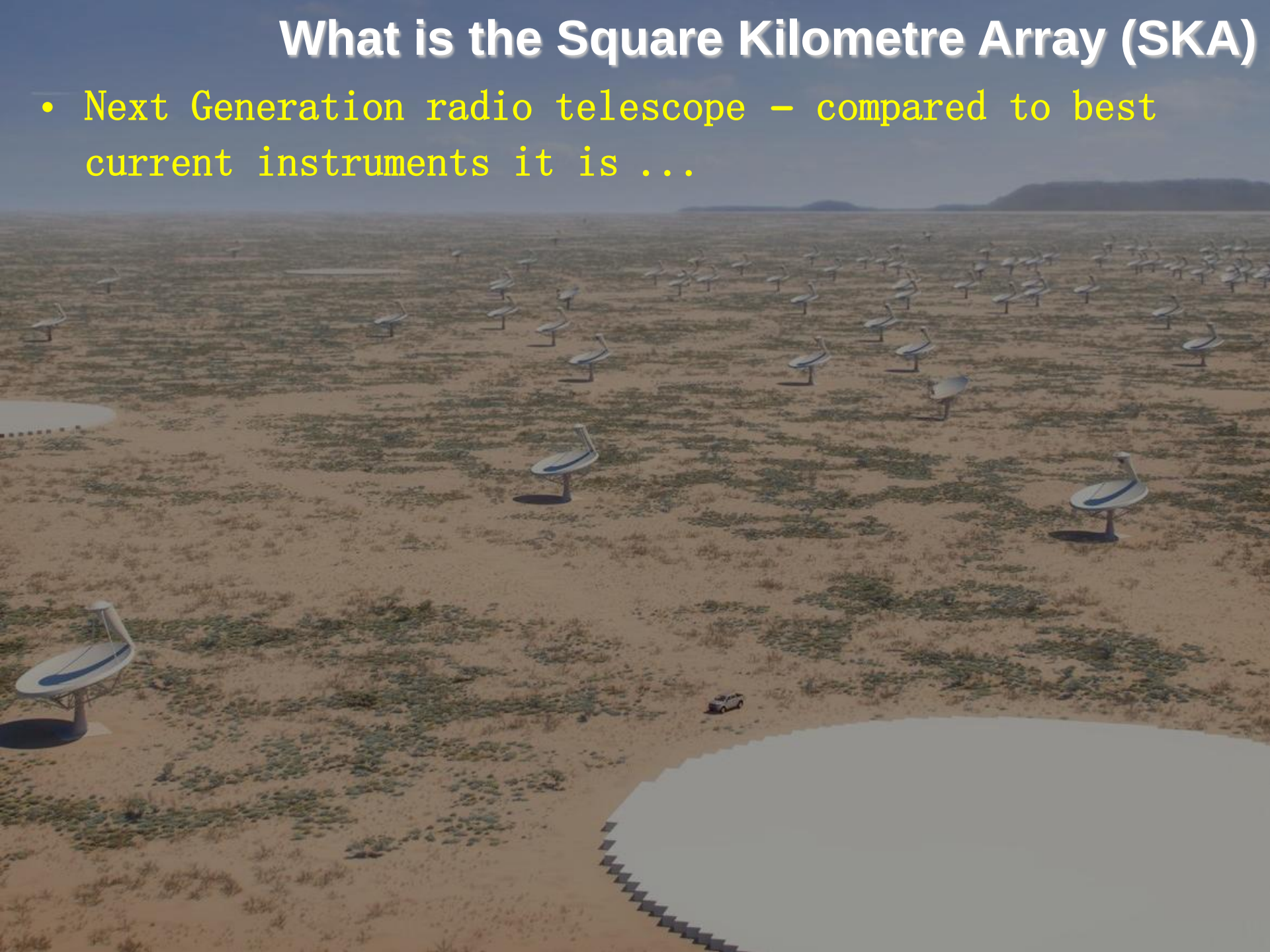
The Square Kilometre Array

Paul Alexander

UK Science Director the SKA Organisation
Leader the Science Data Processor Consortium

What is the Square Kilometre Array (SKA)

- Next Generation radio telescope – compared to best current instruments it is ...



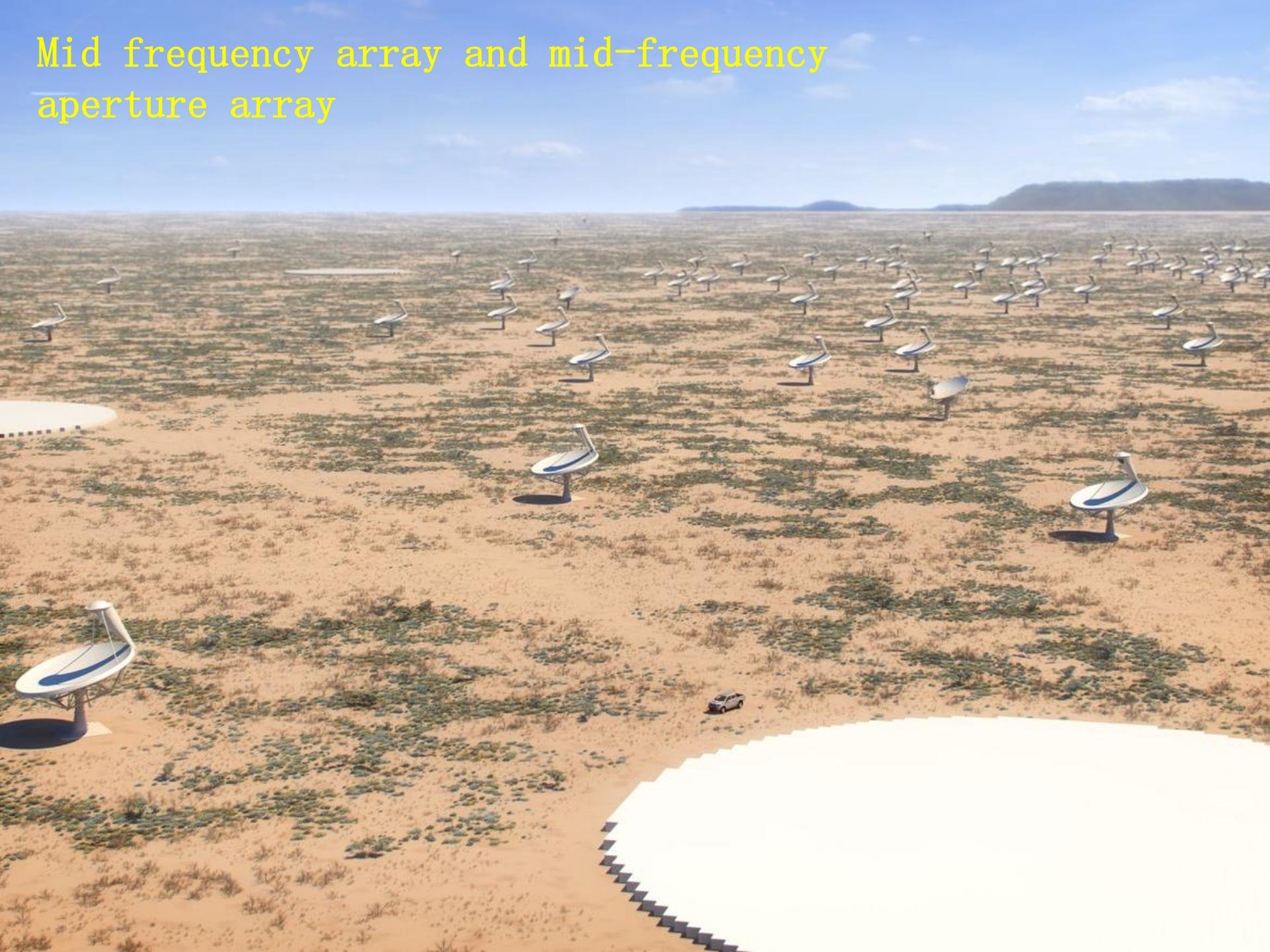
What is the Square Kilometre Array (SKA)

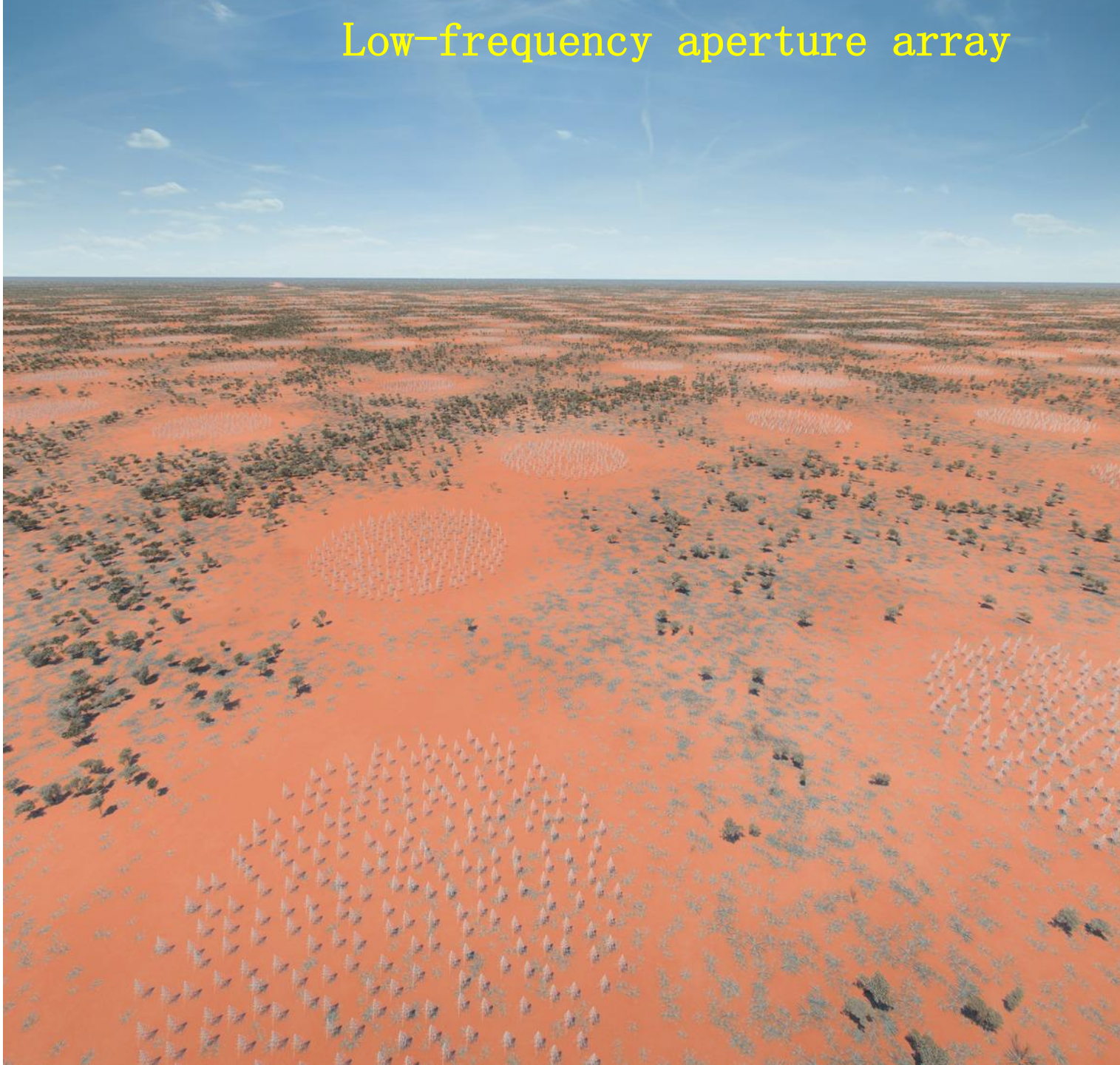
- Next Generation radio telescope – compared to best current instruments it is ...
  - ~100 times sensitivity
  - ~ 10^6 times faster imaging the sky
  - More than 5 square km of collecting area on sizes 3000km

What is the Square Kilometre Array (SKA)

- Next Generation radio telescope – compared to best current instruments it is ...
  - ~100 times sensitivity
  - ~ 10^6 times faster imaging the sky
  - More than 5 square km of collecting area on sizes 3000km
- Will address some of the key problems of astrophysics and cosmology (and physics)
- Builds on techniques developed in Europe
  - It is an interferometer
- Uses innovative technologies...
  - Major ICT project
  - Need performance at low unit cost

Mid frequency array and mid-frequency aperture array



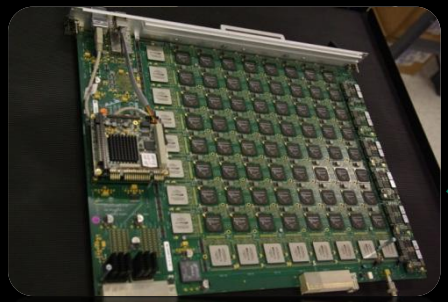

Low-frequency aperture array

SKA: A Leading Big Data Challenge for 2020 decade

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

Antennas

Digital Signal Processing (DSP)



Transfer antennas to DSP
2020: 5,000 PBytes/day
2030: 100,000 PBytes/day

Over 10's to 1000's kms

HPC Processing
2020: 300 PFlop
2028: 30 EFlop

To Process in HPC
2020: 50 PBytes/day
2030: 10,000 PBytes/day

Over 10's to 1000's kms

High Performance Computing Facility (HPC)

SKA Key Science Drivers

ORIGINS

- Neutral hydrogen in the universe from the Epoch of Re-ionisation to now

  When did the first stars and galaxies form?
  How did galaxies evolve?
  Role of Active Galactic Nuclei
  Dark Energy, Dark Matter

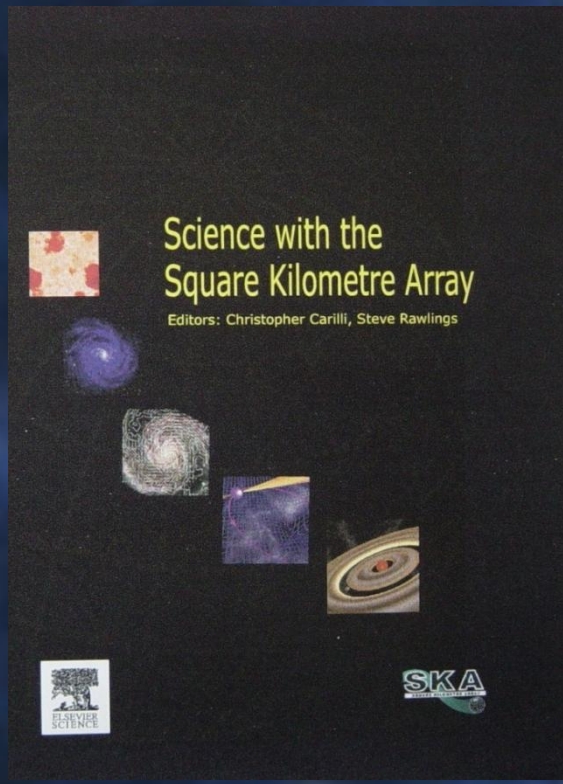
- Cradle of Life

FUNDAMENTAL FORCES

- Pulsars, General Relativity & gravitational waves

- Origin & evolution of cosmic magnetism

TRANSIENTS (NEW PHENOMENA)

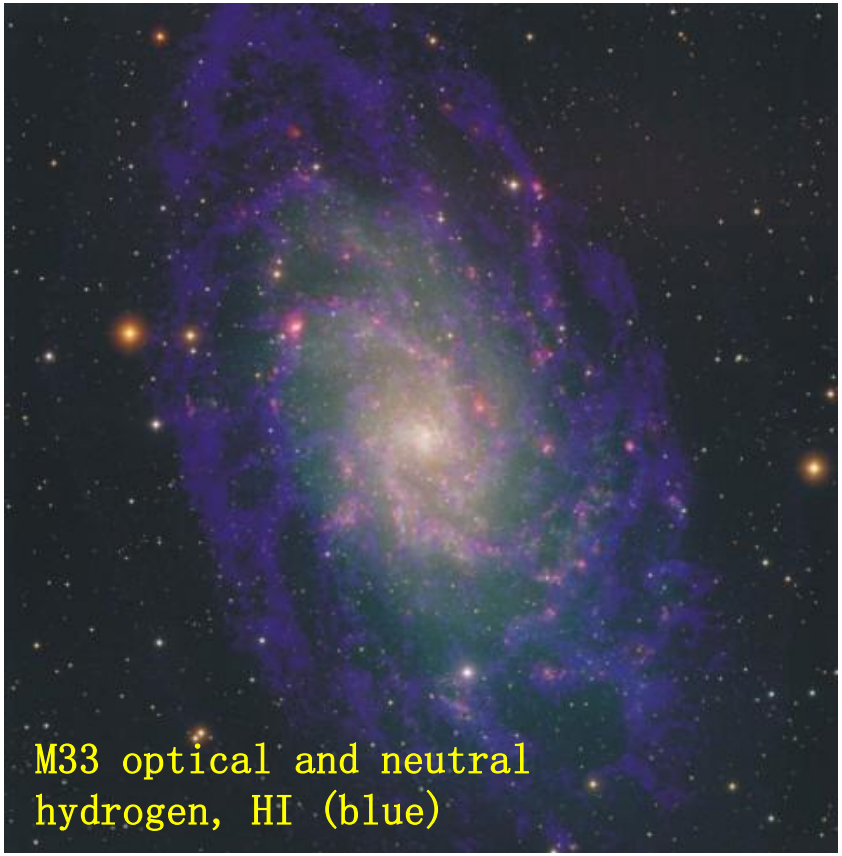


Science with the Square Kilometre Array (2004, **eds. C. Carilli & S. Rawlings, *New Astron. Rev.*, 48**)

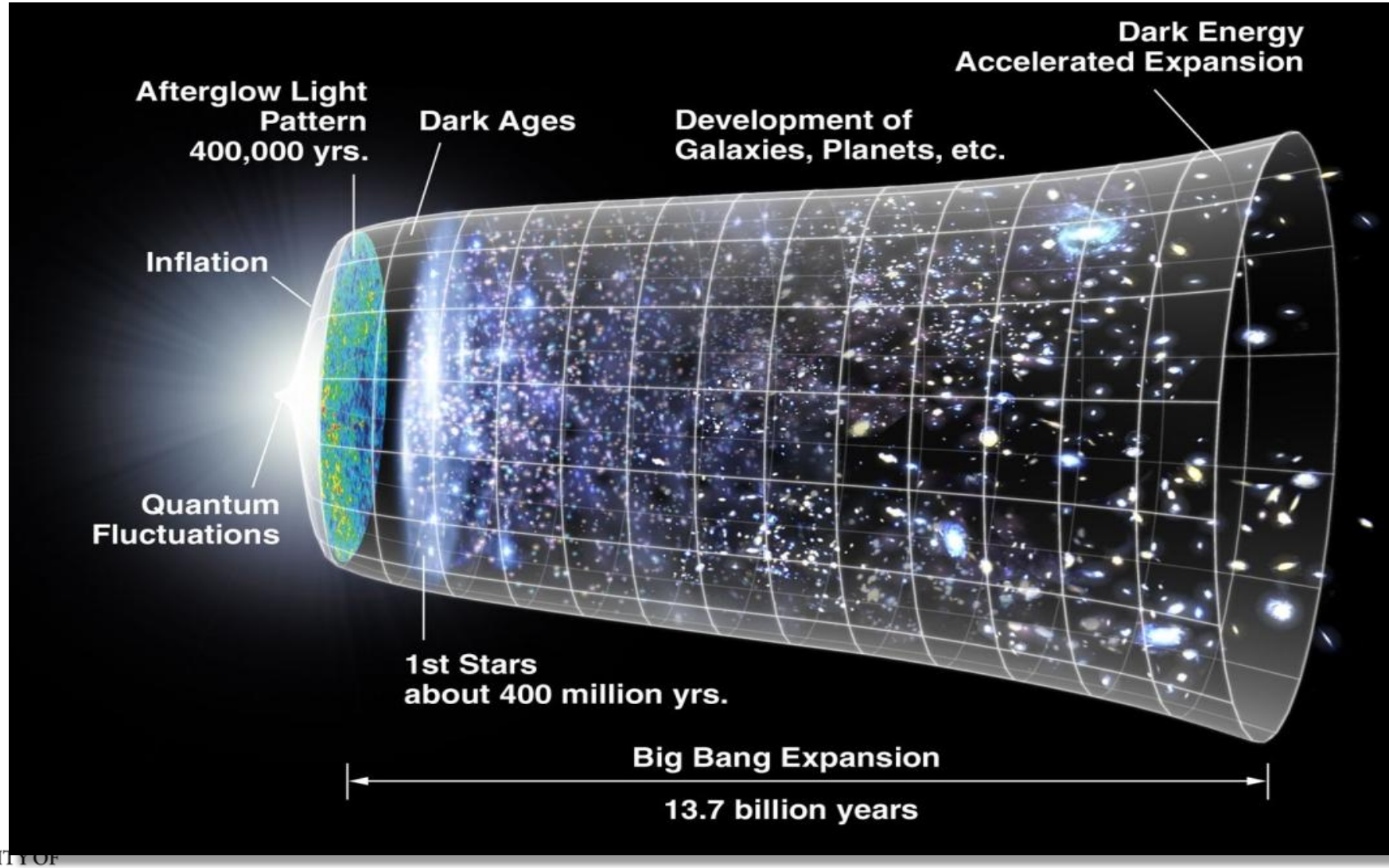
Cosmology & the History of Hydrogen

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

- Hydrogen is the main raw material for building galaxies stars and planetary systems
- Radio astronomy is our tool for observing hydrogen gas via a spectral line at 1420 MHz or a wavelength of 21cm
- Due to finite light travel time distant universe is observed now as it was in the past
  - Spectral line are redshifted to longer wavelengths or lower frequencies.
  - Build telescope to cover frequency range 50-1420 MHz and we can observe the evolution of the formation of structure



21-cm line at 1420 MHz.


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
400,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.7 billion years

SKA SQUARE KILOMETRE ARRAY

SDP

SCIENCE DATA PROCESSOR

Afterglow Light Pattern 400,000 yrs.

Dark Ages

Development of Galaxies, Planets, etc.

Dark Energy Accelerated Expansion

Inflation

Quantum Fluctuations

1st Stars about 400 million yrs.

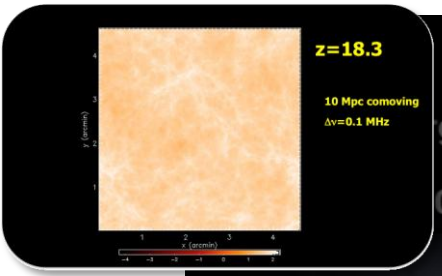
Big Bang Expansion

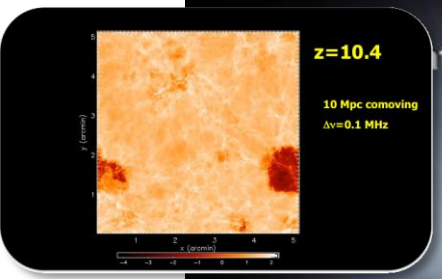
13.7 billion years

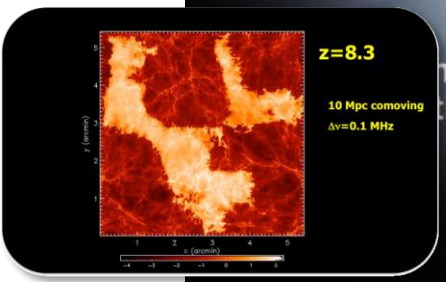
SKA SQUARE KILOMETRE ARRAY

SDP

SCIENCE DATA PROCESSOR







glow Light Pattern

Dark Ages

00,000 yrs.

tion

ntum tions

Dark Energy Accelerated Expansion

Development of Galaxies, Planets, etc.

1st Stars about 400 million yrs.

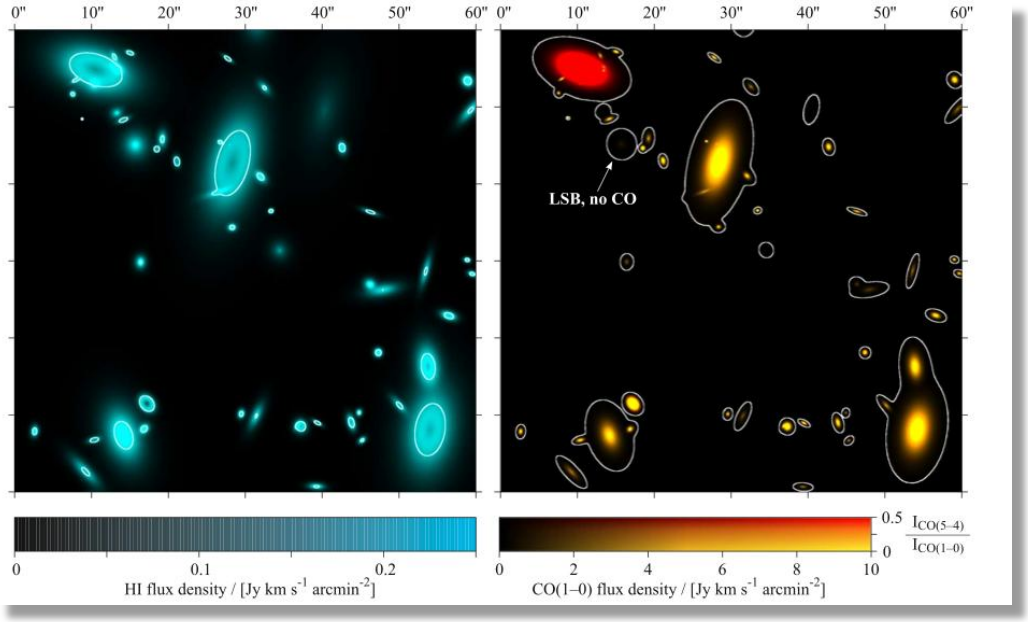
Big Bang Expansion

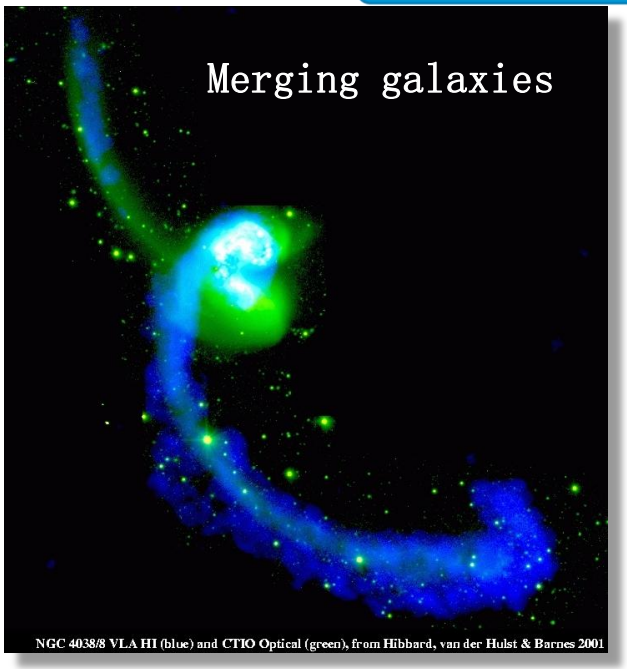
13.7 billion years

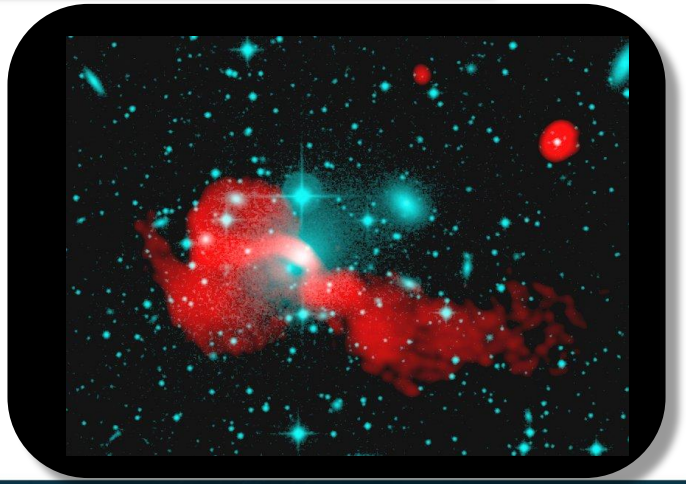
UNIVERSITY OF CAMBRIDGE

Galaxy Evolution

- **At the end of the EoR Neutral hydrogen is in galaxies – fuel for star formation**
- **SKA + ALMA will follow gas content**



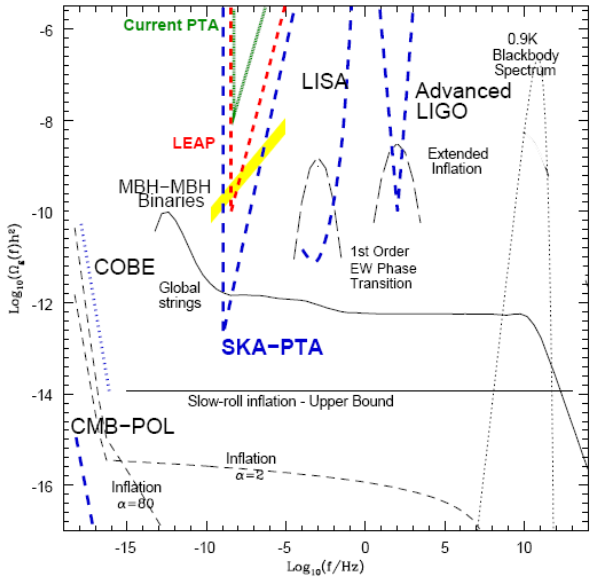


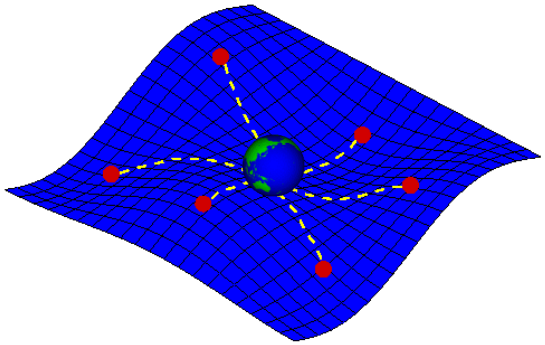


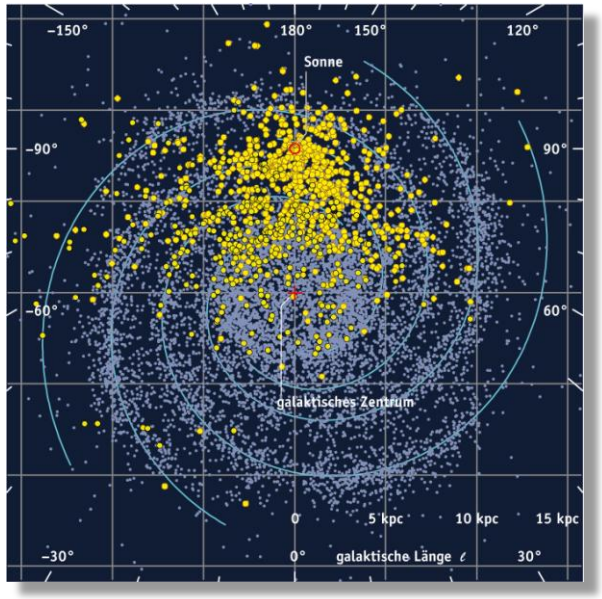
Evolution of gas star formation and AGN in galaxies

- Detect SF galaxies to z = 7 (25 M/yr)
- Distinguish AGN in galaxies to z ~ 7

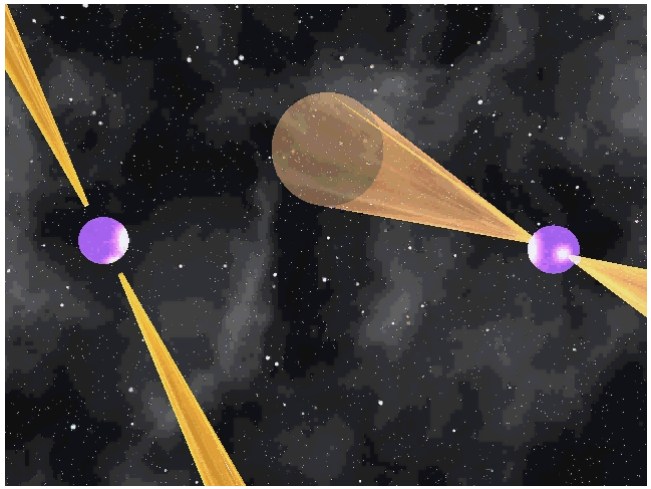
Pulsar as Natural Clocks: Testing gravity







- Pulsars are rotating neutron stars
- Pulse once per revolution → very accurate clocks
- The SKA will detect around 30,000 pulsars in the Galaxy
- Relativistic binaries to test gravity
- Timing net of to detect gravitational waves



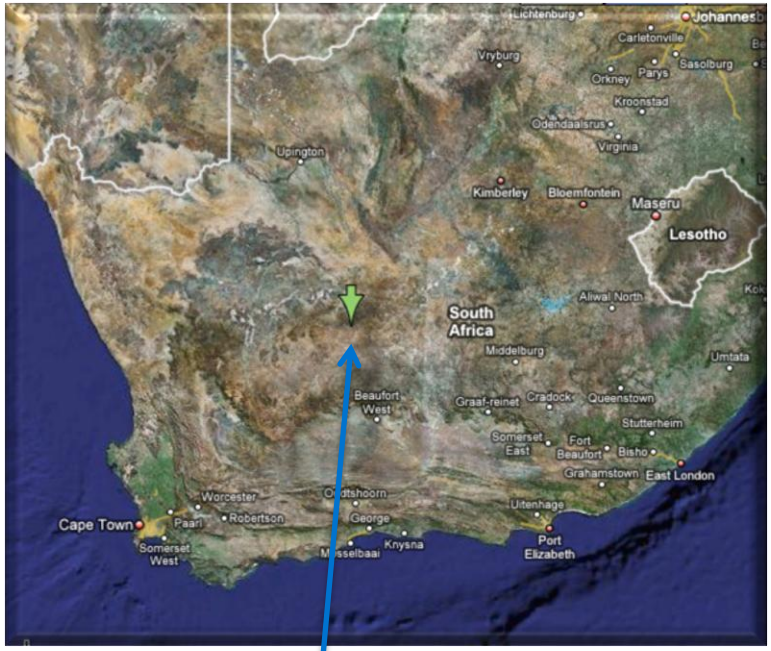
Major Global Partnership

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

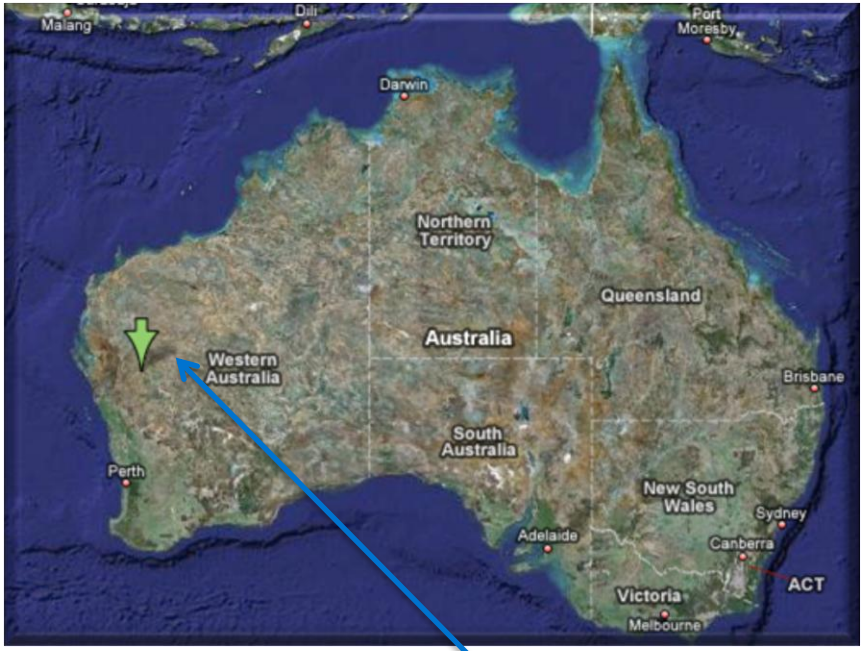
- 11 Countries:
  - UK, RSA, AUS, NZ, Canada, China, NL, Germany, Italy, India, Sweden
- Stage 1 Completion 2023
  - € 650m
- Stage 2 Completion 2030
  - € ?b
- Currently detailed design phase

SKA Phase 1 Implementation

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR



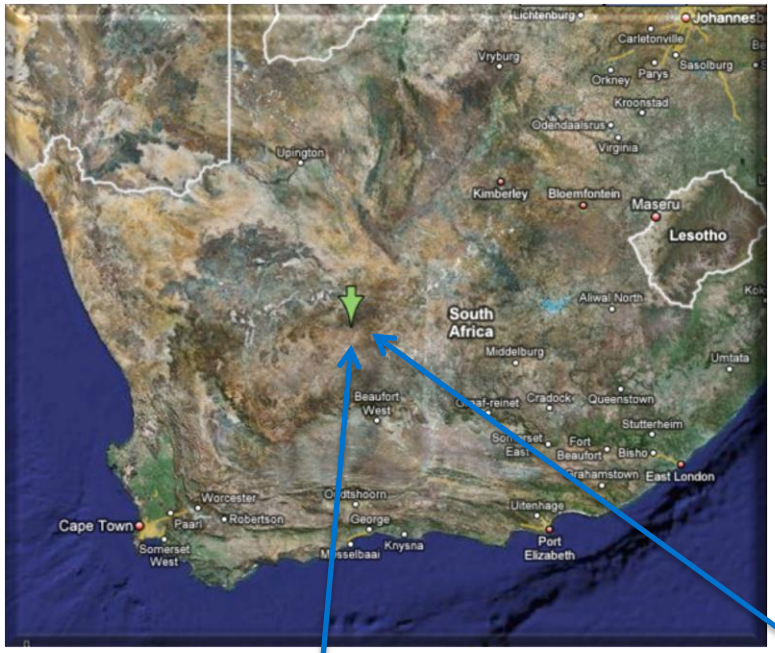
SKA1_Mid_Dish

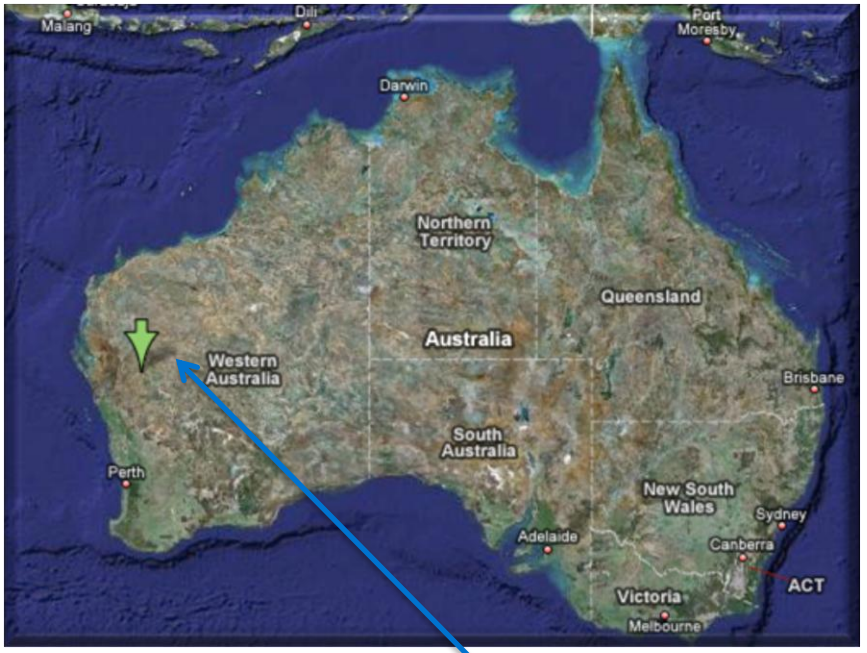


SKA1_Low

SKA Implementation

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR





SKA2_Mid_Dish

SKA2_AIP_AA

SKA2_Low

SKA Timeline

2020	Early science SKA_1	2023: Full Operations SKA_1	
2023-2030	Construction of Full SKA, SKA_2		€?B
2018-2022	10% SKA construction, SKA_1		€650M
2012	Site selection		✔
2013 - 2017	Pre-Construction: 4 yr PEP	Detailed design and production Readiness	€120M
2008 - 2012	System design and refinement of specification		✔
2000 - 2007	Initial concepts stage		✔
1995 - 2000	Preliminary ideas and R&D		✔

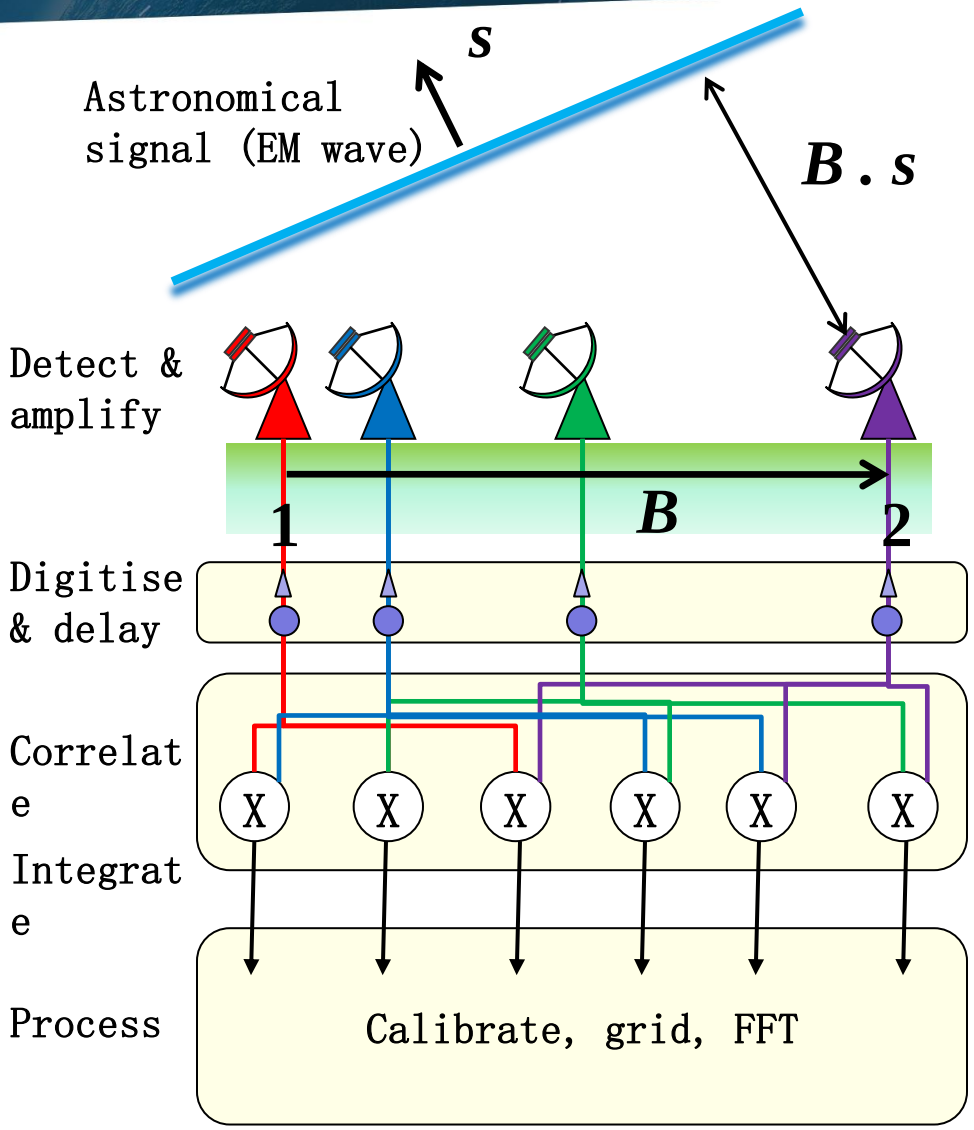
SKA SQUARE KILOMETRE ARRAY

SDP

SCIENCE DATA PROCESSOR

THE PROCESSING CHALLENGE

Standard interferometer



- Visibility:

$$V(B) = E_1 E_2^* = I(s) \exp(i \omega B.s/c)$$

- Resolution determined by maximum baseline

$$\theta_{max} \sim \lambda / B_{max}$$

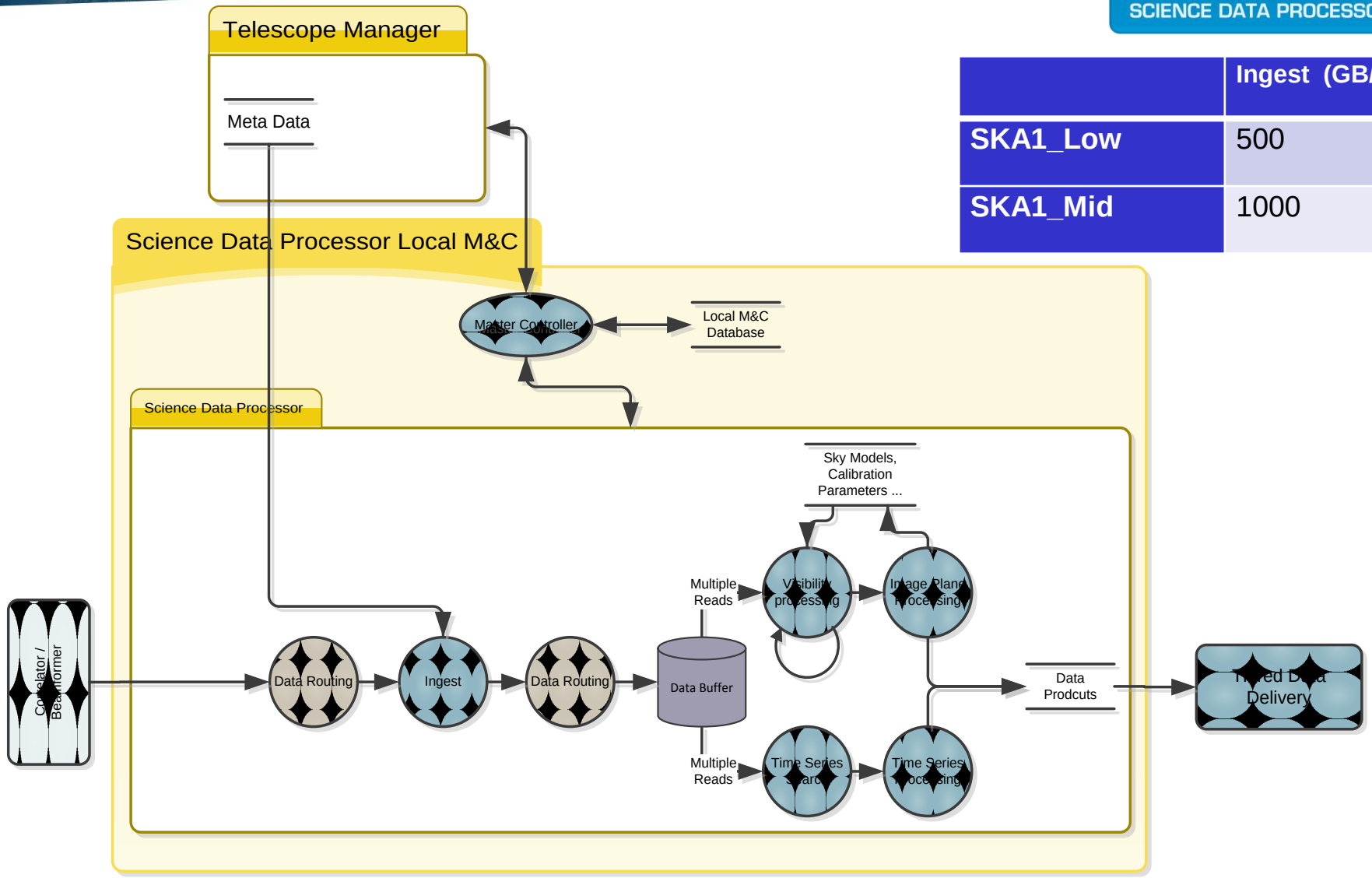
- Field of View (FoV) determined by the size of each dish

$$\theta_{dish} \sim \lambda / D$$

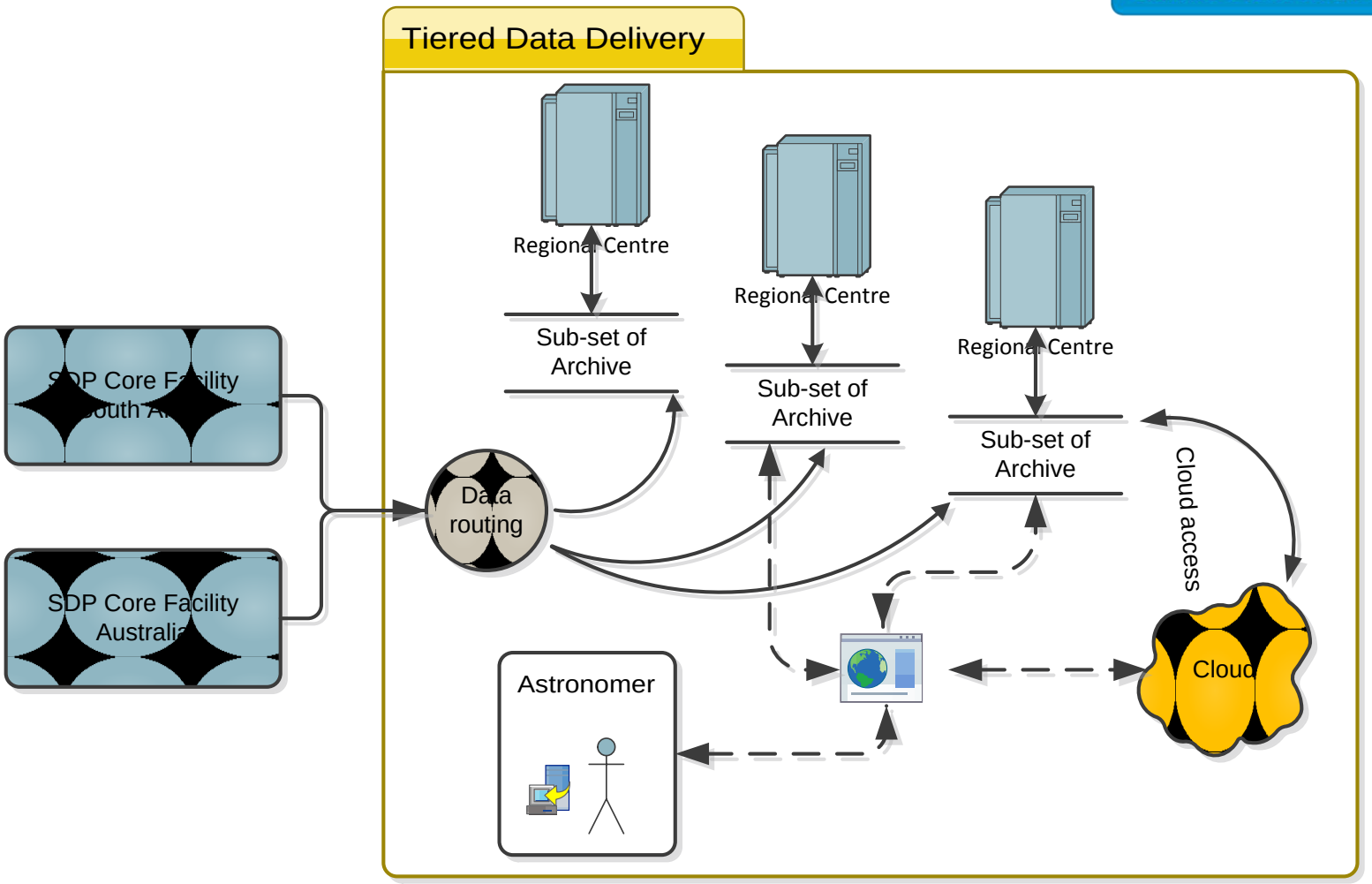
SKY Image

High Level Description

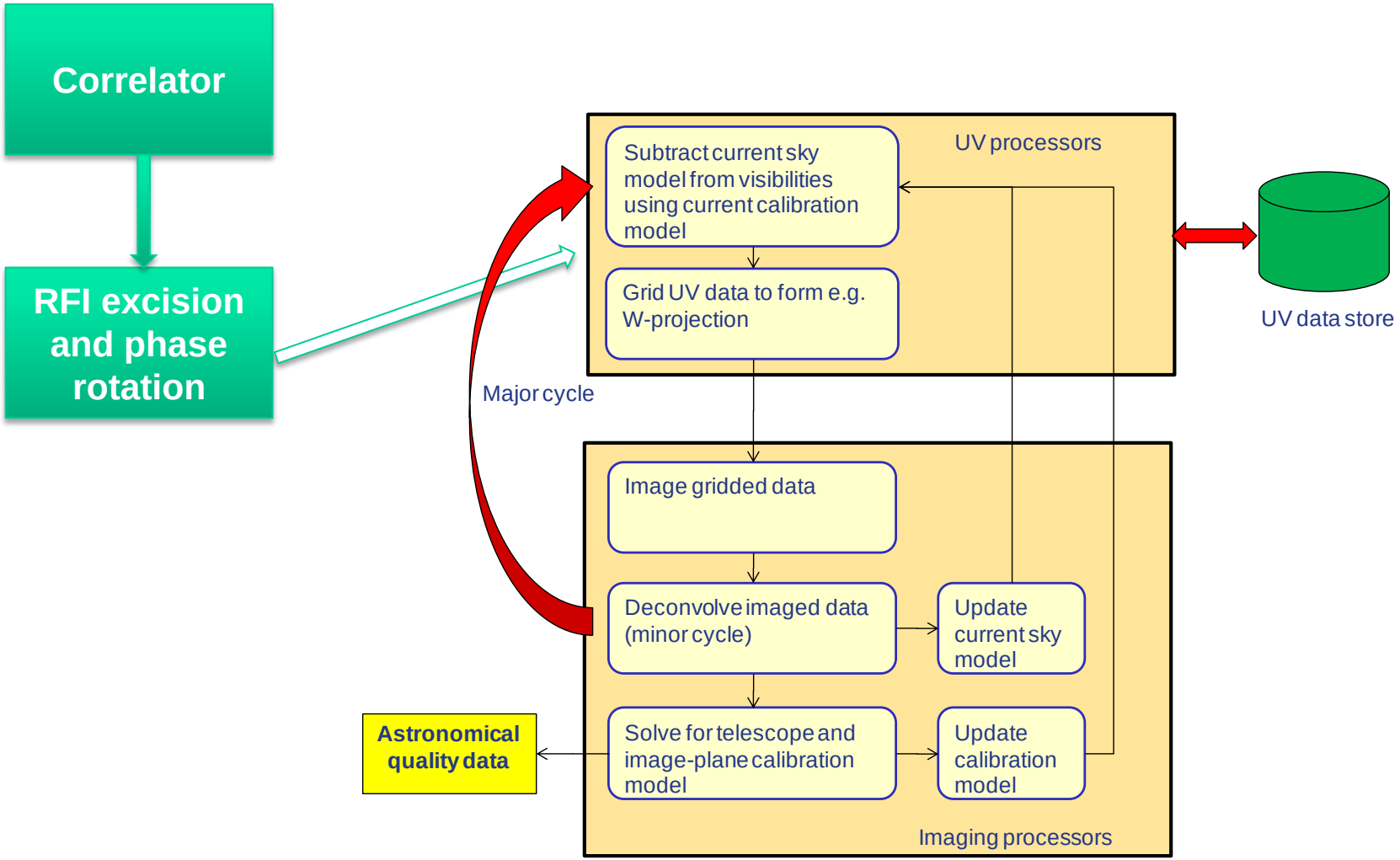
	Ingest (GB/s)
SKA1_Low	500
SKA1_Mid	1000



High Level Description



Imaging Processing Model



Performance Requirements

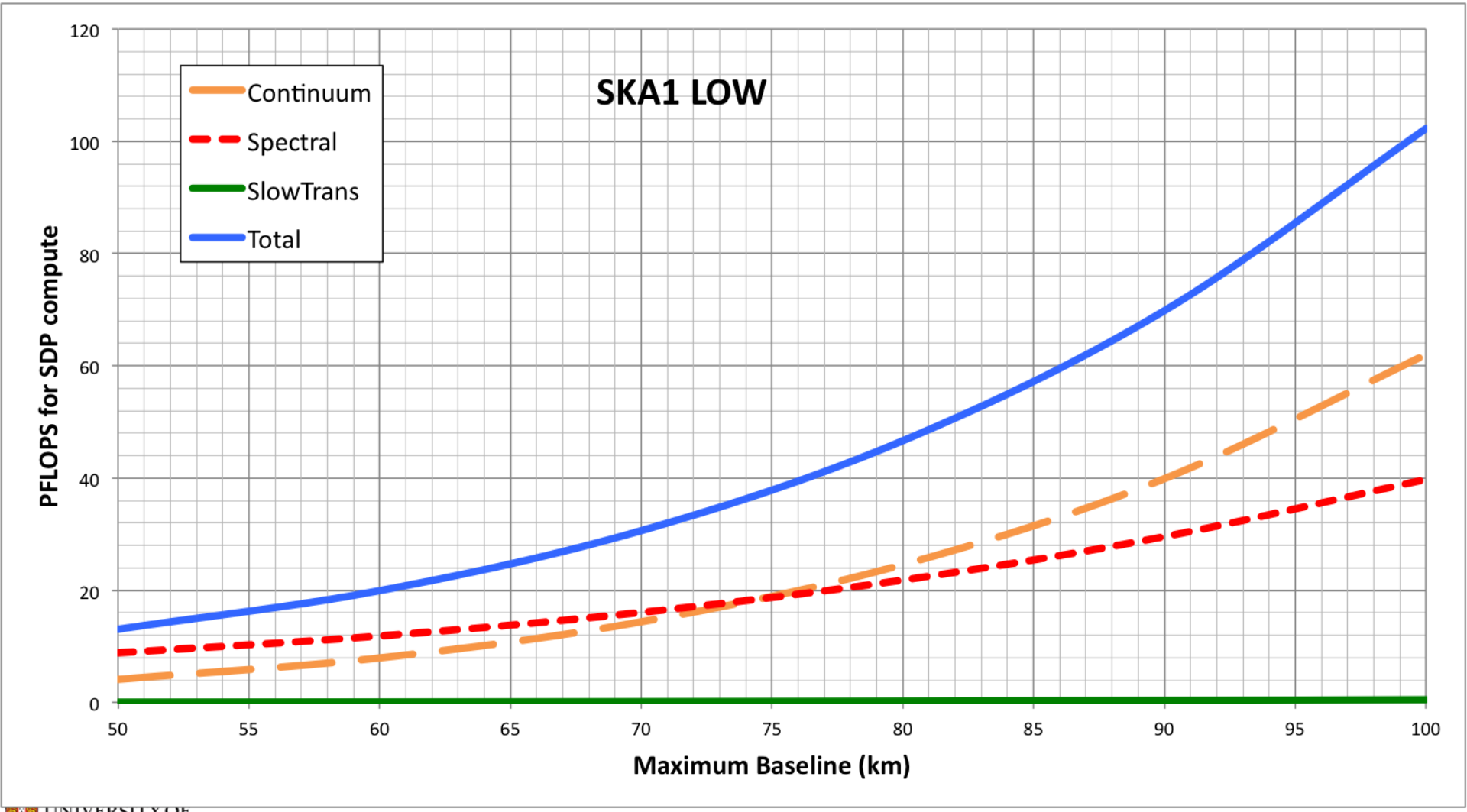
SCIENCE DATA PROCESSOR

- Imaging and calibration determines system sizing
- Data is Buffered after Ingest
  - Double buffered
  - Buffer size >100 PByte
- Imaging must account for very large field of view
  - Algorithms complex and computationally expensive
  - Will need to evolve algorithms during life of the telescope
- Data products will be calibrated multi-dimensional images and time-series data
  - Volume of potential data products very large
  - May only be able archive data specific to observation requested
- Commensal fast-imaging mode for slow transients

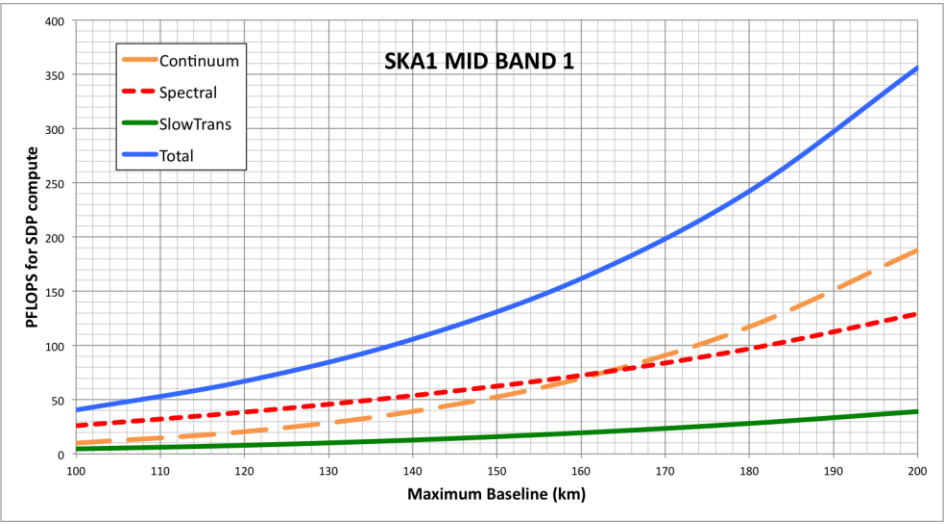
	Processing maximal (Pflop)	Ingest (GB/s)
SKA1_LOW	100	500
SKA1_Mid	360	1000

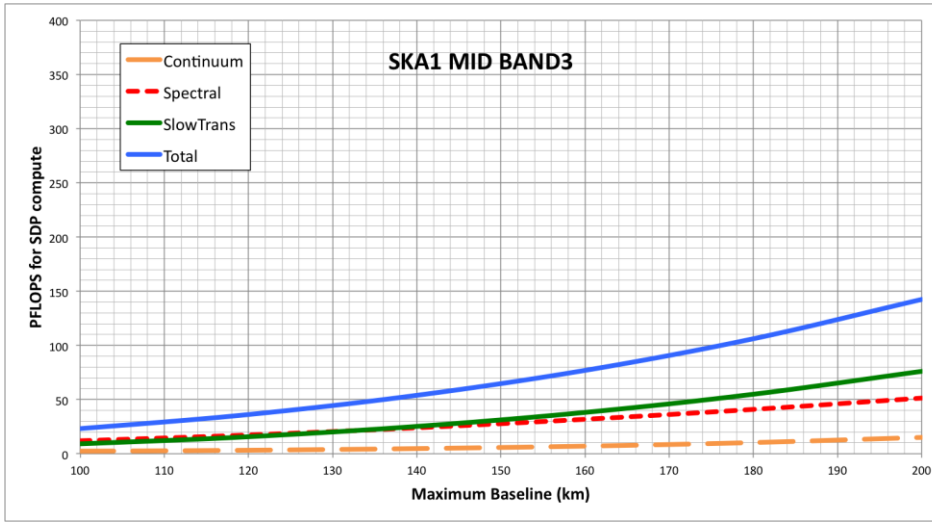
Detailed analysis is complex

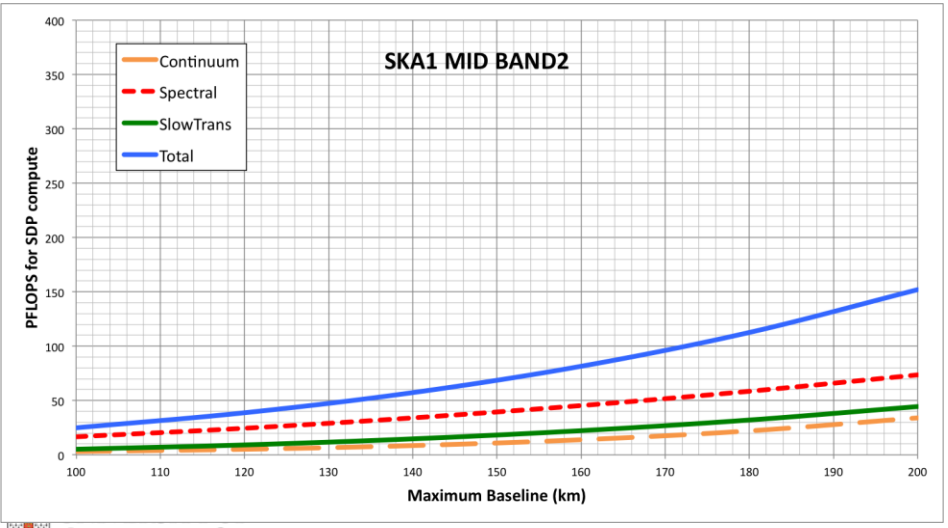
Processing a critical engineering driver
SKA1 Low

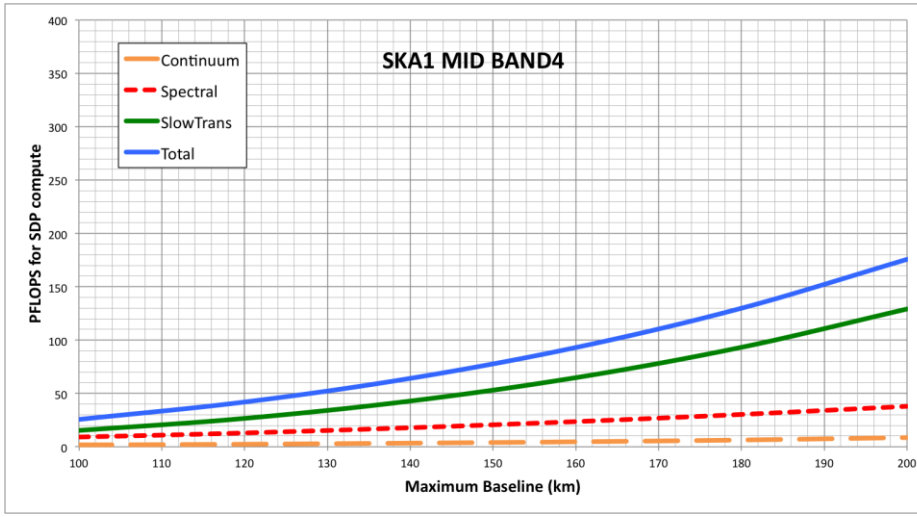


Processing a critical engineering driver
SKA1 Mid









Architectural Principles

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

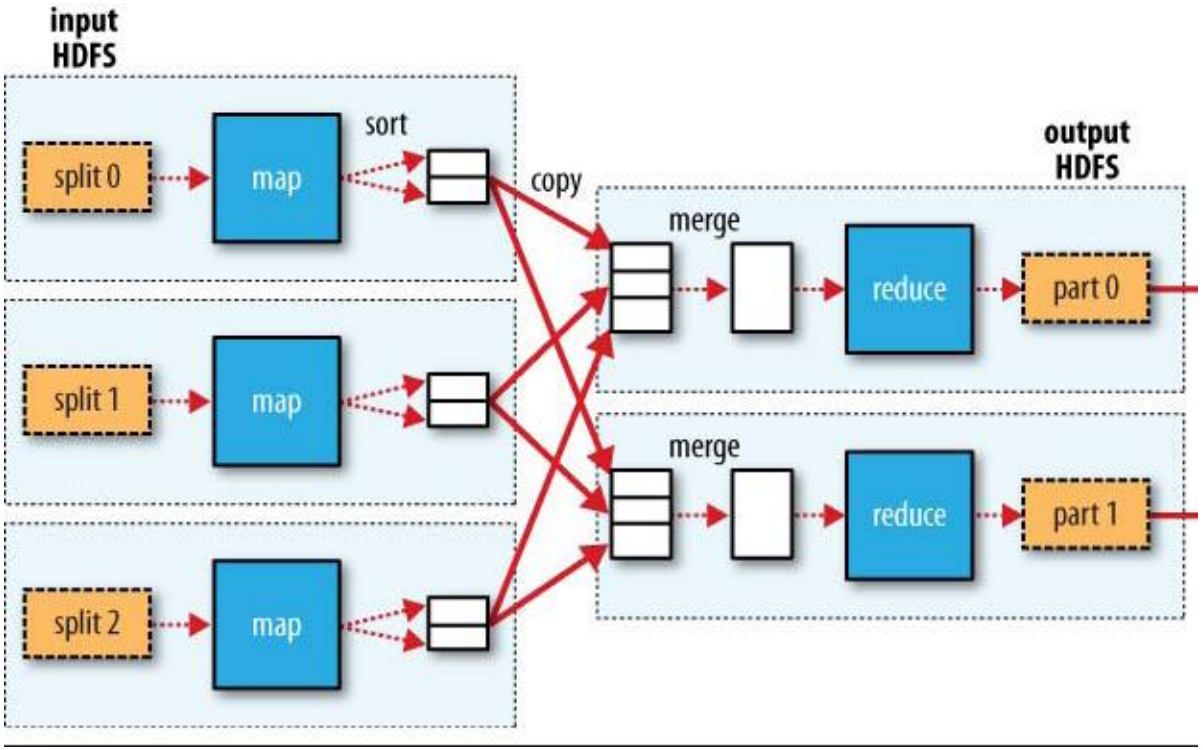
- Main principles:
  - **Ensure scalability**
  - **Ensure affordability**
  - **Ensure Maintainability**
  - **Support current state-of-the-art algorithms**

- Exploit data parallelism, not just in frequency but also other dimensions
  - We have only two fundamental/bulk data structures
  - Raster grids and key-value-value stream records [e.g. u,v,w, -> visibility]

- Emphasis is on the framework to manage the throughput
  - Hardware platform will be replaced on a short duty cycle c.f. any HPC facility
  - Algorithms and workflow will evolve as we learn about telescopes

Approach: Co-design of software and physical layer architectures

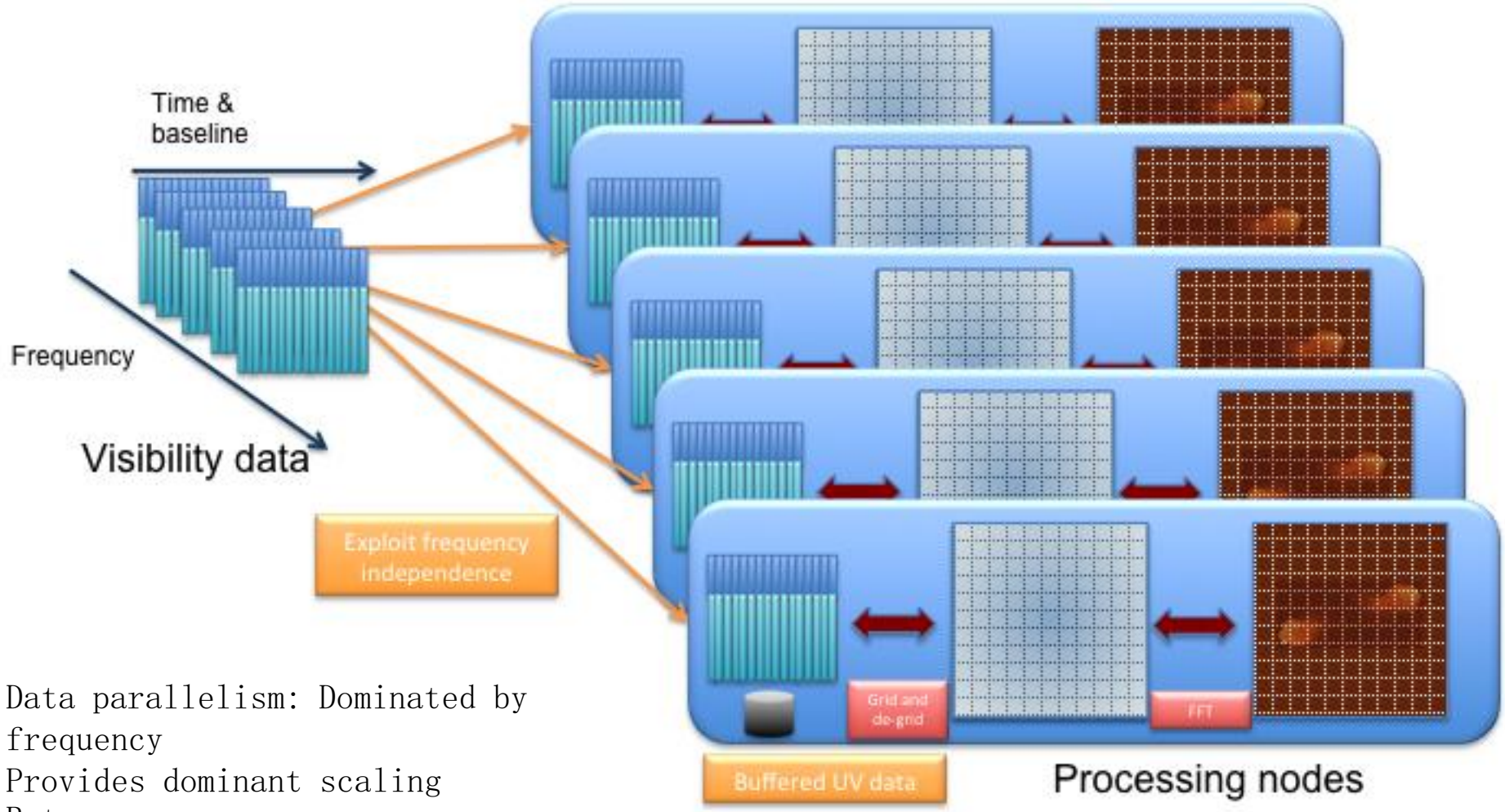
Data Driven Design: Hadoop



Approach: Build on BigData Concepts

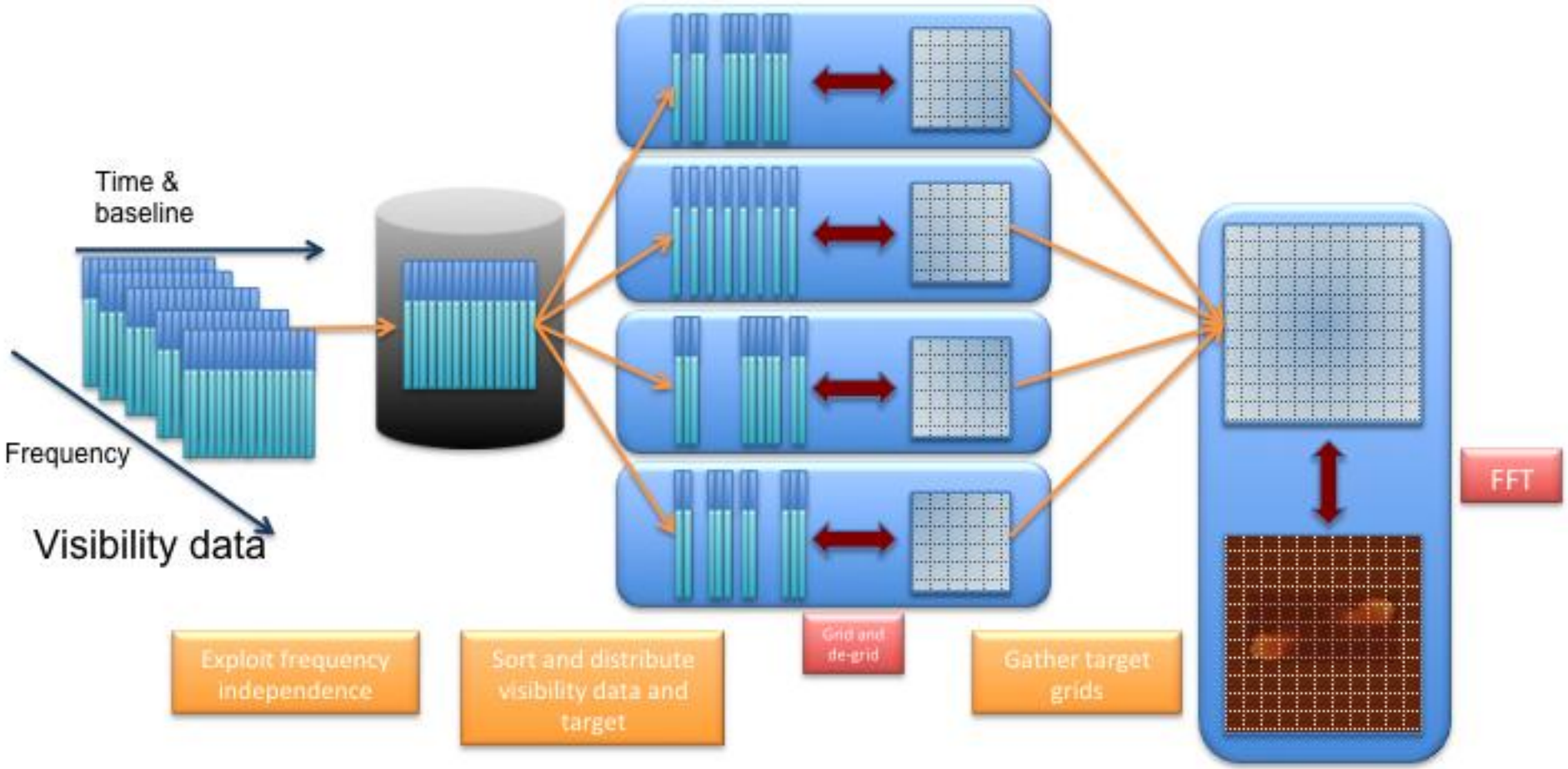
- **"data driven" → key element is the data flow manager**
- **Inspired by Hadoop but for our complex data flow**

Data Driven Architecture



- Data parallelism: Dominated by frequency
- Provides dominant scaling
- But …

Data Driven Architecture

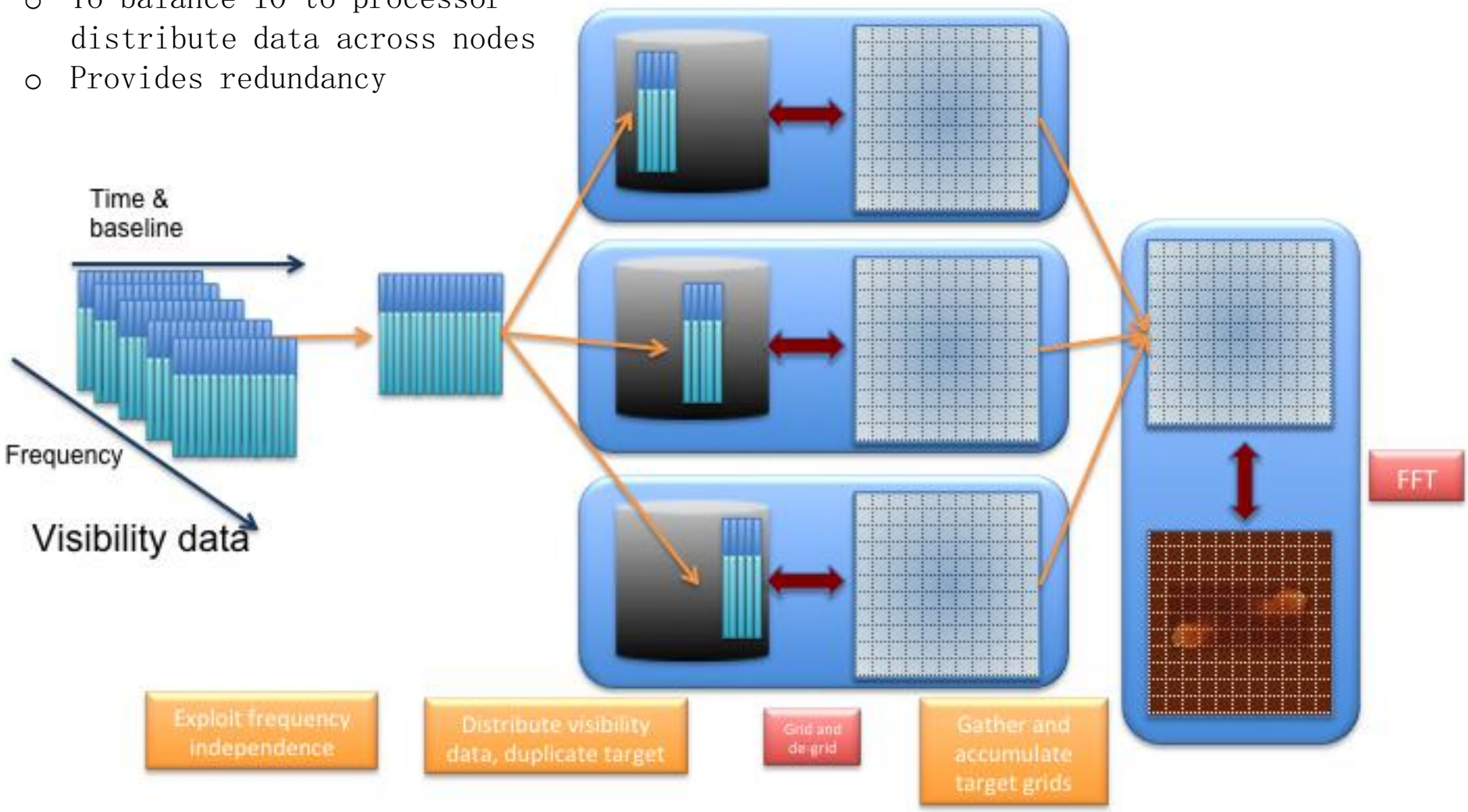


- Further data parallelism in spatial indexing (UVW-space)
- Use to balance memory bandwidth per node
- Some overlap regions on target grids needed

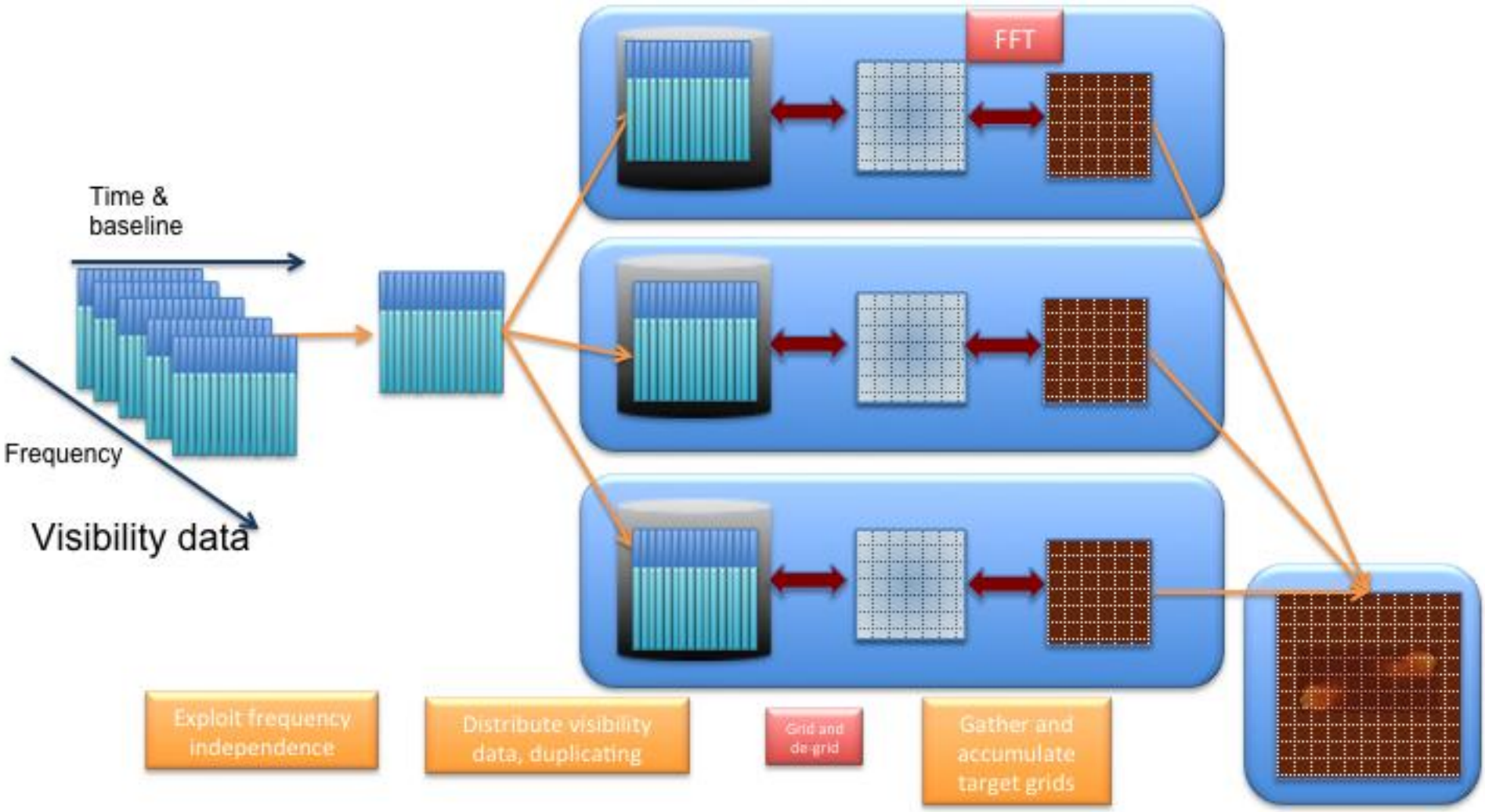
Data Driven Architecture

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

- To balance IO to processor distribute data across nodes
- Provides redundancy

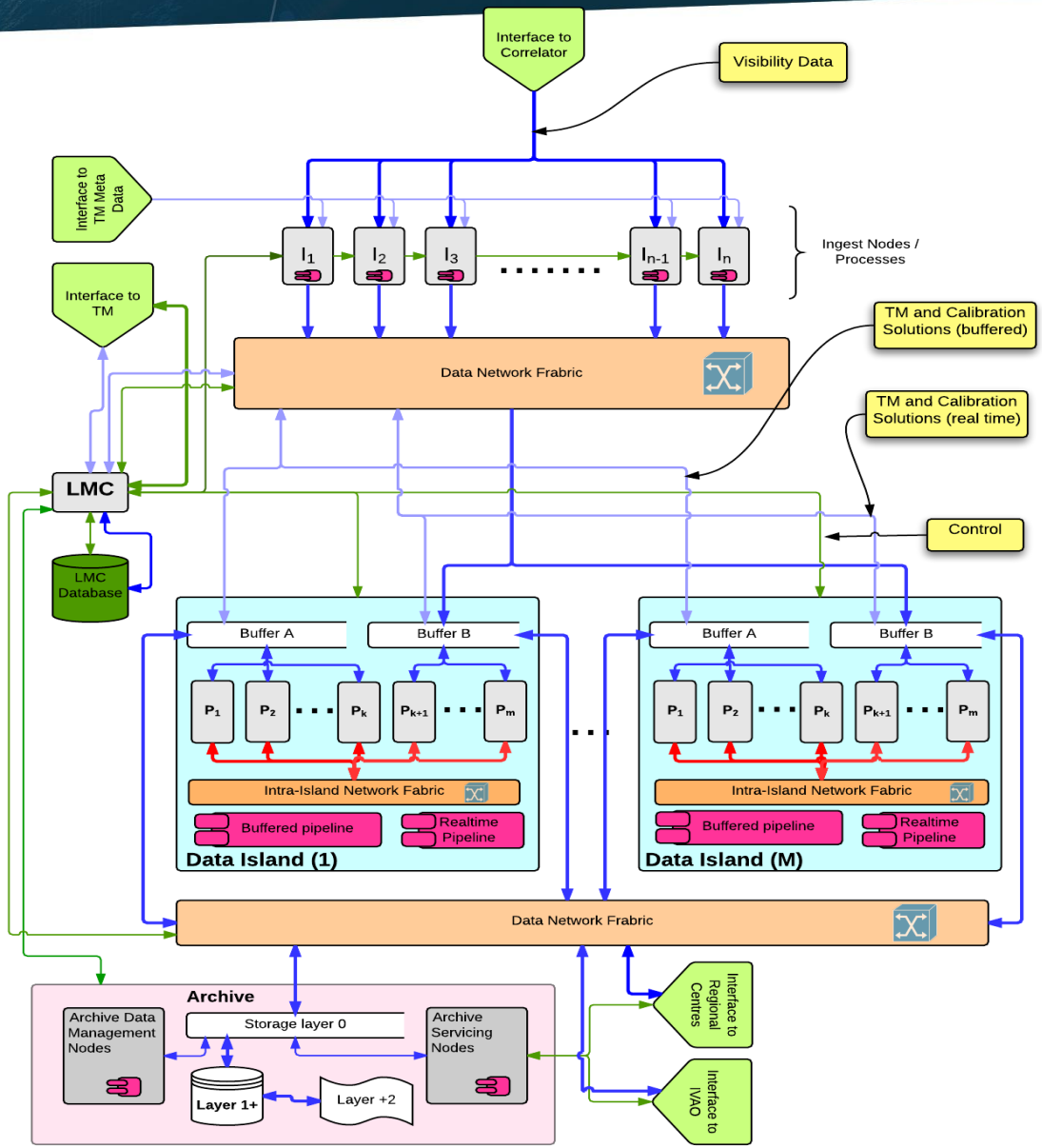


Data Driven Architecture

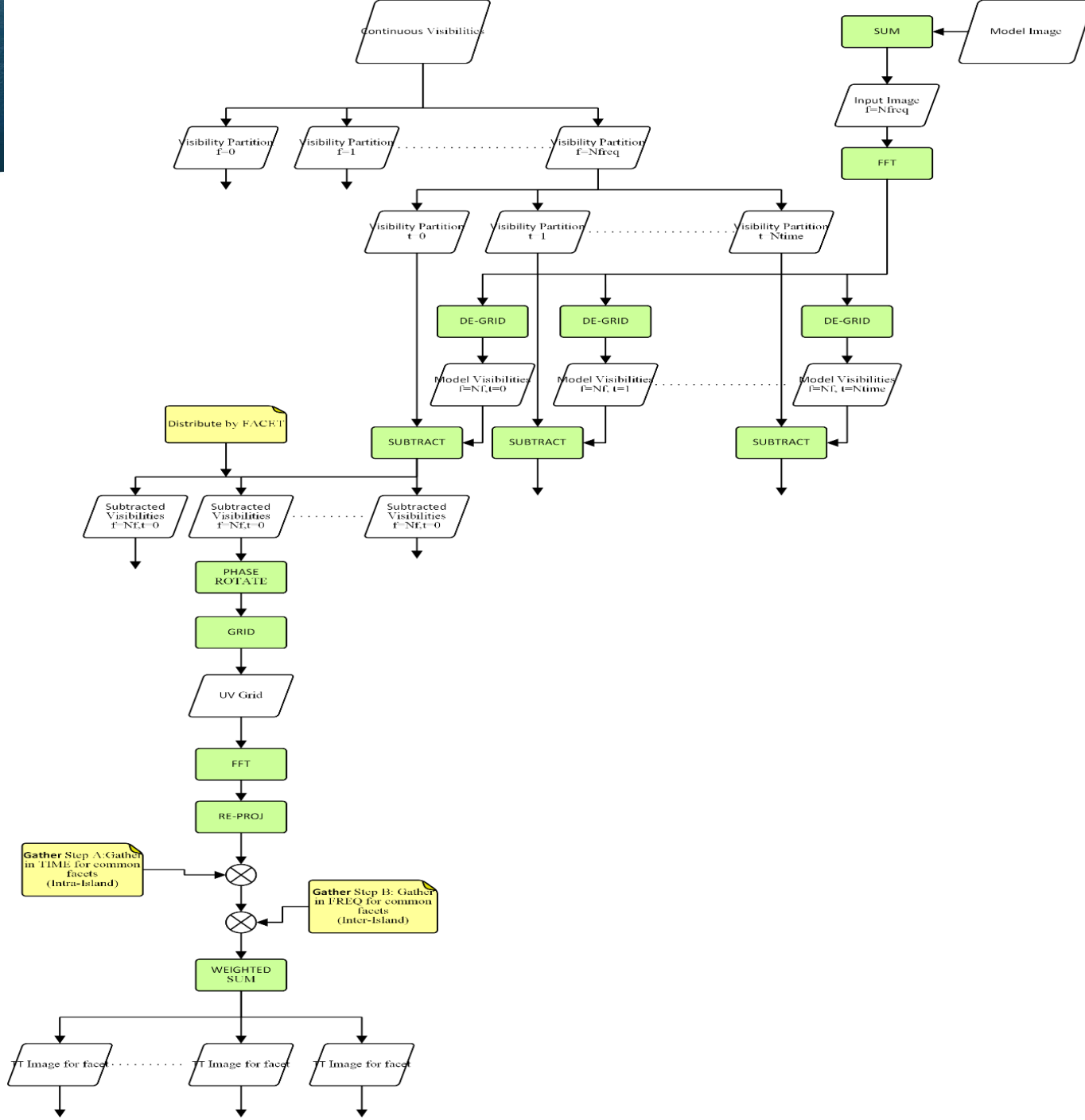


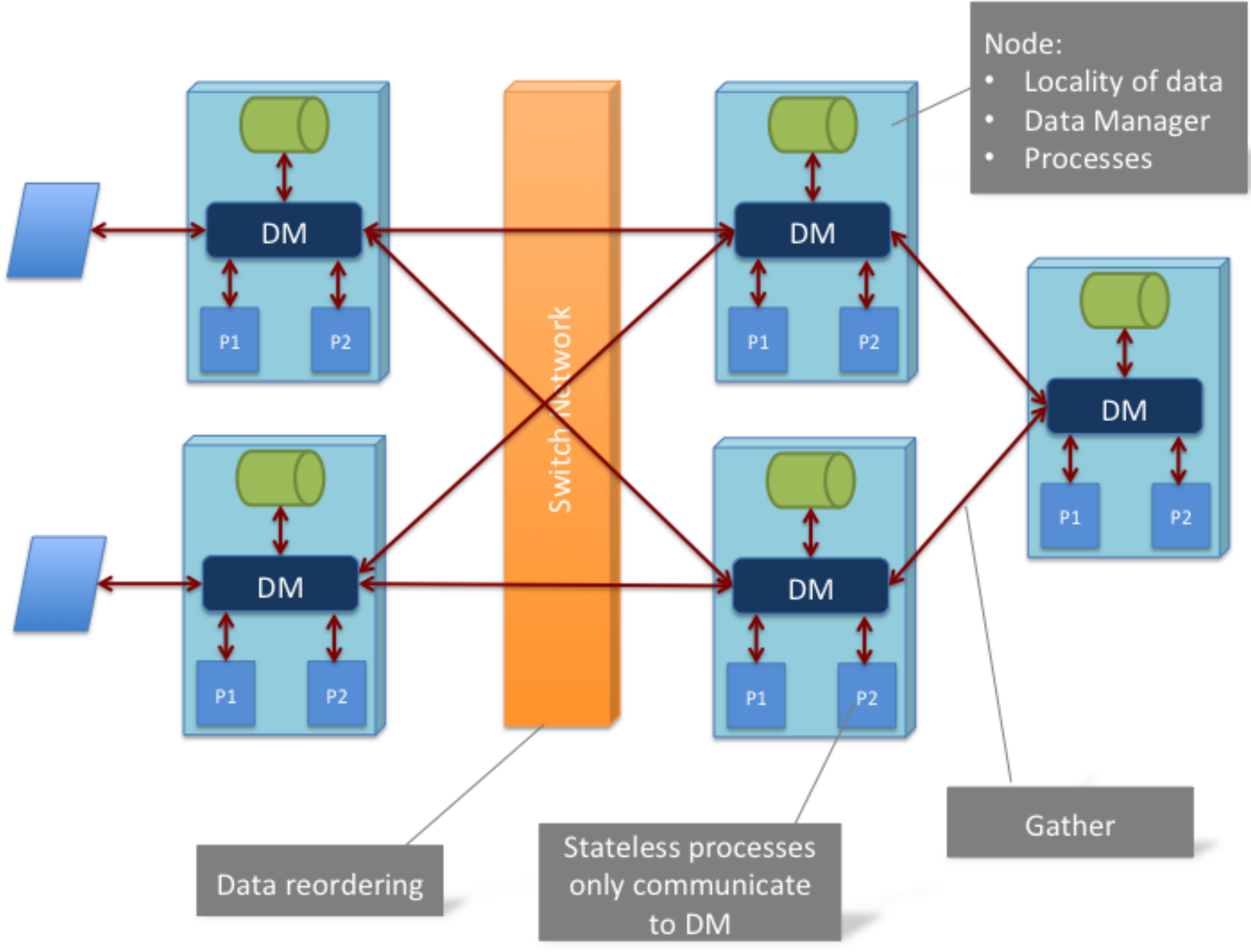
- Smaller FFT size at cost of data duplication

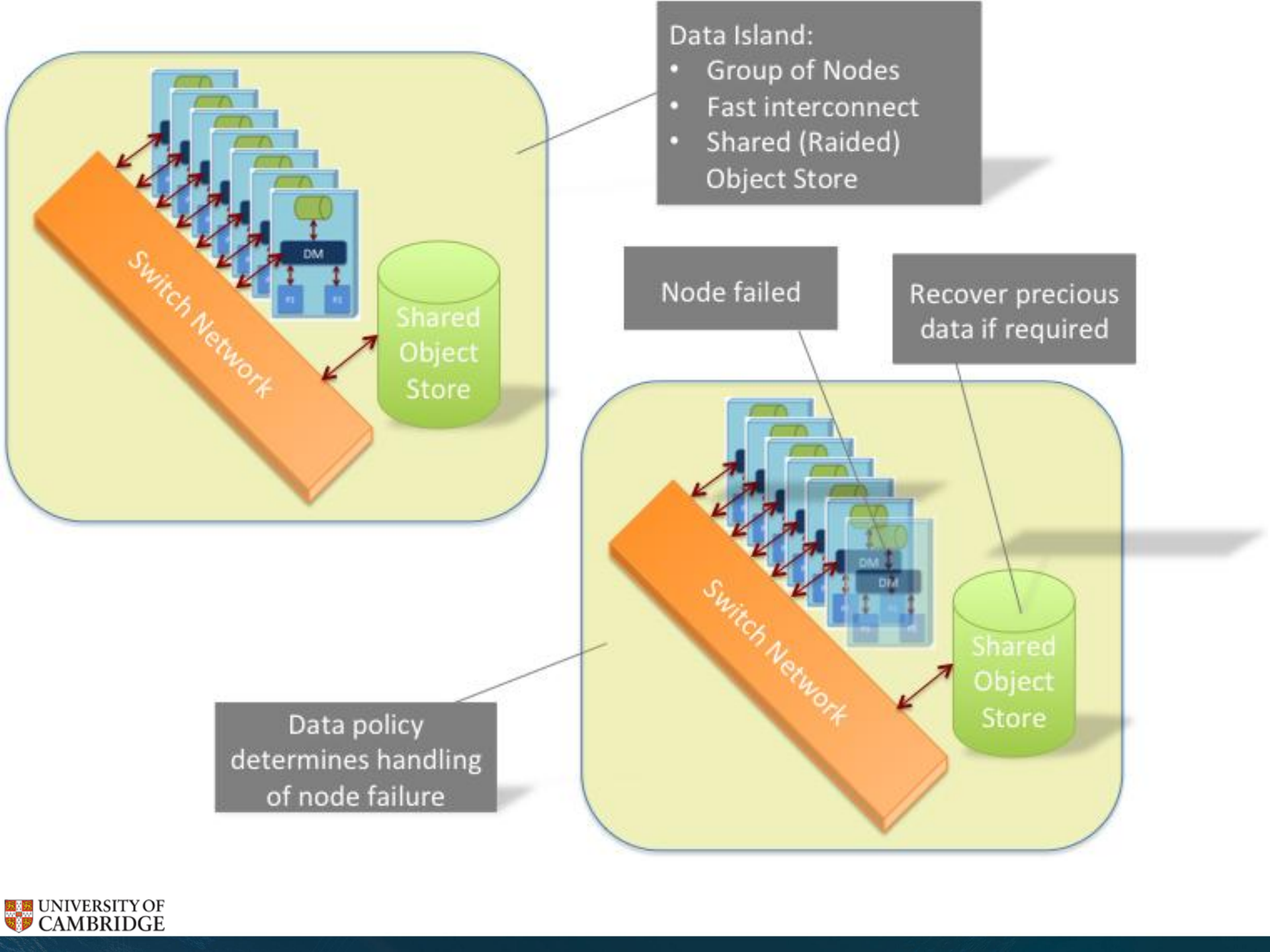
Architectural view of data and information flow



Data Distribution Diagram




Node:
• Locality of data
• Data Manager
• Processes
DM
P1
P2
Switch Network
Data reordering
Stateless processes only communicate to DM
Gather


Data Island:
• Group of Nodes
• Fast interconnect
• Shared (Raided) Object Store
Switch Network
DM
Shared Object Store
Node failed
Recover precious data if required
Switch Network
DM
DM
Shared Object Store
Data policy determines handling of node failure

Regional Centres

- Provide regional data centres and High Performance Computing services
- Software development
  - For the SKA generally
- Data distribution and access for astronomers
- Support for key science teams
- Technology development and engineering support

Archive Estimates SKA1

SCIENCE DATA PROCESSOR

Archive Growth Rate: ~ 50 – 300 Pbytes/yr

Data Products

- Standard products
  - calibrated multi-dimensional sky images;
  - time-series data
  - catalogued data for discrete sources (global sky model) and pulsar candidates
- Requested products for specific experiments / observations
  - calibrated visibility data
  - rotation-measure synthesis images

Further Processing and Science Extraction at Regional Centres

What about SKA2

Data rates and processing increase by FACTOR ~100 for SKA2

3-30 EBytes / year of fully processed data for SKA2

Thank You

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR