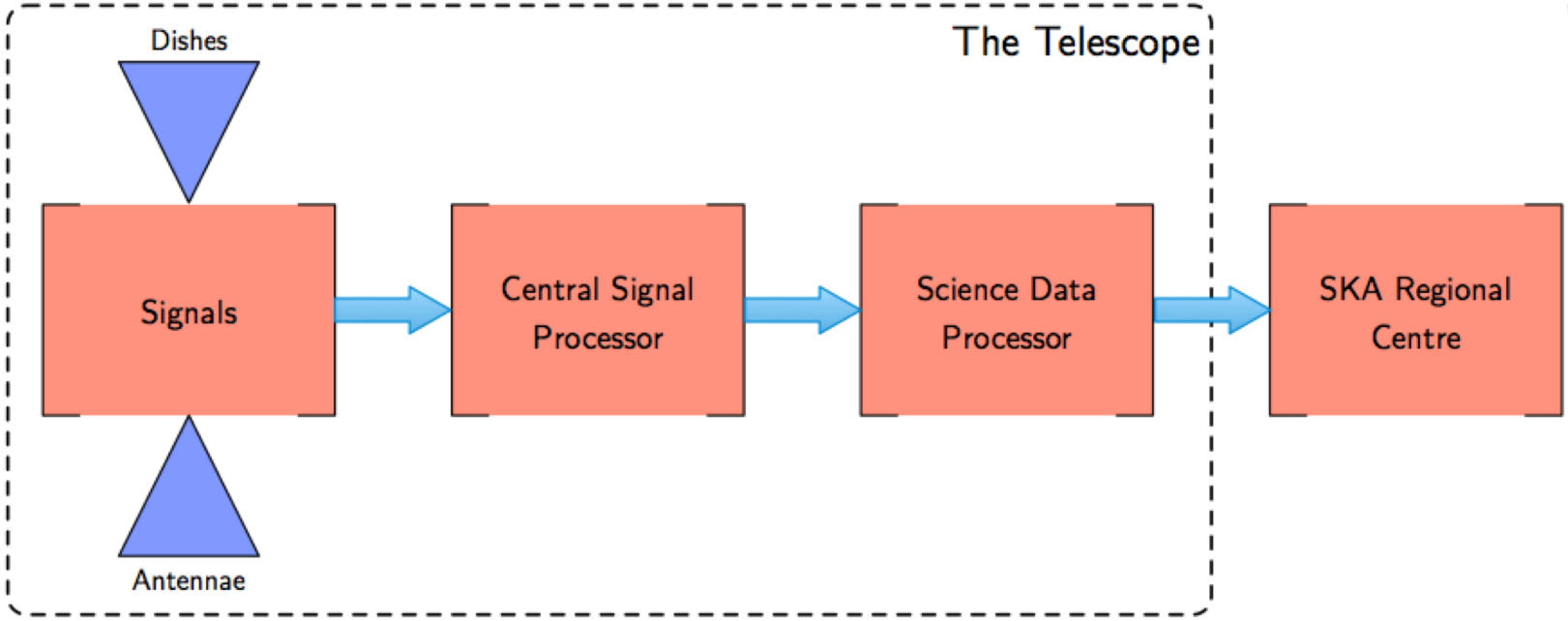
Solar power station: 2.6 MW-hr Lithium-ion battery

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

How?

Data flow challenges



Ordinarily, the Science Data Processor is not considered a part of the telescope

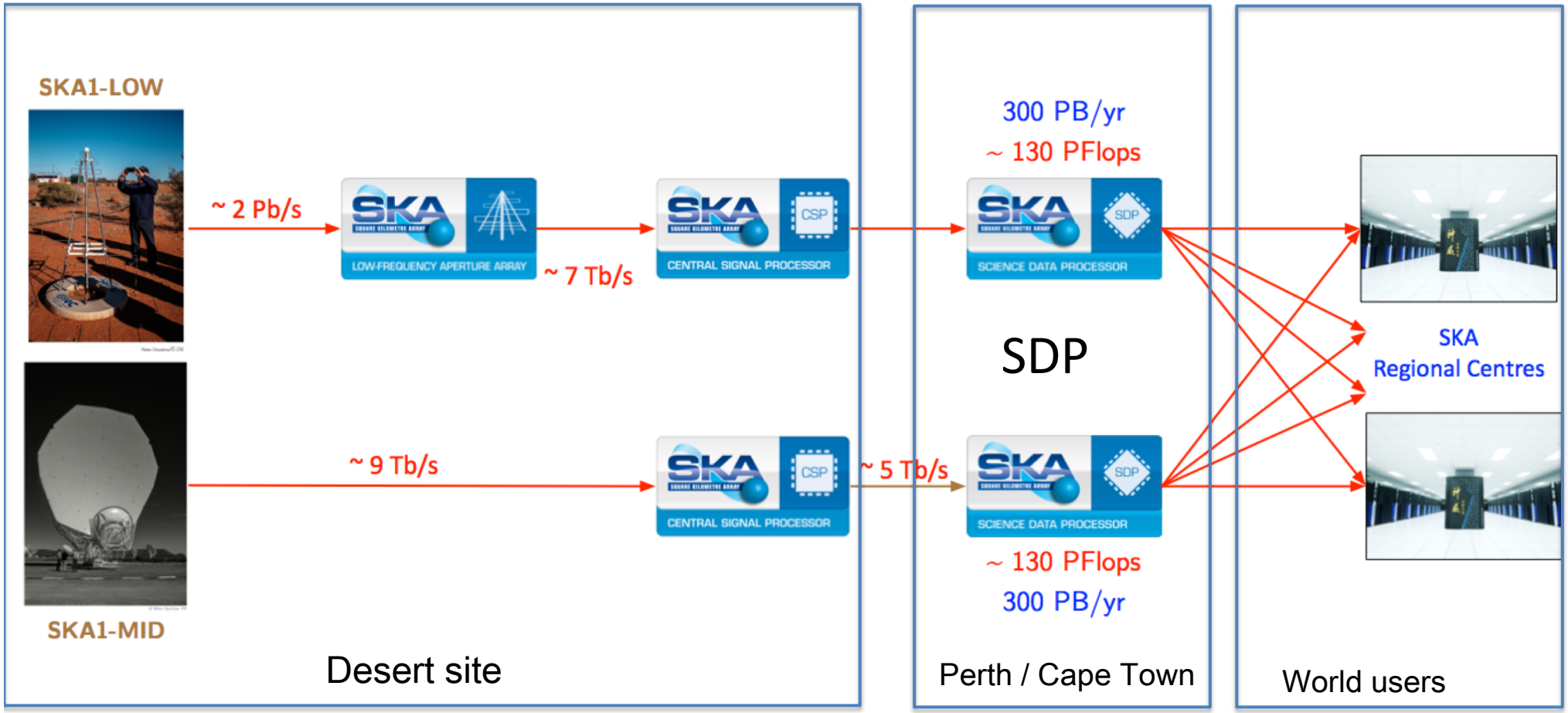
- data reduction should never interrupt data acquisition

But for SKA:

- data rates and volumes emerging from central signal processor are so high that we will not be in a position to store the raw data from the CSP
- it will be cheaper to re-observe than store the raw data indefinitely

The science data processor becomes a schedulable resource of the telescope for observation planning

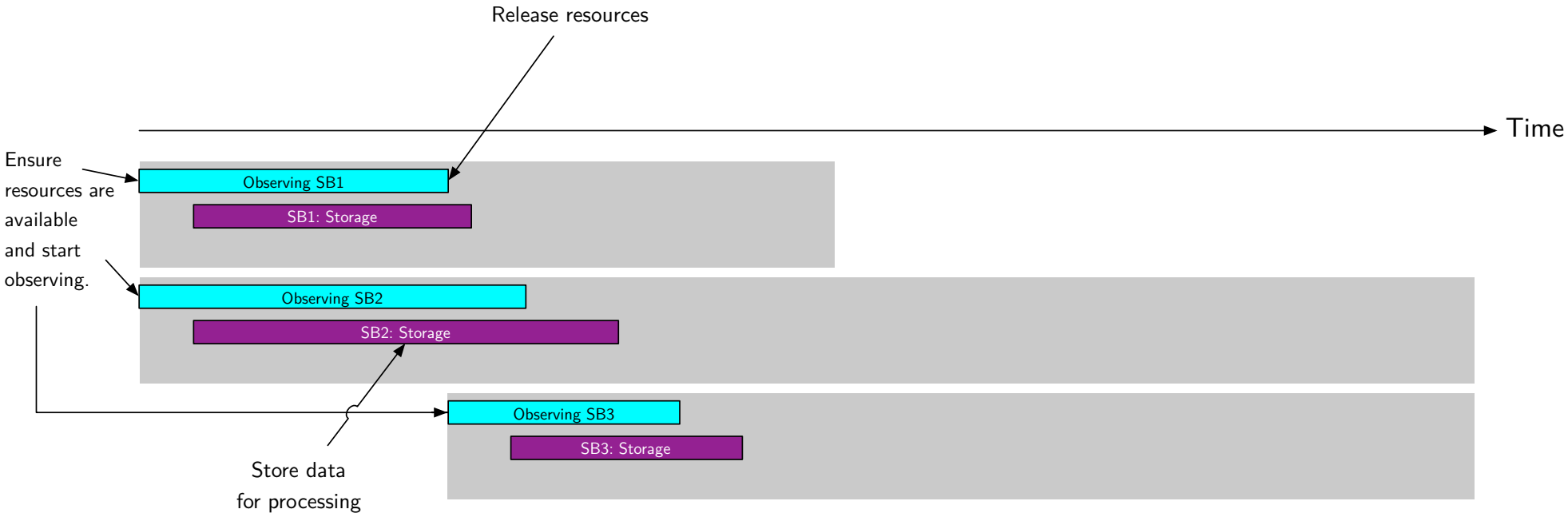
Data flow challenges



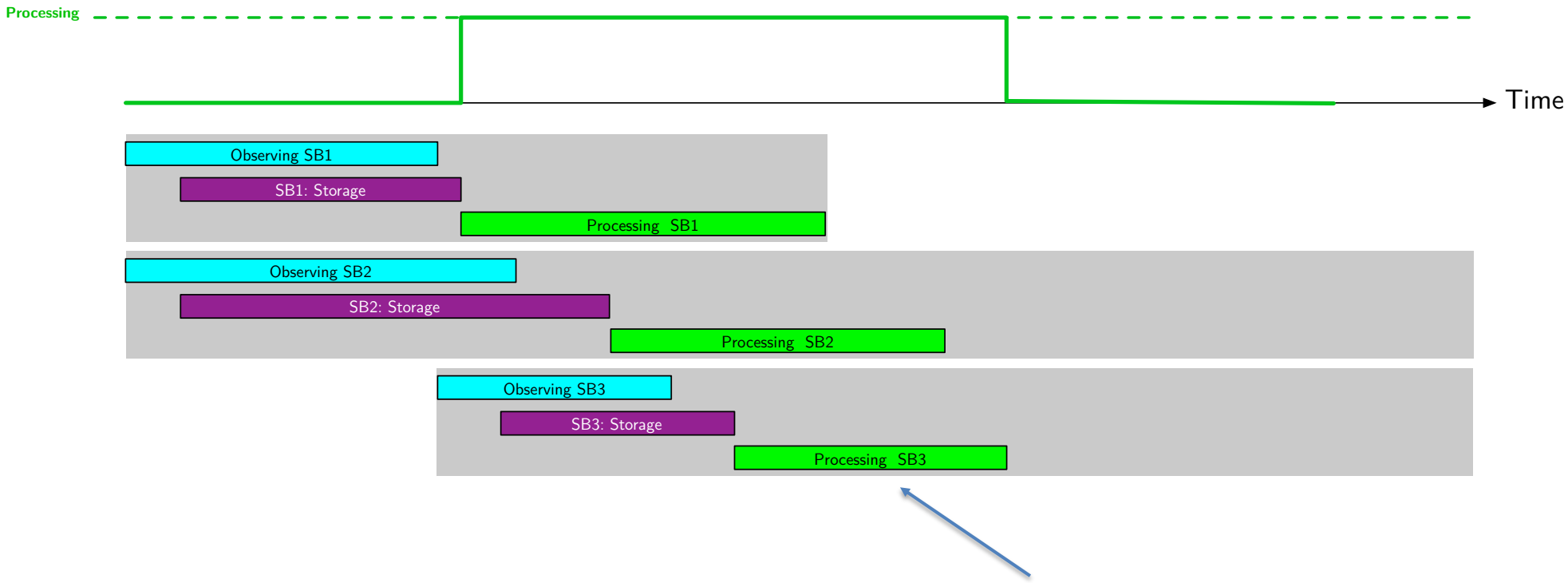
10 – 50 x data rate reduction by SDP

How does the SDP need to function?

Traditional observatory model – external compute

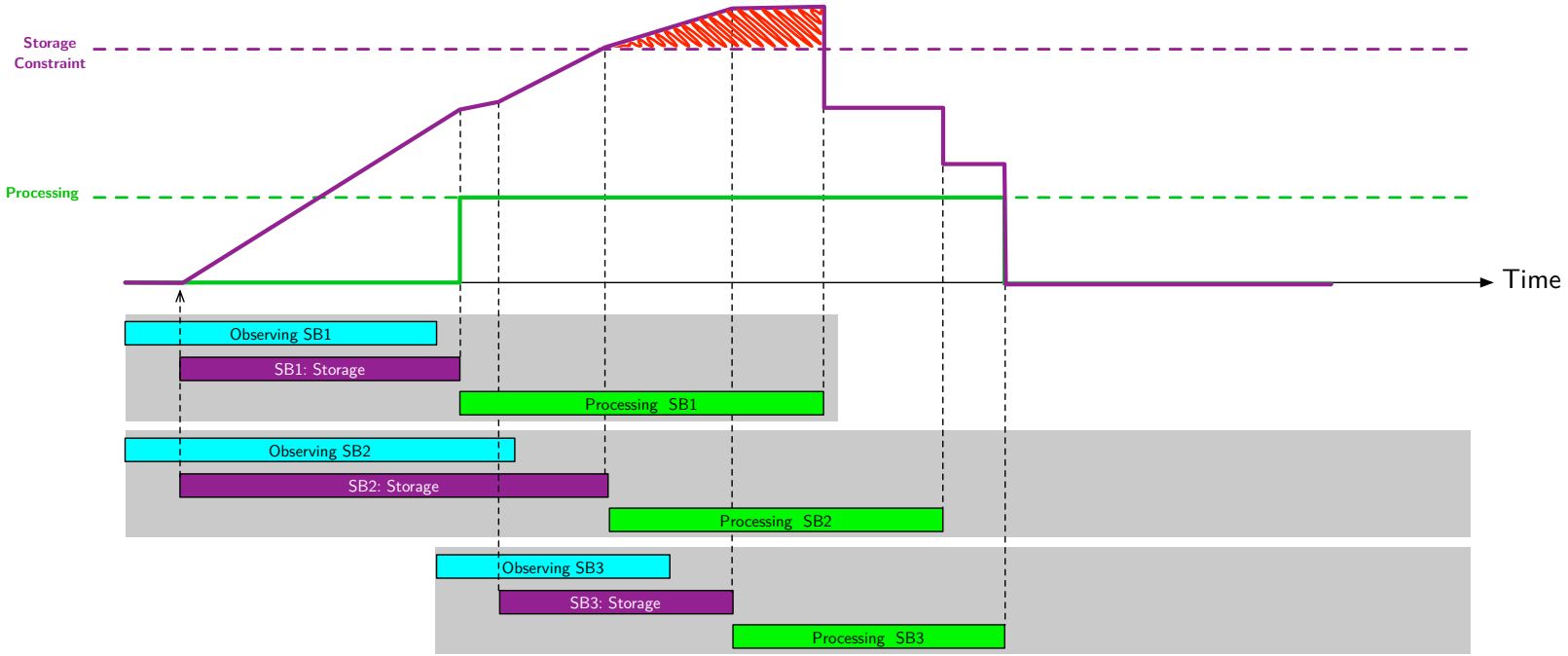


…make processing part of the system

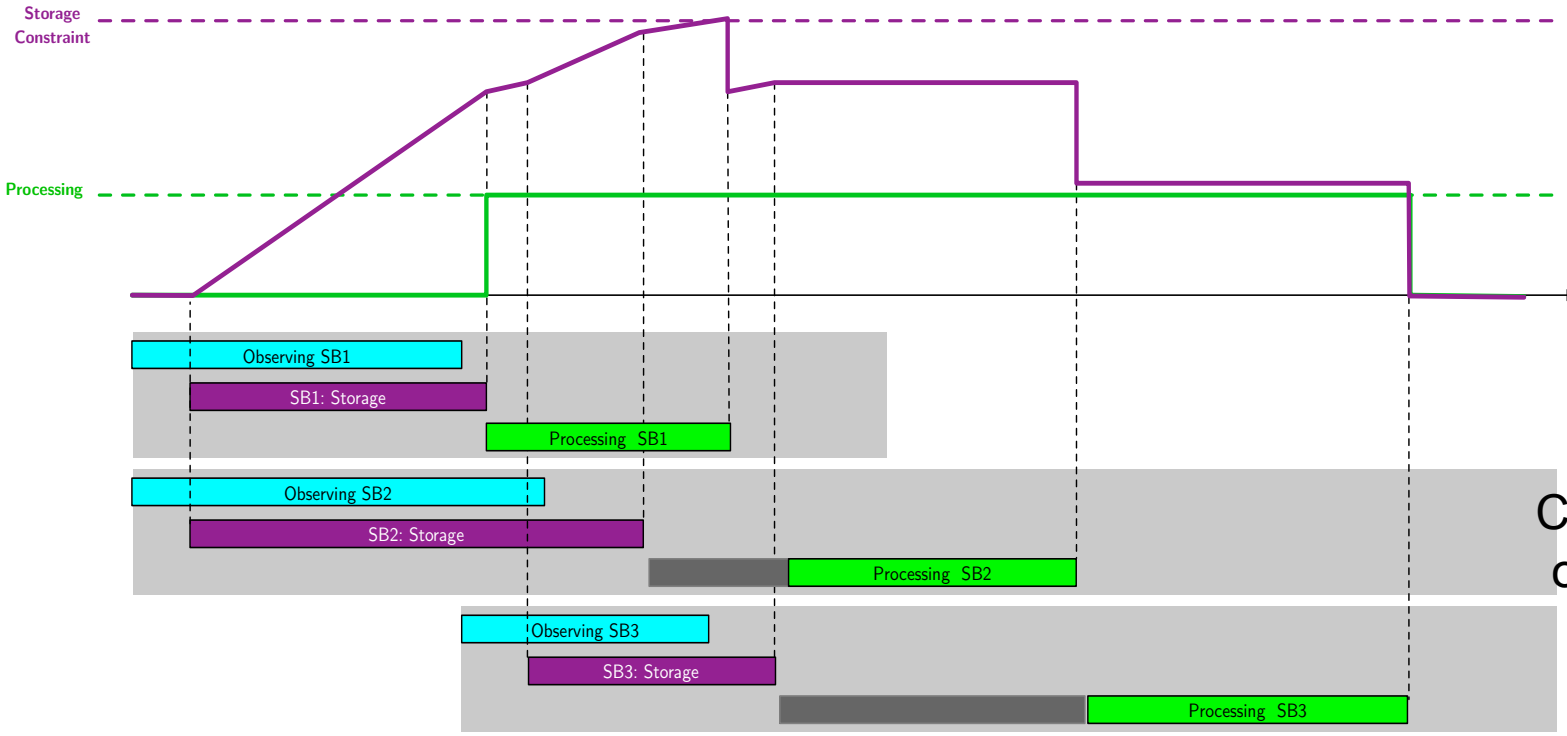


More processing power = shorter processing time

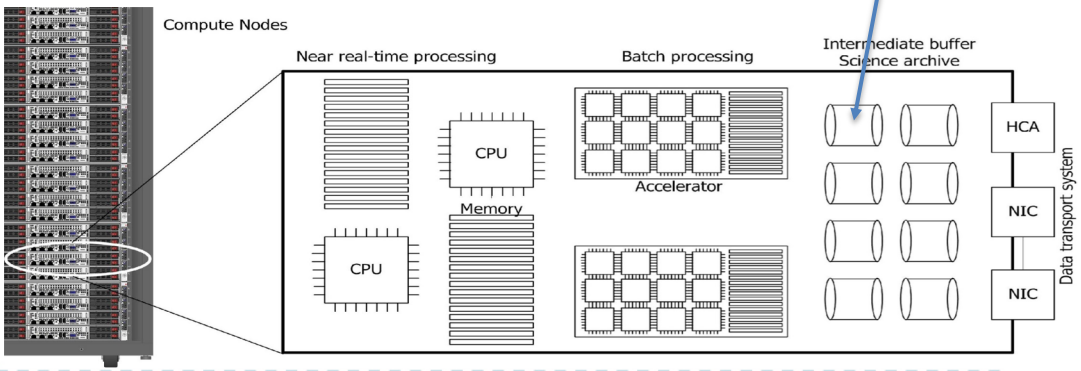
...but need to put data somewhere to process it



Add ”cold” storage to enable load balancing

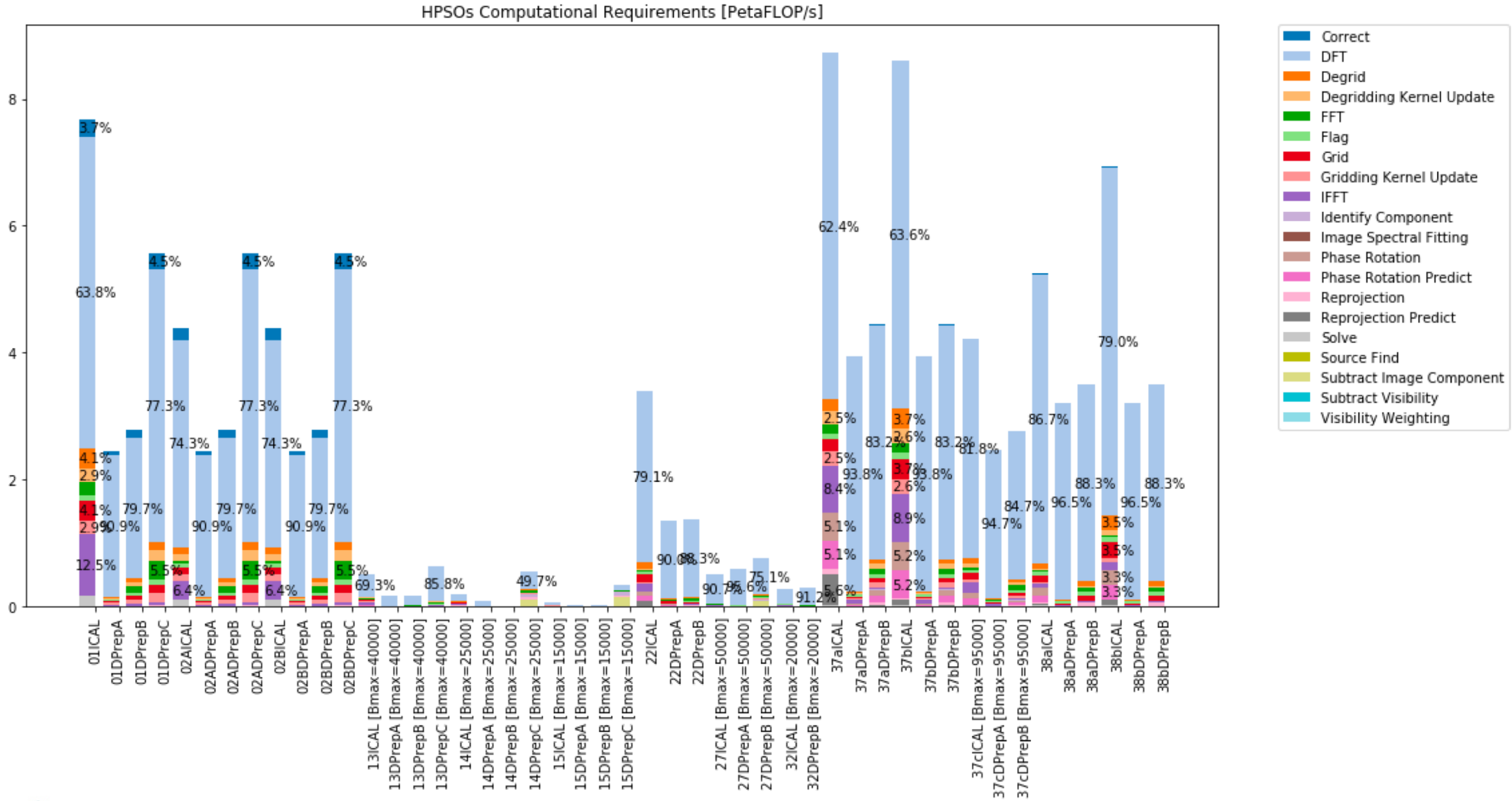


Cold buffer is on compute nodes



SDP – processing variation

System sizing for SDP is based on parametric model and has many assumptions about the pipelines employed by different experiments.



SDP Challenges

Complexity

Multi-axis datasets

Iterative convergent pipelines

Must be predictable for scheduling – but each scheduling block different

~0.25 Exaflop of compute (10% efficiency)

•Orchestration required for data ingest, processing, control, preservation and delivery

SDP Challenges

Capital Cost

constrained on capital cost (SKA1 is a 670M EURO project)

SDP around 10% of that (HW and SW)

SDP Challenges

- Power constraints (i.e running cost)
- Current Exascale roadmap indicates ~40*MW / ExaFLOP in 2023
- We need about 4 times better…

*This number just got worse (was ~25MW)

SDP Challenges

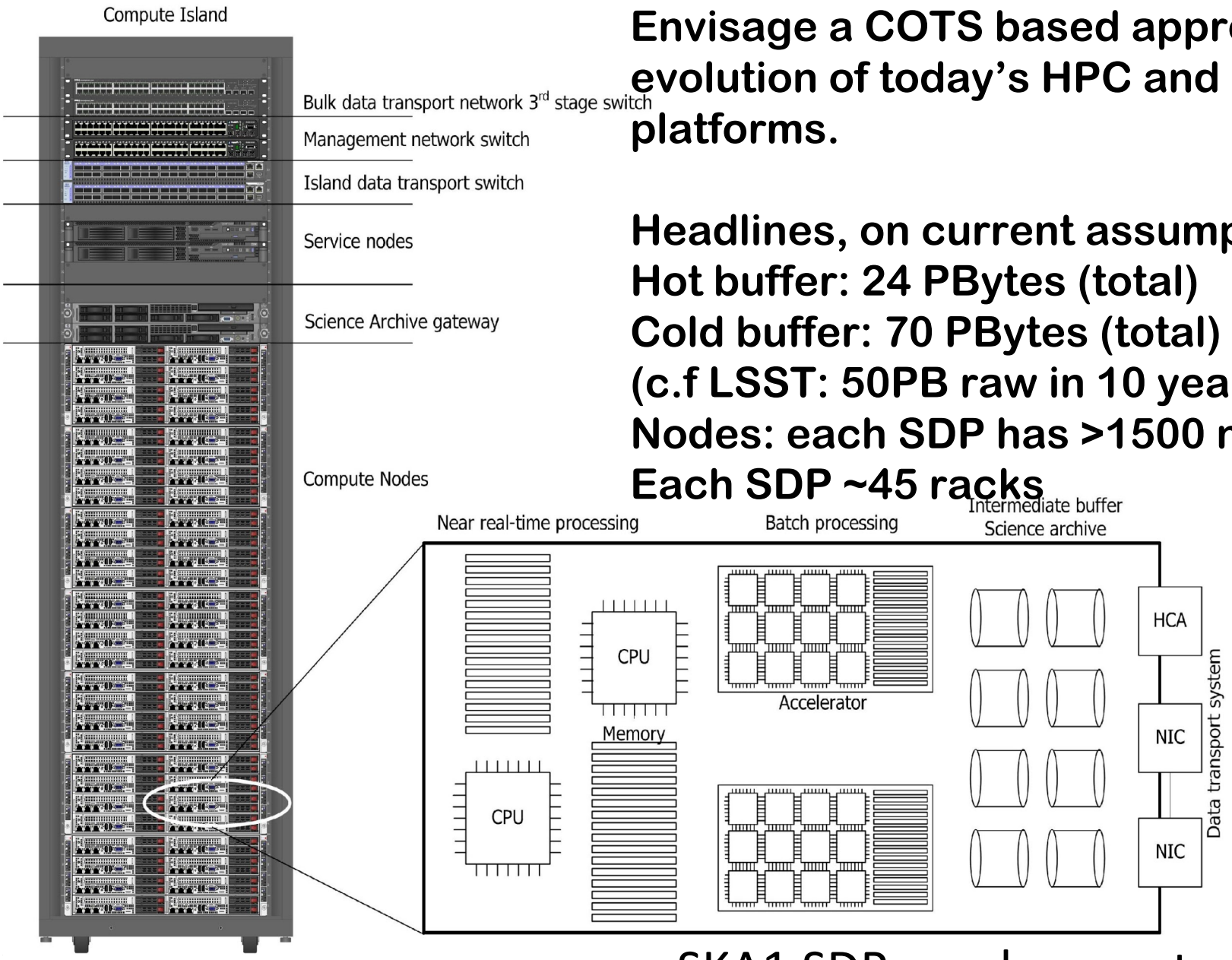
Long lifetime

SW and HW will change – refresh cycles for HW and new algorithms

365 days a year operation, 50 year lifetime

HPC and HTC landscape will evolve

Current SDP Hardware concept



Envisage a COTS based approach – an evolution of today's HPC and Cloud platforms.

Headlines, on current assumptions
Hot buffer: 24 PBytes (total)
Cold buffer: 70 PBytes (total)
(c.f LSST: 50PB raw in 10 years…)
Nodes: each SDP has >1500 nodes
Each SDP ~45 racks

SKA1 SDP purchase not until ~2024

SDP headline design numbers*:

Input

- ~900 GByte/s INGEST

Process

- 200 PByte/s aggregate bandwidth to fast working memory
- 250 Pflop/s total peak
- 10% efficiency assumed

Temporarily store)

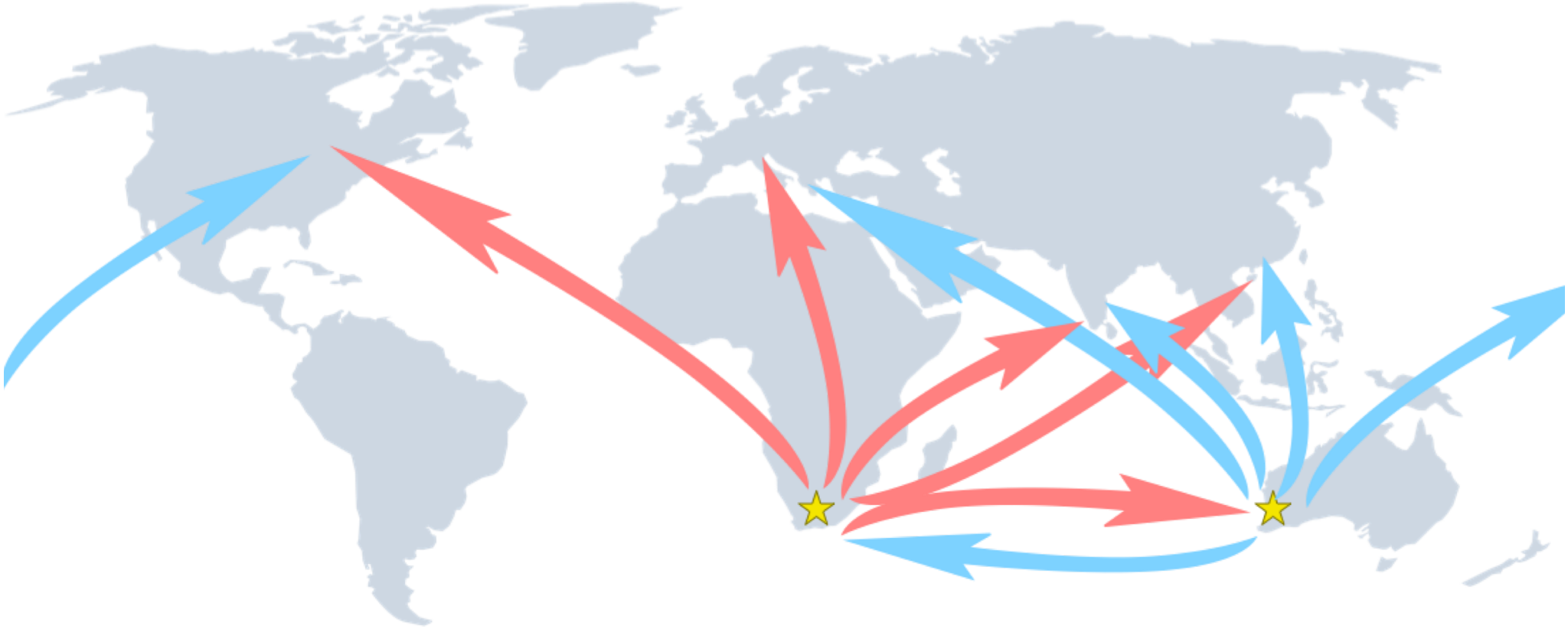
- 100PBytes total distributed buffer

Preserve and ship

- (up to) 2 PetaByte per day of Science Data Products

*all numbers subject to change!

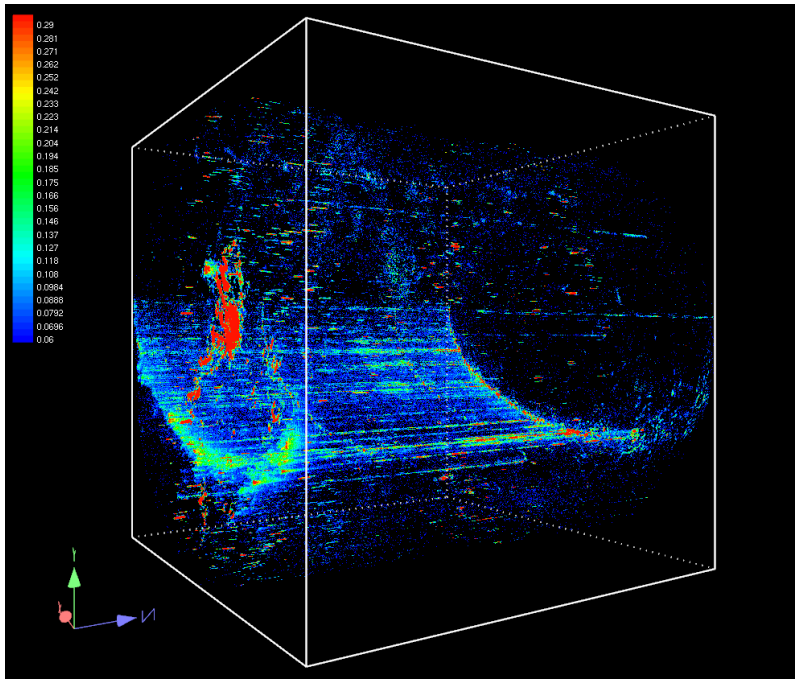
Delivery to users...



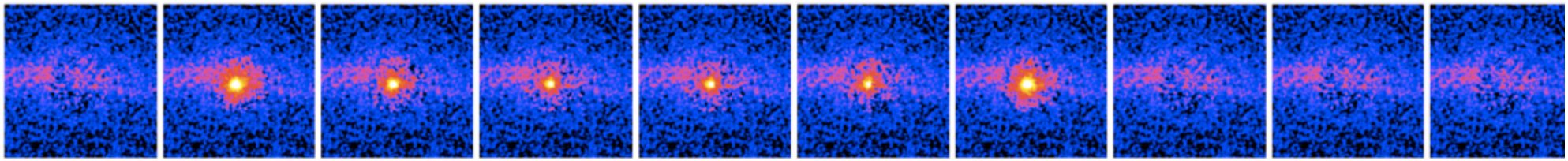
Observatory Data Products flow from the Science Data Processors in Perth and Cape Town to SRCs around the globe

SKA Observatory data products – data rates

Image cubes (2 spatial dimensions, plus radio spectral frequency, polarization)



- Deep-cube: per 6 hours integration, O(50k x 50k) pixels, 50k channels, 4 polarisations: 6 Petabytes (!). 2 Tbit/s
- Image plane searching: per 1 second, O(5k x 5k) pixels, 10 channels, 1 polarisation: each cube 25 Gbytes, Data rate: 200 Gbit/s



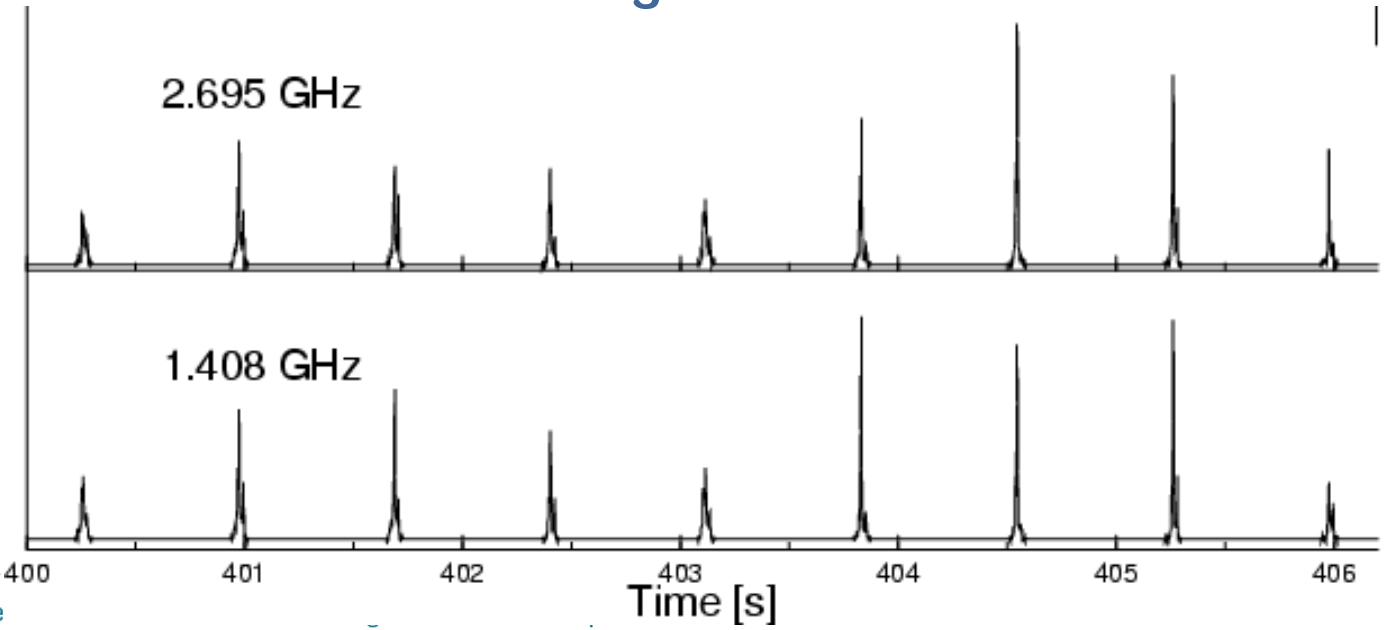
SKA Observatory data products – data rates

Correlated visibility samples

- Few seconds to Sub-second time resolution
- Data rate out of SKA depends on averaging but could be between a few terabits/s and ~ 200 gbit/s

Time-series data for pencil-beams: search for milli-second to second periodic variability in pulsars

- Data rates out of SKA are much lower than image-plane or correlation data products. However, might want to always run transients search mode alongside other modes.



SKA Regional Centres

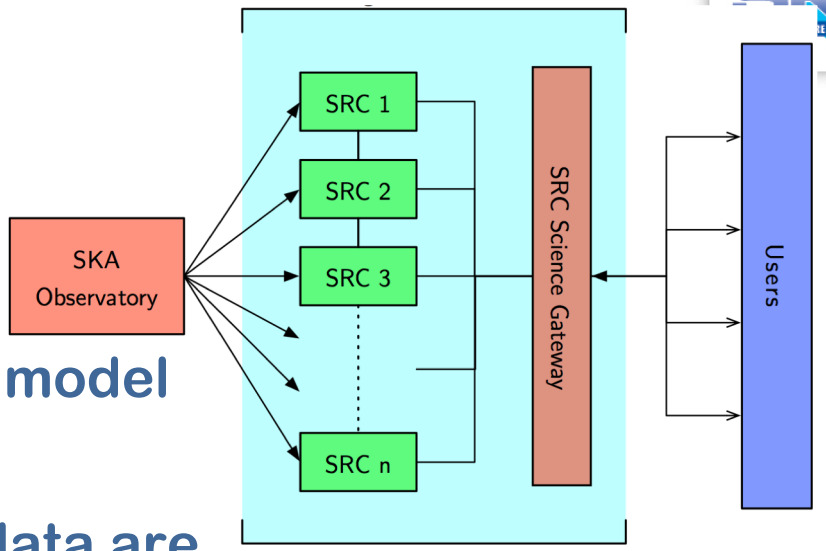
Work in progress!

Outside the cost cap for the project

Essential for delivery of science!

Hope for something "like" the WLCG model

- Resources pledged into system
- Users do not know where "their" data are
- Access given by user-linked data privileges
- Accounting to track resource use



- a collaborative network for collaborative science
- transparent and location-agnostic interface for users
- all SKA users access their project data via SRCs
- a forum for development of software tools: analysis, modelling, visualisation
- Local user support functions

Similar to the DACs for LSST but with open access to archival products

Summary

SKA will be the biggest radio observatory in the world

- 50 lifetime – commissioning data ~2024, Key Science start 2028
- Use cases - science led, very open

Data processing within the observatory ~250 Pflop/s (600x LSST)

Delivery to Regional Centers up to 6/700 Pbytes/year (100 gbit/s out of each site)

Users do not “get” their data

- SRC science gateway to visualise data products, develop SRC pipelines and extract final products (e.g. plots)
- SRCs will be community led, hopefully based on a collaborative model
- SRCs are unlikely to be dedicated HW – must support a mix of different models and architectures.

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

- end

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

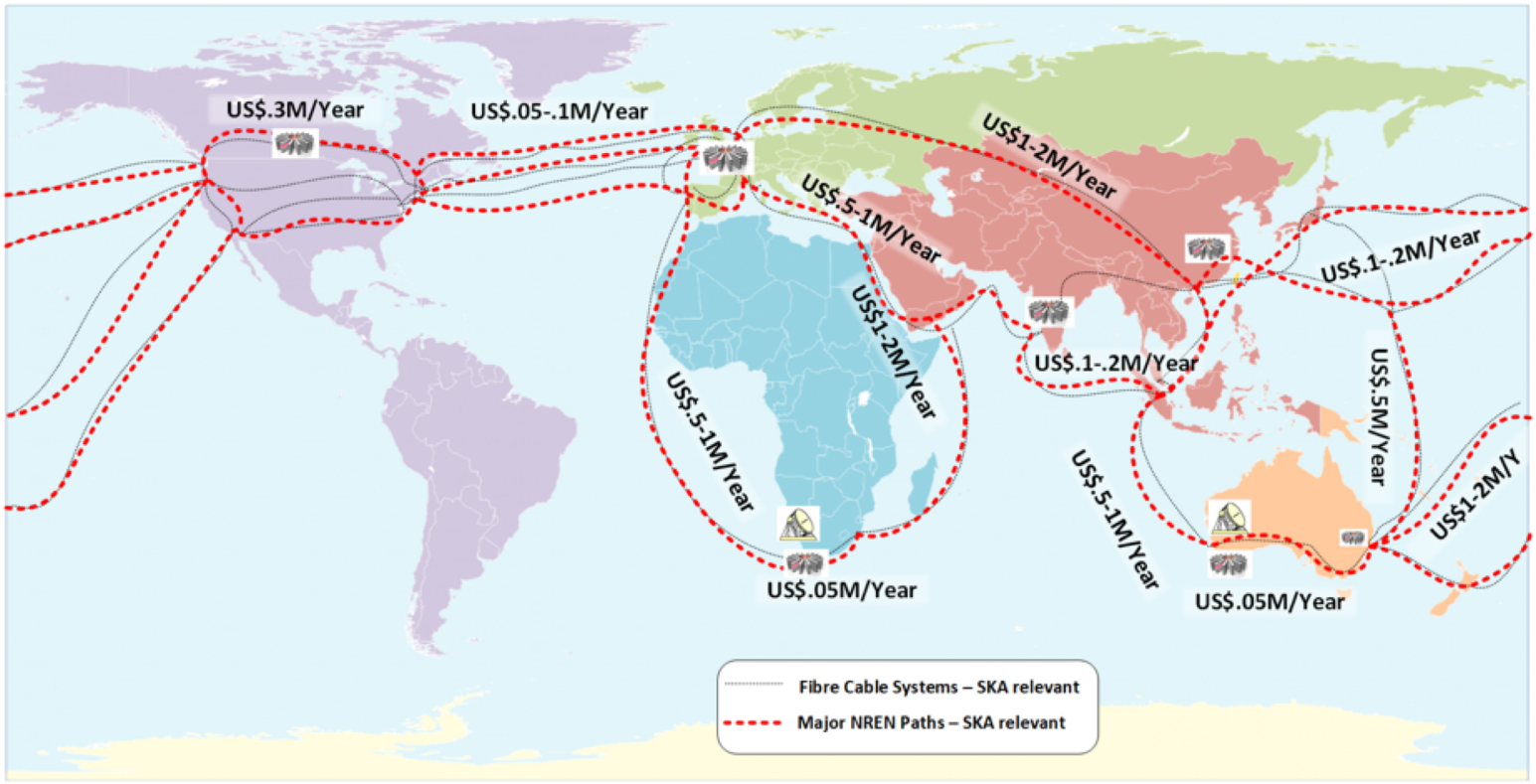
- Spare slides for more info...

SKA Observatory data rates

Data rates out of the observatory are Highly tunable

Our estimates based on simple model of experiments are ~20gbits/s continuous

But we could easily imagine scenarios ten or hundreds of times faster than this (which we couldn't afford to transport or store yet)



Advanced data products

Advanced data products (4) will be generated by the users in the SRCs. When "finished" these will be included in the discoverable SKA archive which will be housed in across the SRCs. We need to understand all of these so we can work out what the SRC workflows look like.

Input from the community is needed to determine what these products should be.

List does not need to be complete - must be able to generate these products from the ODPs…

Co-added images
Rotation measure maps
Moment maps
Power spectra
Cutouts
Pulsar timing residuals
Variance maps
Spatially integrated line spectra
Cross-matched catalogues
Shear maps

The pipelines for generating these products from the ODPs will be developed by the user community, presumably with guidance and input from user support at SRCs

A collaborative model for SKA Regional Centres

There are three main factors that lead to a global collaborative model for SKA Regional Centres (SRCs)

1. The science data products that emerge from the SKA observatory are not in the final state required for science analysis and publication

2. The data volumes are so large that direct delivery to end users is unfeasible

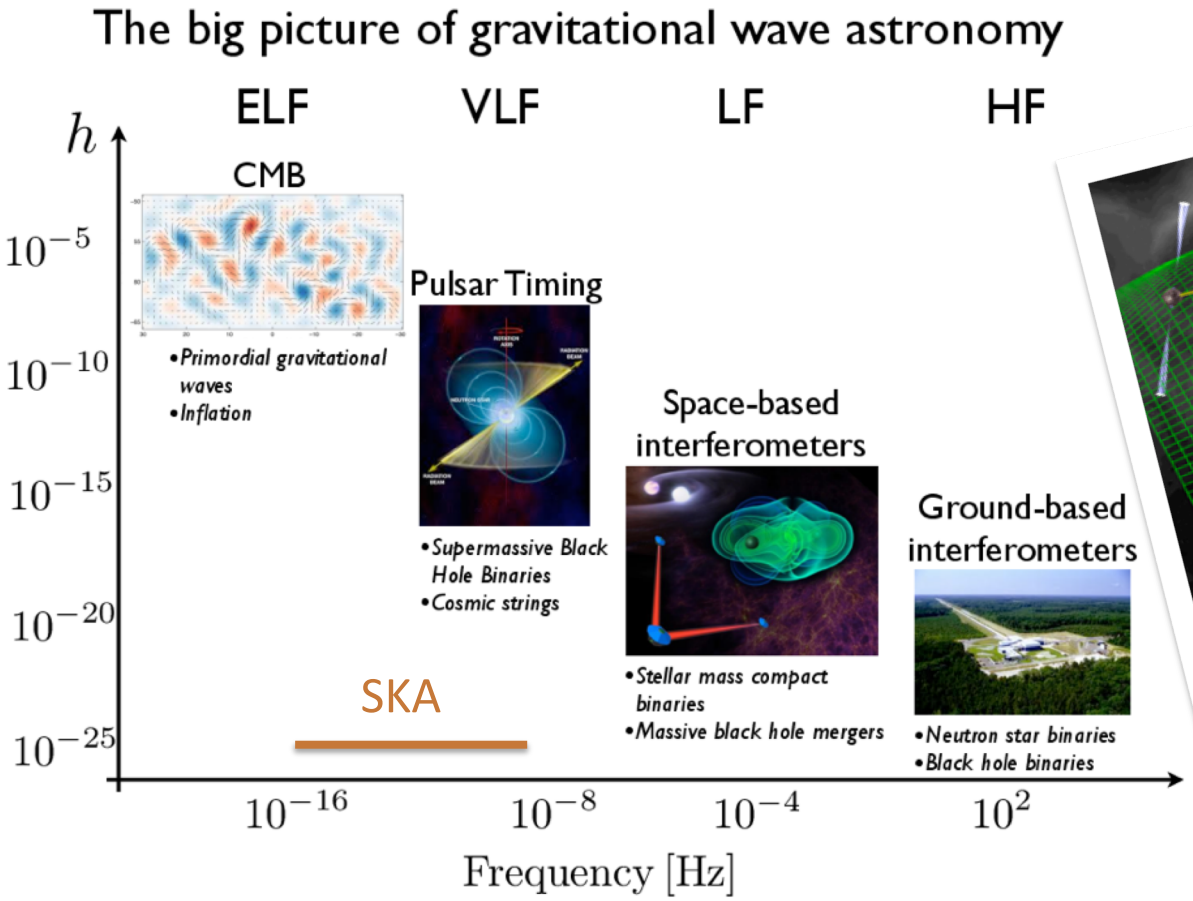
3. The community of scientists working on SKA science data will be geographically distributed

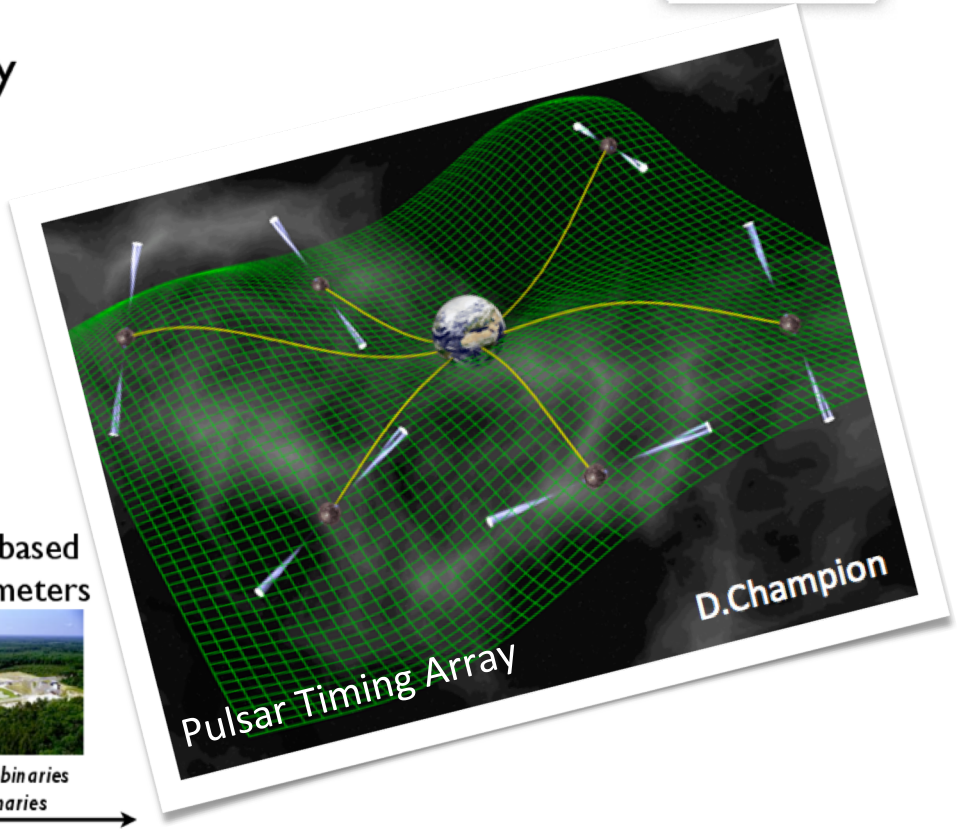
SKA Science Data Processor (SDP)

- Design work has focused on overall SDP architecture:
  - Not “what is it” but “what does it need to do
- Design for:
  - Flexibility
  - Scalability
  - Upgradability
  - COTS (i.e. as generic as possible to enable competitive costings)
- Data driven architecture
- Decoupling of real-time, streaming processes and off-line batch processing – enable load balancing
- Dynamic scheduling of telescope
- SDP batch system is *loosely* coupled to telescope schedule

- Understand Requirements (in System Engineering) early
- Put off finalising design as long as possible
  - Staged approach to SDP roll out – de-risk
  - Full SKA1 SDP purchase not until 2024

Using Pulsars to detect Gravitational Waves





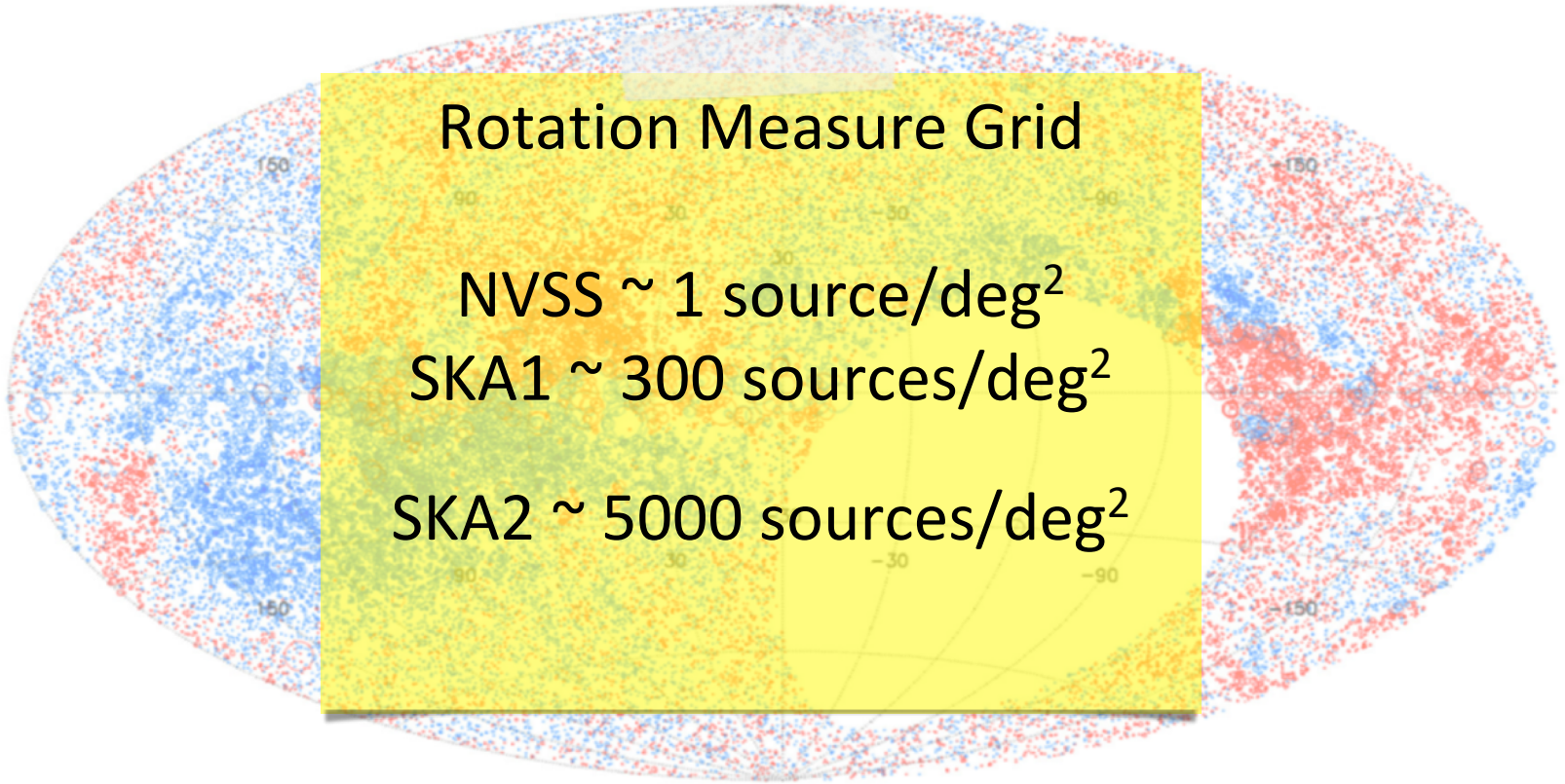
Advanced LIGO: NS-NS to 350 Mpc, NS-BH to 650 Mpc (z=0.15), BH-BH to z=0.4 (1500 Mpc)

PTAs – SMBH mergers in early galaxy evolution

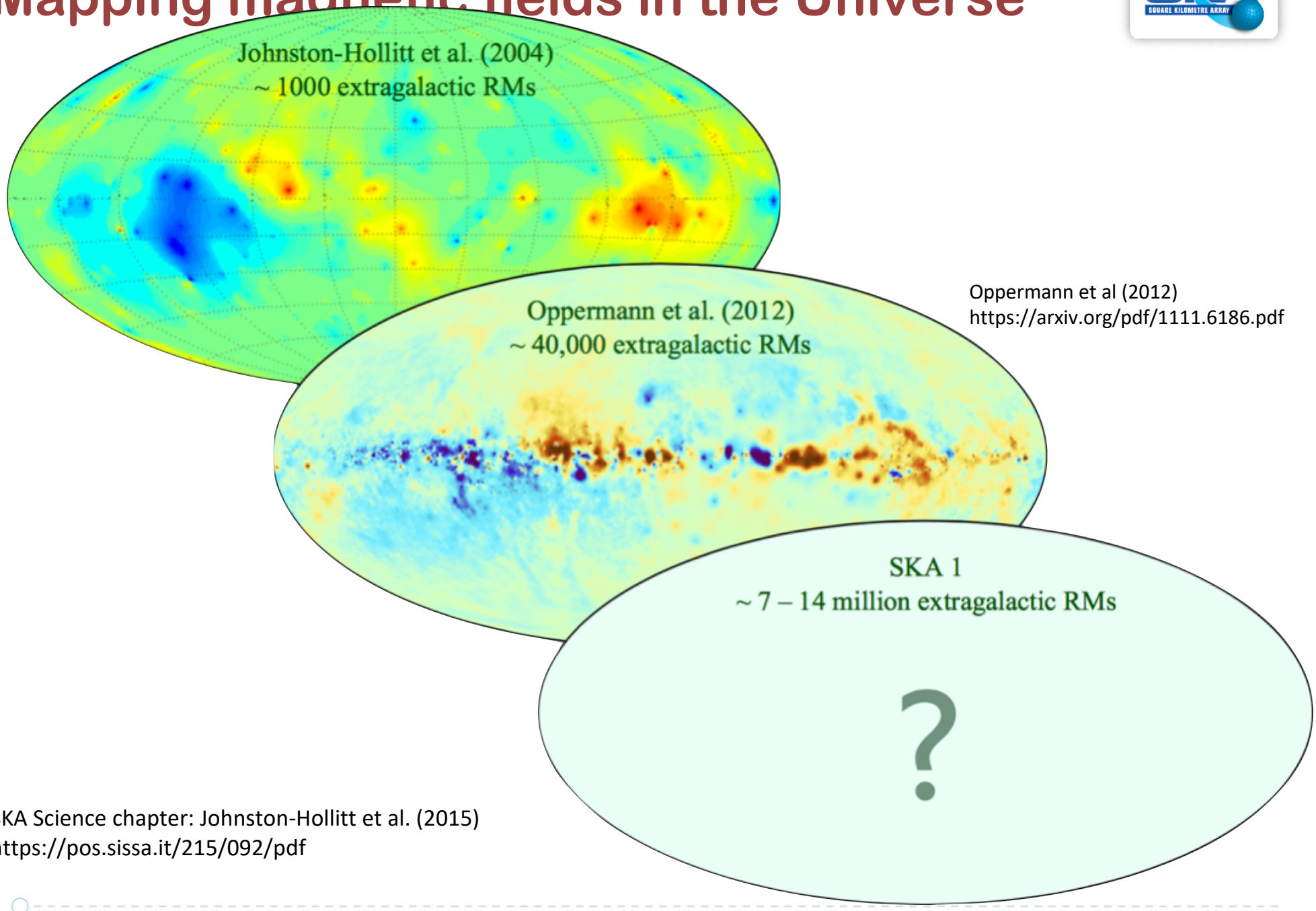
Mapping magnetic fields in the Universe

Measure the magnetic field structure in our Galaxy, nearby galaxies, galaxy clusters and the cosmic web

- Turbulent component of Galactic field and small-scale structure → 3D structure
- Structure of fields in nearby galaxies (along with direct imaging) → Field Origin
- Structure of fields in and around galaxy clusters → their role in cluster dynamics



Mapping magnetic fields in the Universe



Oppermann et al (2012)
https://arxiv.org/pdf/1111.6186.pdf

SKA Science chapter: Johnston-Hollitt et al. (2015)
https://pos.sissa.it/215/092/pdf

Transients and the SKA

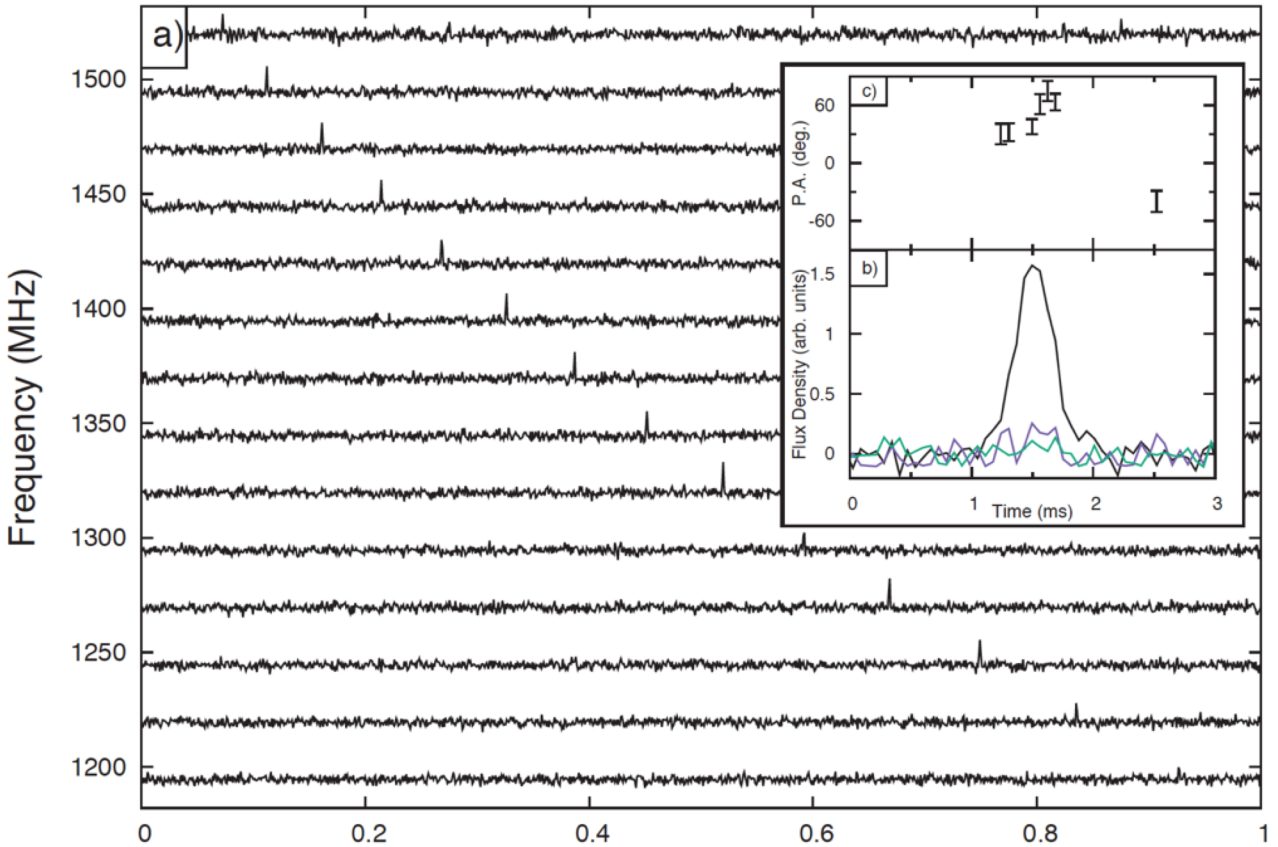
Pulsars and Transients are an important part of the SKA science case

- High Priority Science Objectives (HPSOs)

- Pulsar Searches
- Pulsar Timing
- Transients – FRBs

- Precision Astrophysics

- Gravitational Waves



Keane et al., 2016, Nature, 530, 453

Transients and the SKA

There is a big emphasis on Pulsars and Transients in the SKA science case

- High Priority Science Objectives (HPSOs)

- Pulsar Searches
- Pulsar Timing
- Transients – FRBs

- Precision Astrophysics

- Gravitational Waves

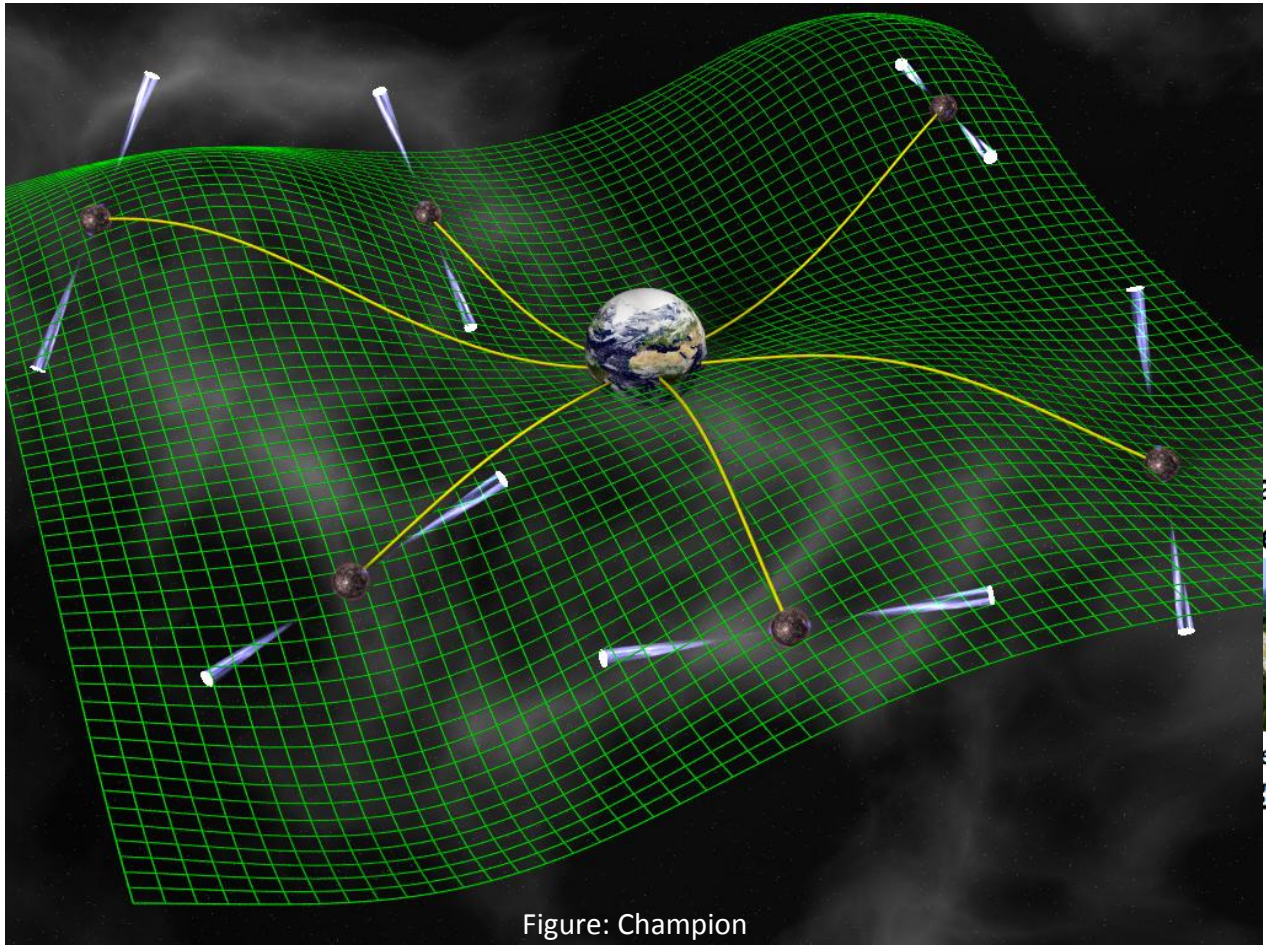


Figure: Champion

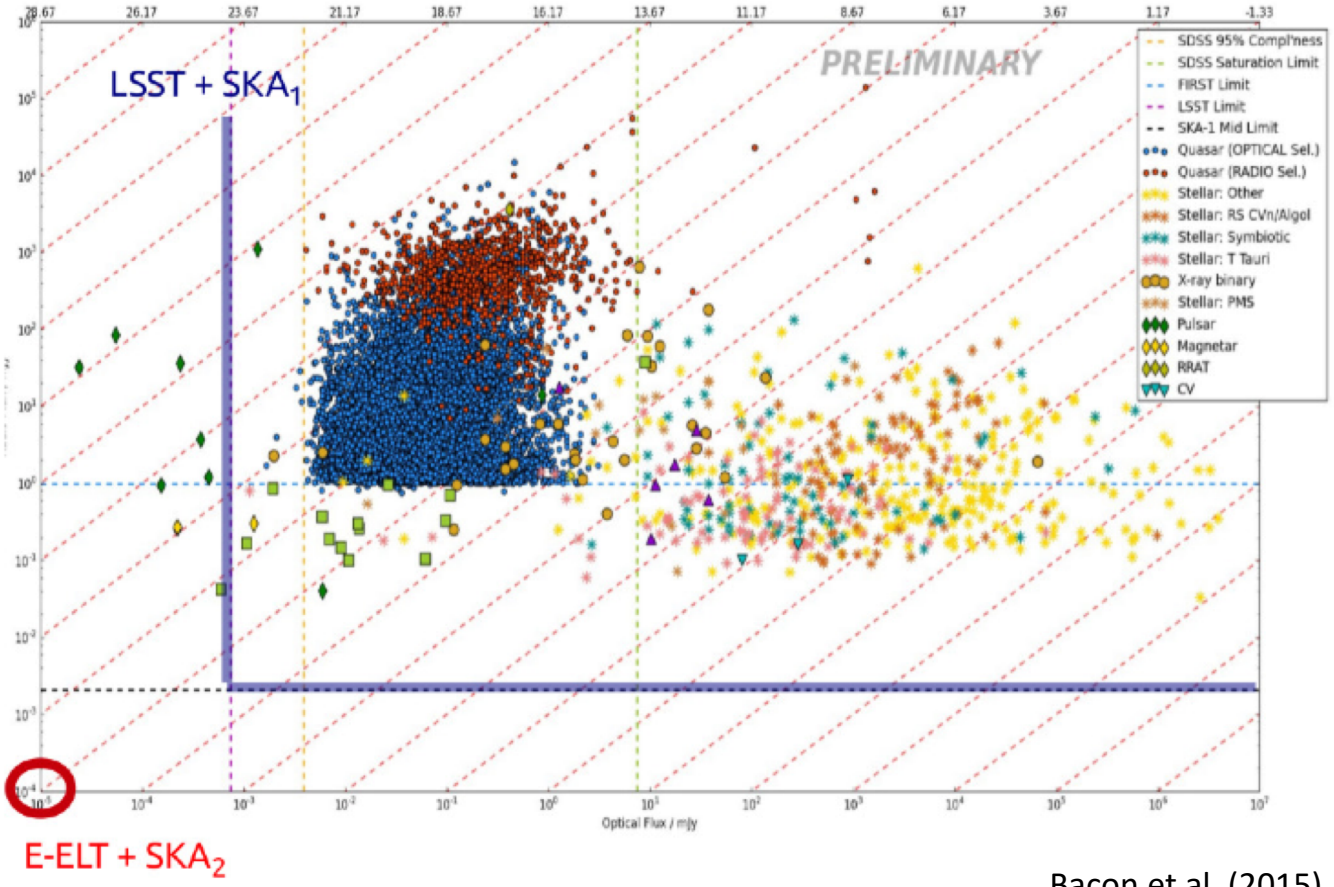
Synergies: SKA & LSST

SKA and LSST provide complementary physical constraints.

- Removal of systematics
- Cross checks of results
- Mutual support (e.g. redshifts – photometric and HI)
- A more complete picture (e.g. galaxy evolution)
- Exploitation of the time domain.

Main themes:

- Cosmology and large-scale structure
- Galaxy evolution
- Time domain



Bacon et al. (2015)
https://pos.sissa.it/215/145/pdf