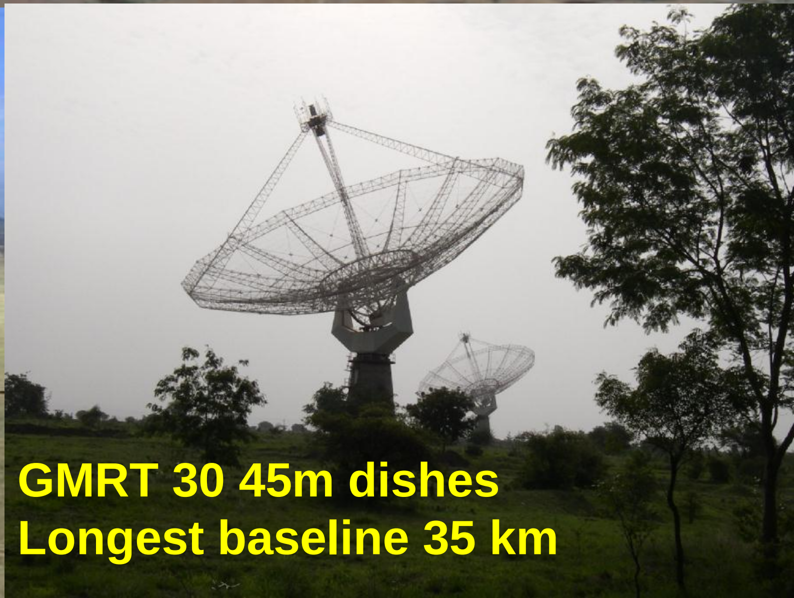
Square Kilometre Array Computational Challenges

Paul Alexander

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)

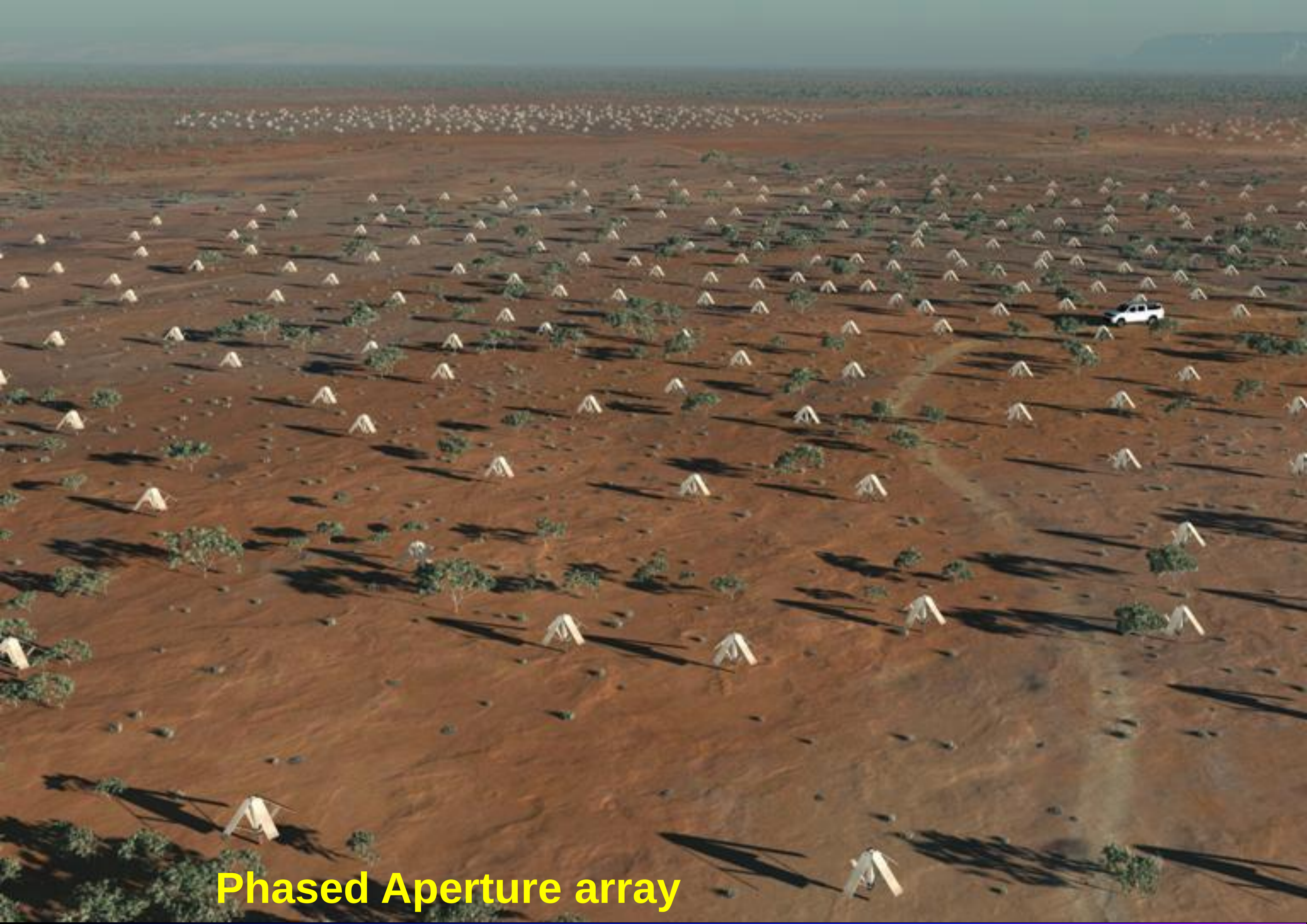
- **Next Generation radio telescope – compared to best current instruments it is ...**
  - **~100 times sensitivity**
  - **~ 10^6 times faster imaging the sky**
  - **More than 5 square km of collecting area on sizes 3000km**

- **Will address some of the key problems of astrophysics and cosmology (and physics)**

- **Builds on techniques developed in Cambridge**
  - **It is an interferometer**

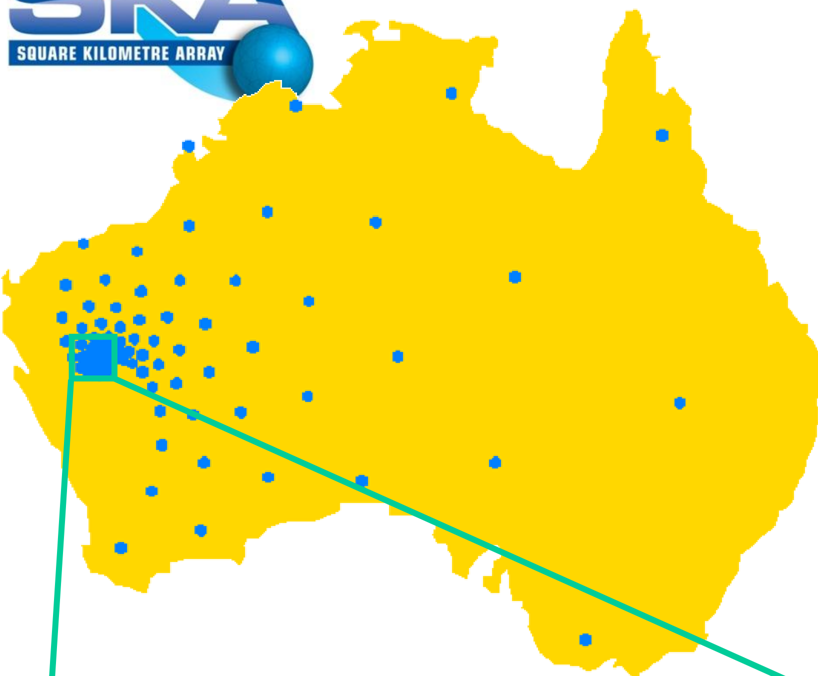
- **Uses innovative technologies...**
  - **Major ICT project**
  - **Need performance at low unit cost**


Dishes

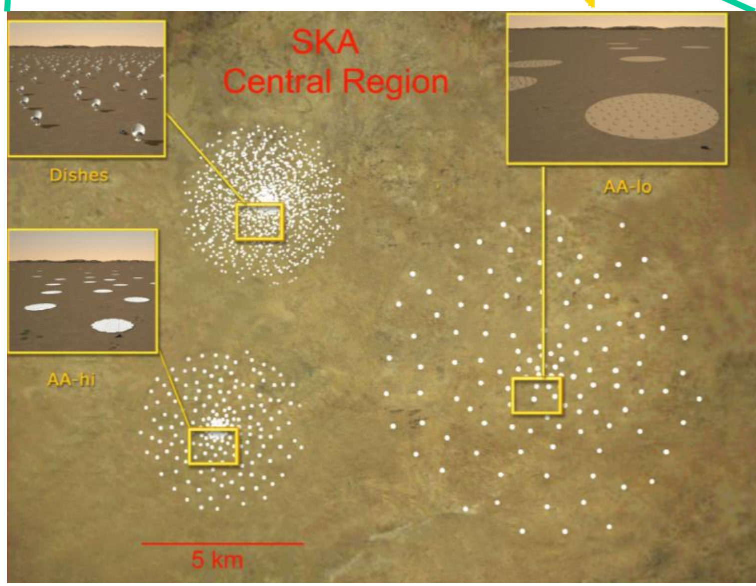


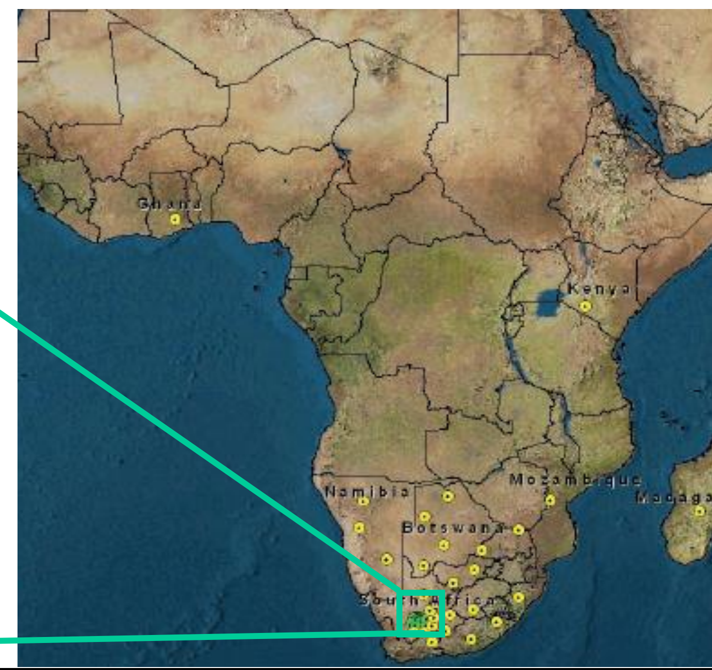
Phased Aperture array

also a Continental sized Radio Telescope

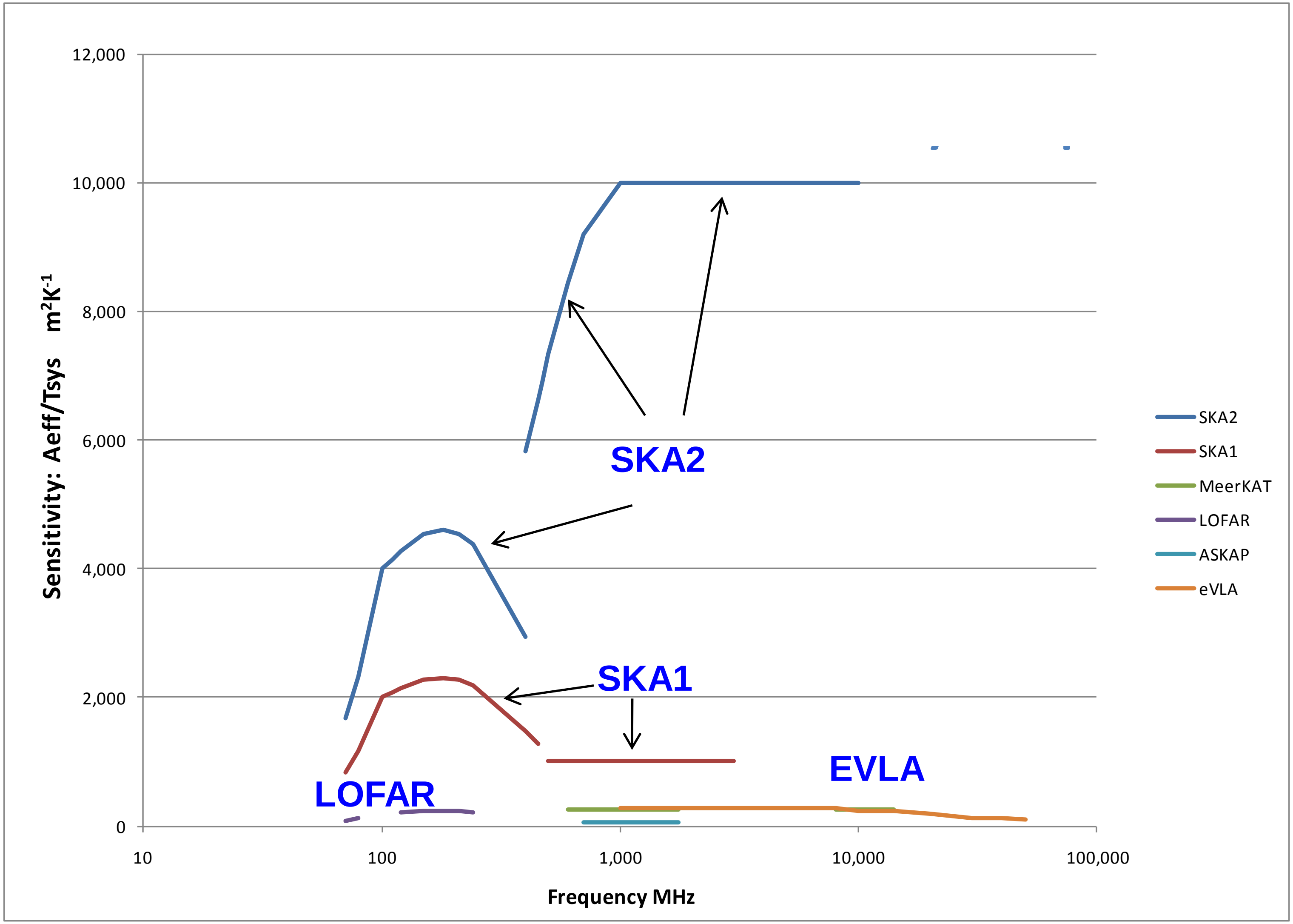


- Need a radio-quiet site
- Very low population density
- Large amount of space
- Possible sites (decision 2012)
  - Western Australia
  - Karoo Desert RSA





Sensitivity comparison

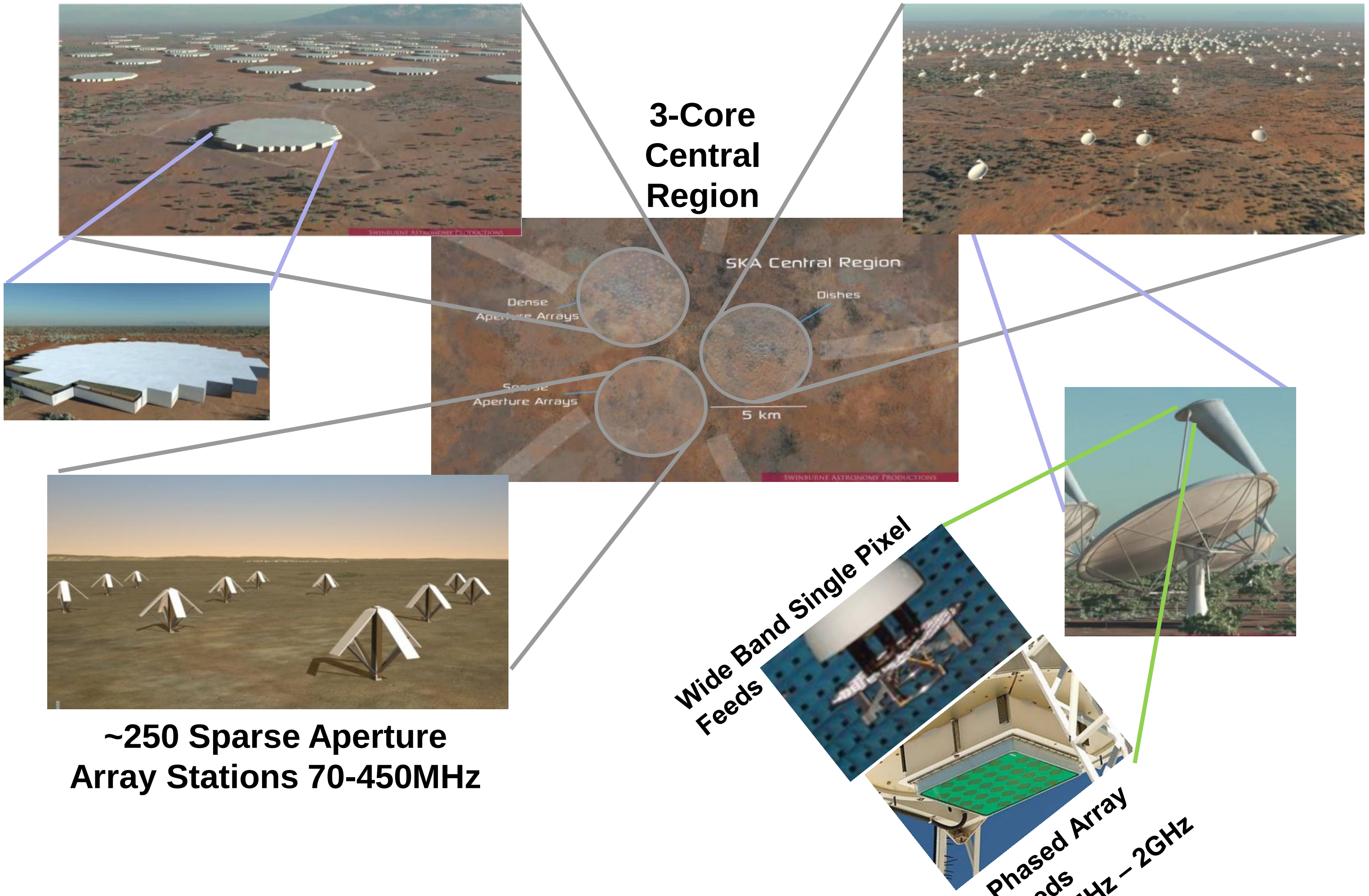


SKA2

~ 250 Dense Aperture Array Stations 300-1400MHz

~ 2700 Dishes

3-Core Central Region



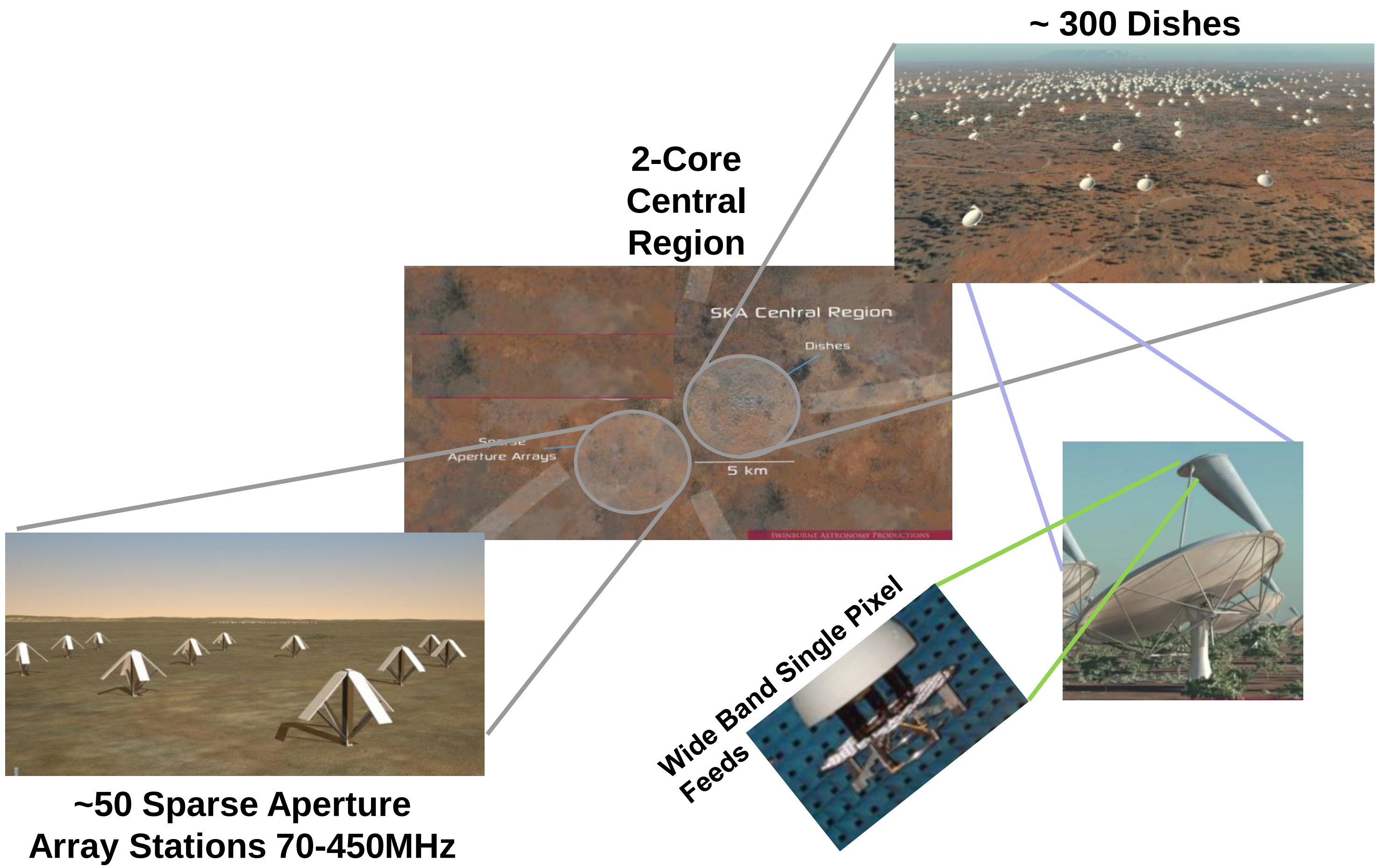
~250 Sparse Aperture Array Stations 70-450MHz

Wide Band Single Pixel Feeds

Phased Array Feeds 800 MHz – 2GHz

Artist renditions from Swinburne Astronomy Productions

SKA1



Artist renditions from Swinburne Astronomy Productions

SKA Timeline

2019	Operations SKA_1 2024: Operations SKA_2	
2019-2023	Construction of Full SKA, SKA_2	€1.5B
2016-2019	10% SKA construction, SKA_1	€300M
2012	Site selection	
2012 - 2016	Pre-Construction: 1 yr Detailed design PEP 3 yr Production Readiness	€90M
2008 - 2012	System design and refinement of specification	
2000 - 2007	Initial concepts stage	
1995 - 2000	Preliminary ideas and R&D	

Work Packages in the PEP

1. Management
2. System
3. Science
4. Maintenance and support /Operations Plan
5. Site preparation
6. Dishes
7. Aperture arrays
8. Signal transport & networks
9. Signal processing
10. Science data processor
11. Telescope manager
12. Power

SPO (items 1–5)

Work Package Contractors (items 6–11)

Work Packages in the PEP

1. Management
2. System
3. Science
4. Maintenance and support /Operations Plan
5. Site preparation

SPO

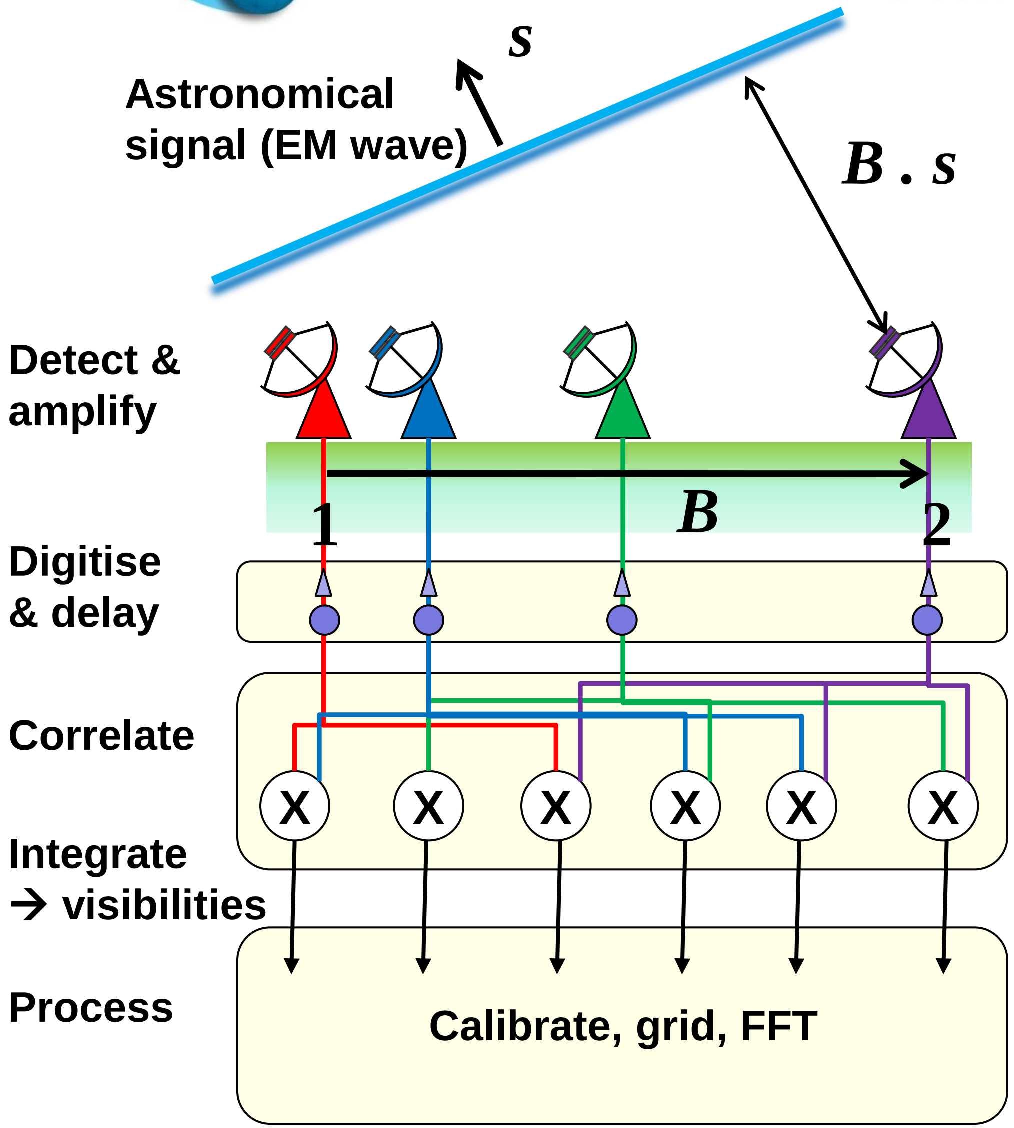
6. Dishes
7. Aperture arrays
8. Signal transport & networks
9. Signal processing
10. Science data processor
11. Telescope manager
12. Power

Work Package Contractors

UK (lead), AU (CSIRO…), NL (ASTRON…)
South Africa SKA, Industry (Intel, IBM…)

Very Brief Introduction to Radio Astronomy Imaging

Standard interferometer



- **Visibility:**

  $V(B) = E_1 E_2^*$

  $= I(s) \exp(i\, \omega\, B.s/c)$

- **Resolution determined by maximum baseline**

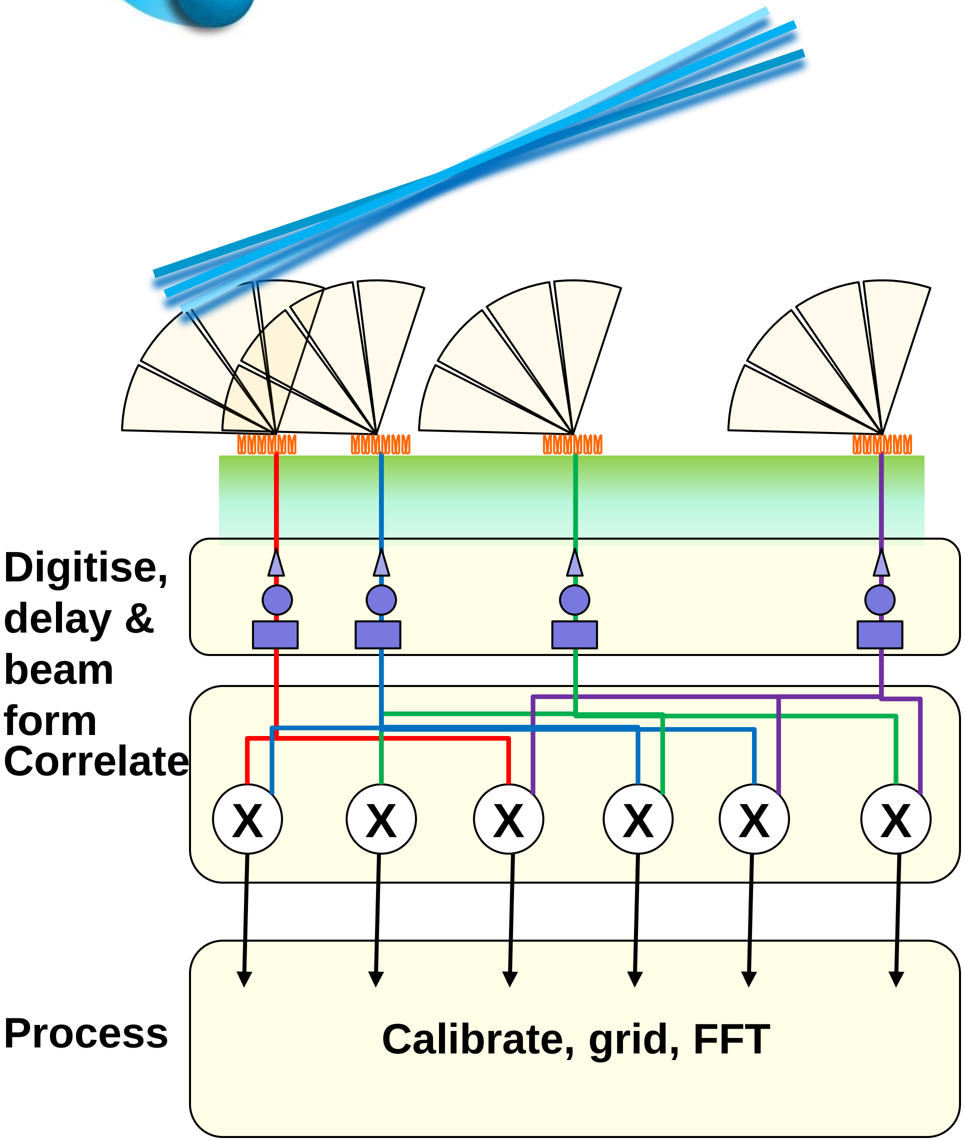
  $\theta_{max} \sim \lambda / B_{max}$

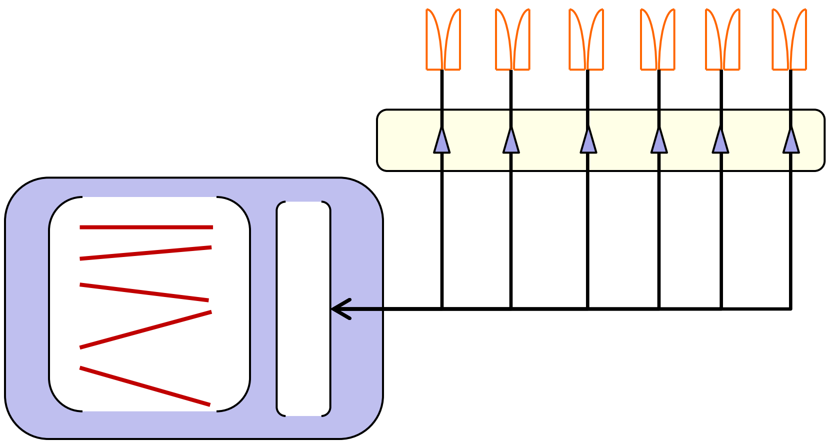
- **Field of View (FoV) determined by the size of each dish**

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

SKY Image

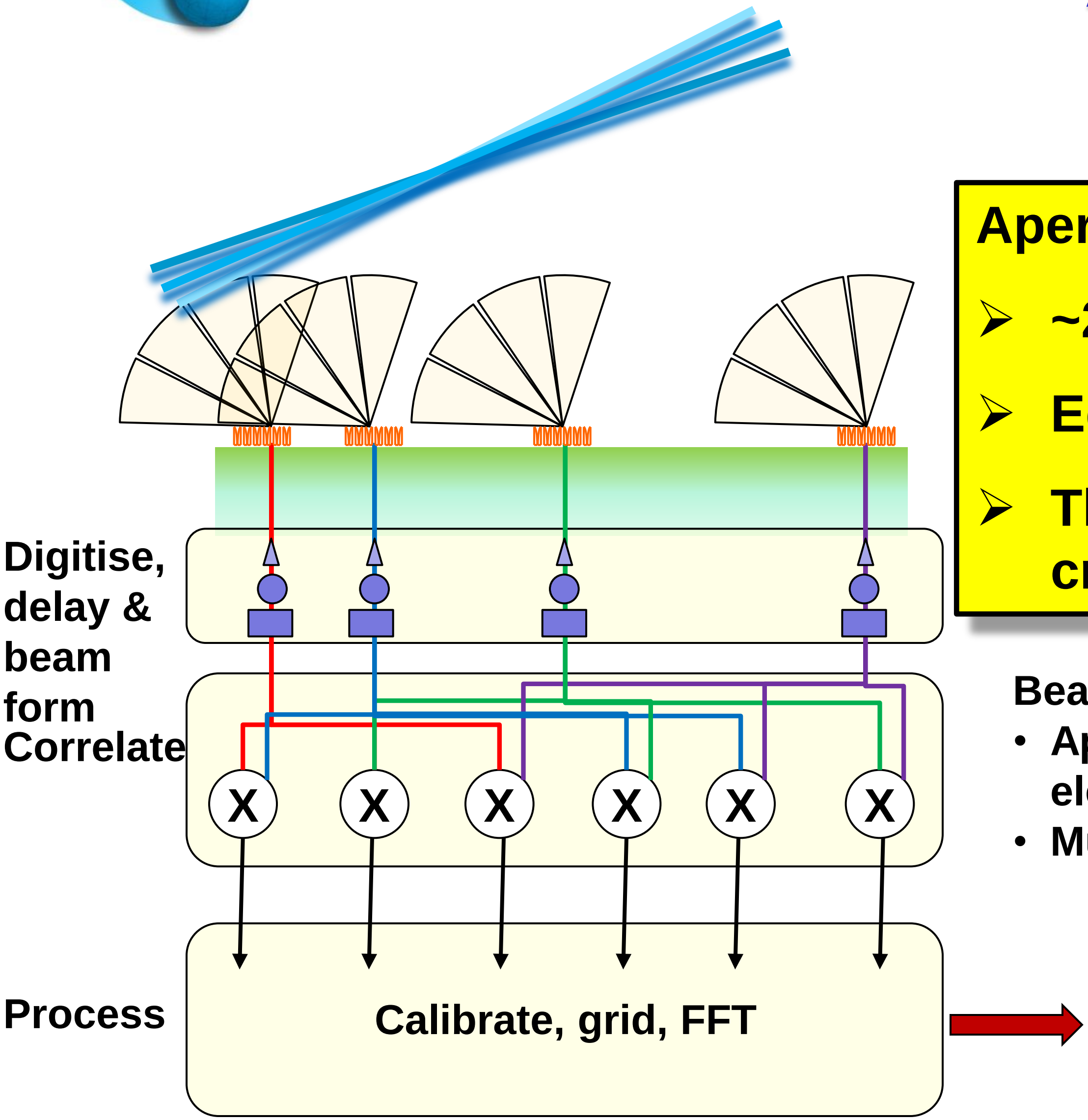
Aperture arrays





Beam form:
- **Apply delay gradients to point electrically**
- **Multiple delay gradients**
  - ➢ **many beams and large FoV**

Aperture arrays



SKY Image

Aperture-Array station

- **~25000 phased elements**
- **Equivalent to one dish**
- **These are then cross-correlated**

Beam form:
- **Apply delay gradients to point electrically**
- **Multiple delay gradients**
  - **many beams and large FoV**

Formulation

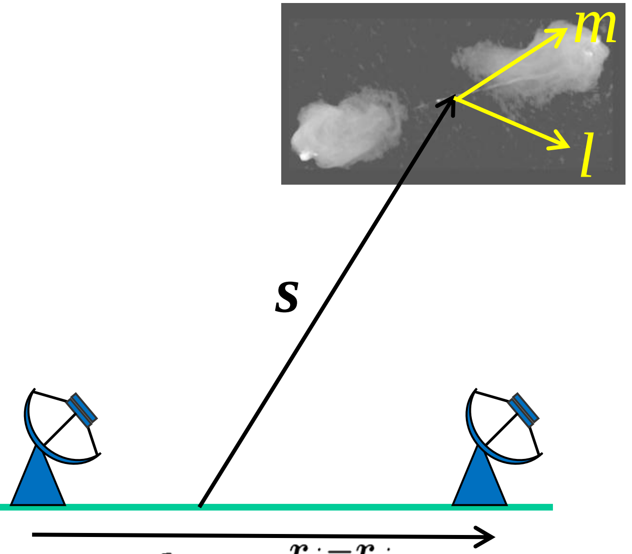
- What we measure from a pair of telescopes:

$$V_{ij} = \int \boldsymbol{J}_i \otimes \boldsymbol{J}_j^* < \boldsymbol{E}_0 \otimes \boldsymbol{E}_0 > \exp(i2\pi(\boldsymbol{s}\cdot\boldsymbol{b}_{ij} + f\tau_i - f\tau_j u))\mathrm{d}^2\boldsymbol{s}$$

- In practice we have to deal with this equation, but for simplicity consider a scalar model

$$V_{ij} = \int A_{ij} I \exp(i2\pi(\boldsymbol{s}\cdot\boldsymbol{b}_{ij} + f\tau_i - f\tau_j))\mathrm{d}^2\boldsymbol{s}$$

- The delays allow us to follow a point on the sky
- The $\boldsymbol{J}_i$ are direction dependent Jones matrices which include the effects of:
  - propagation from the sky through the atmosphere
  - scattering
  - coupling to the antenna/detector
  - gain



$$\boldsymbol{b}_{ij} = \frac{\boldsymbol{r}_i - \boldsymbol{r}_j}{\lambda}$$

Formulation

- Where we include time delays to follow a central point. In terms of direction cosines relative to the point we follow and for a $\boldsymbol{b} = (u,v,w)$

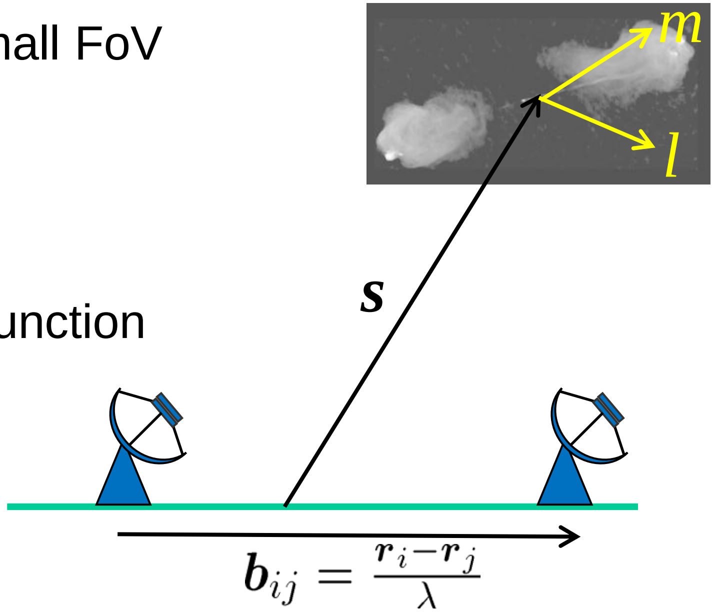
$$V_{ij} = \int A_{ij}(l,m)I(l,m)\exp(i2\pi(ul + vm + w(\sqrt{1 - l^2 - m^2} - 1)))\frac{\mathrm{d}l\mathrm{d}m}{\sqrt{1-l^2-m^2}}$$

- If we can calibrate our system we can apply our telescope-dependent calibrations and then for small FoV we can approximate

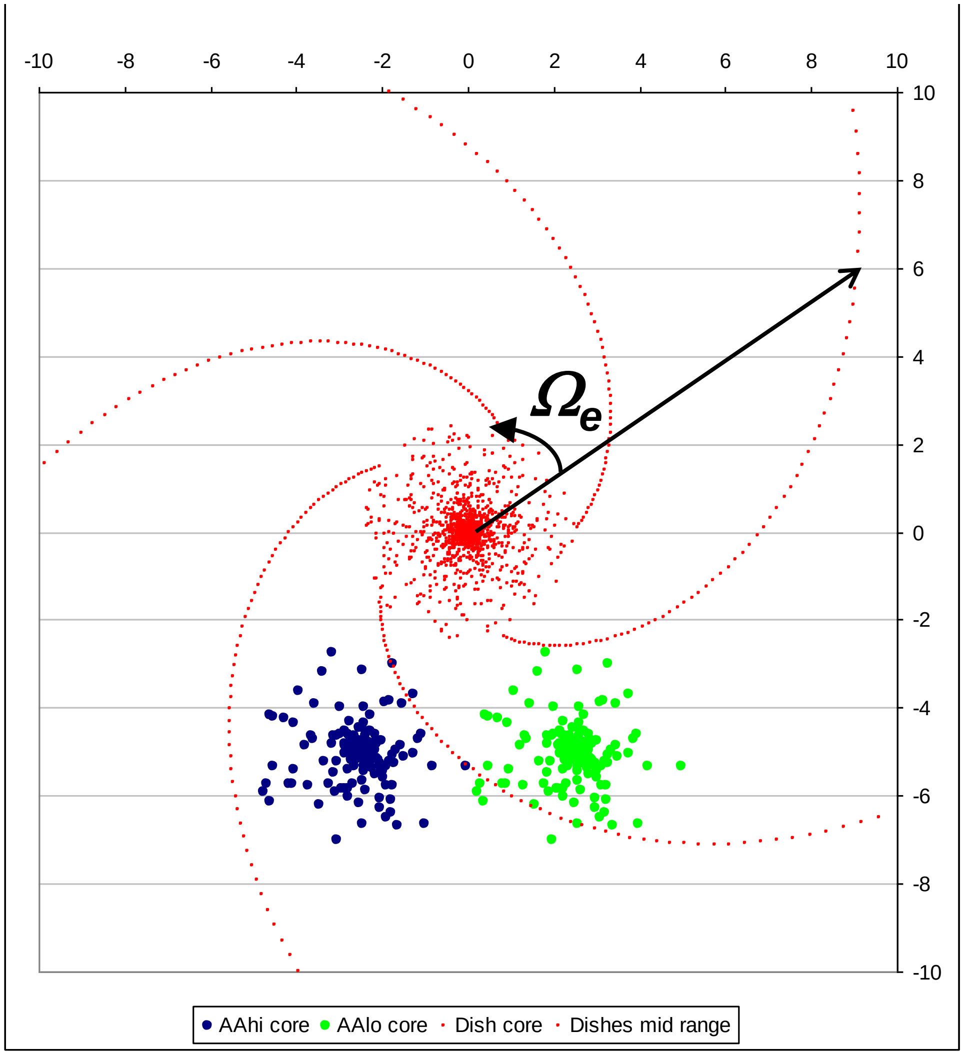
$$V(u,v) \approx \int A(l,m)I(l,m)\exp(i2\pi(ul + vm))$$

- And the measured data are just samples of this function

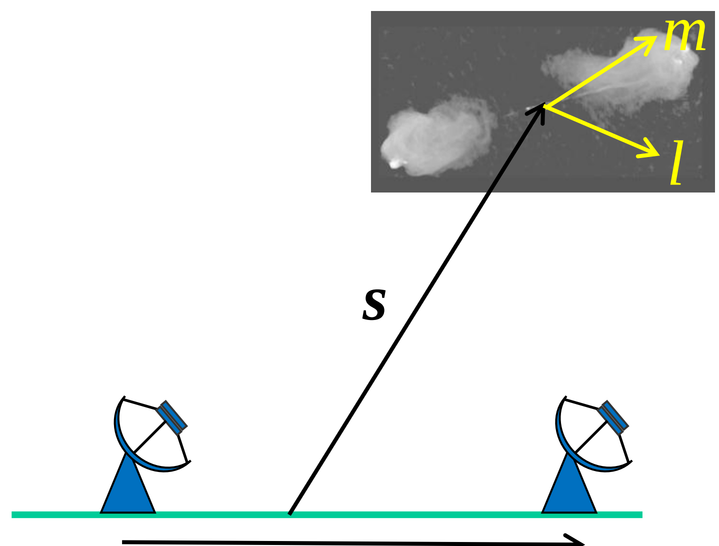
- In this case we can estimate the sky via Fourier inversion and deconvolution



Formulation



- Sampling of the Fourier plane is determined by the positioning of the antennas and improved by the rotation of the earth



$$\mathbf{b}_{ij} = \frac{\mathbf{r}_i - \mathbf{r}_j}{\lambda}$$

The Science Aims and the Imaging Challenges

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

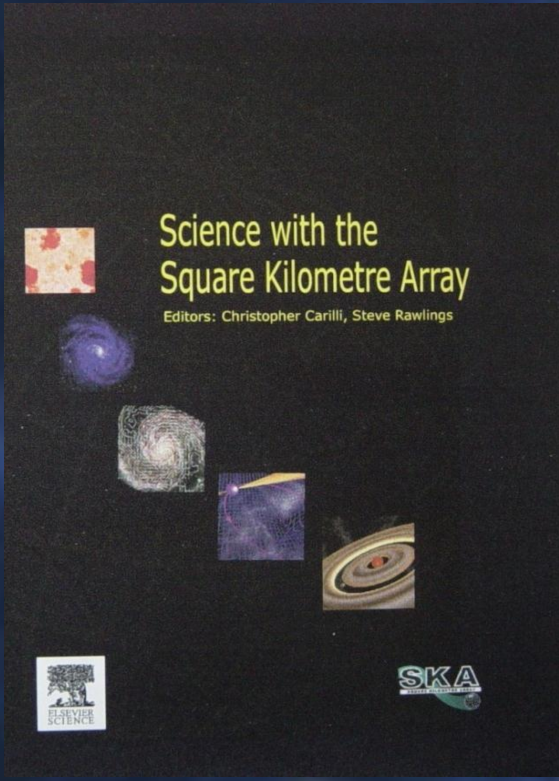
- **Cradle of Life**

FUNDAMENTAL FORCES

- **Pulsars, General Relativity & gravitational waves**

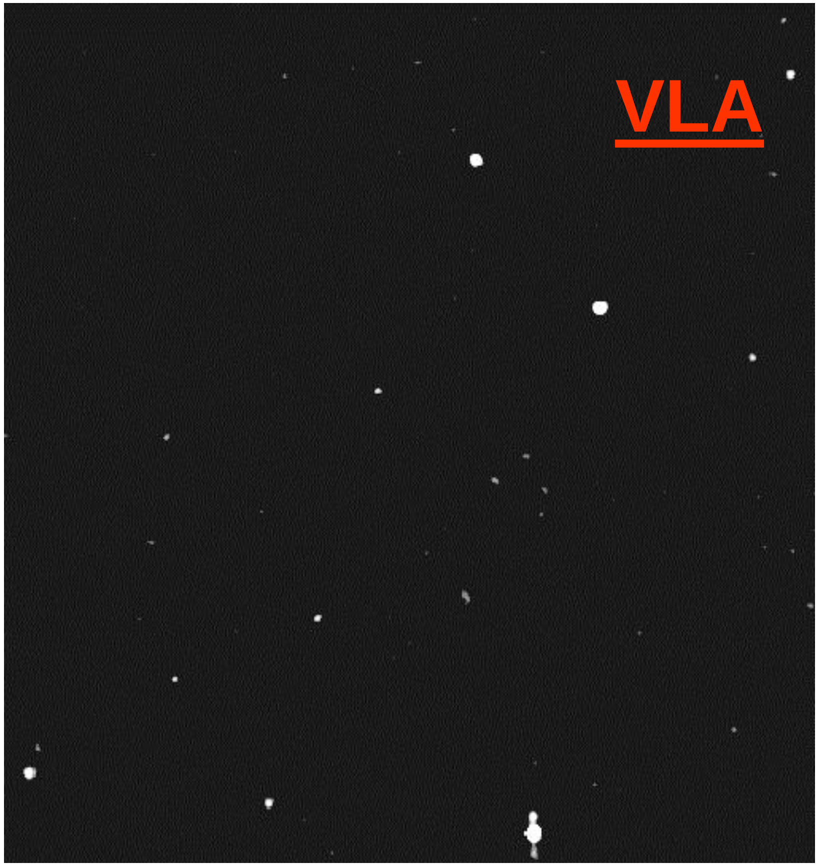
- **Origin & evolution of cosmic magnetism**

TRANSIENTS (NEW PHENOMENA)



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

Galaxy Evolution back to z~10?

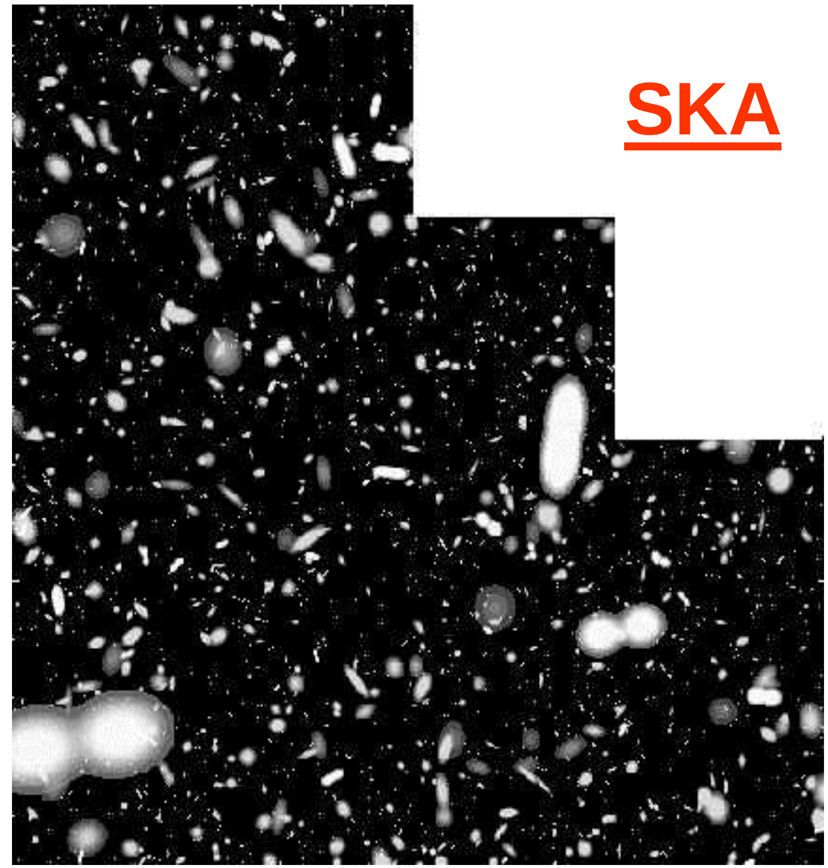


~ 3000 galaxies

~15 radio sources

Galaxy Evolution back to z~10?

HDF



SKA

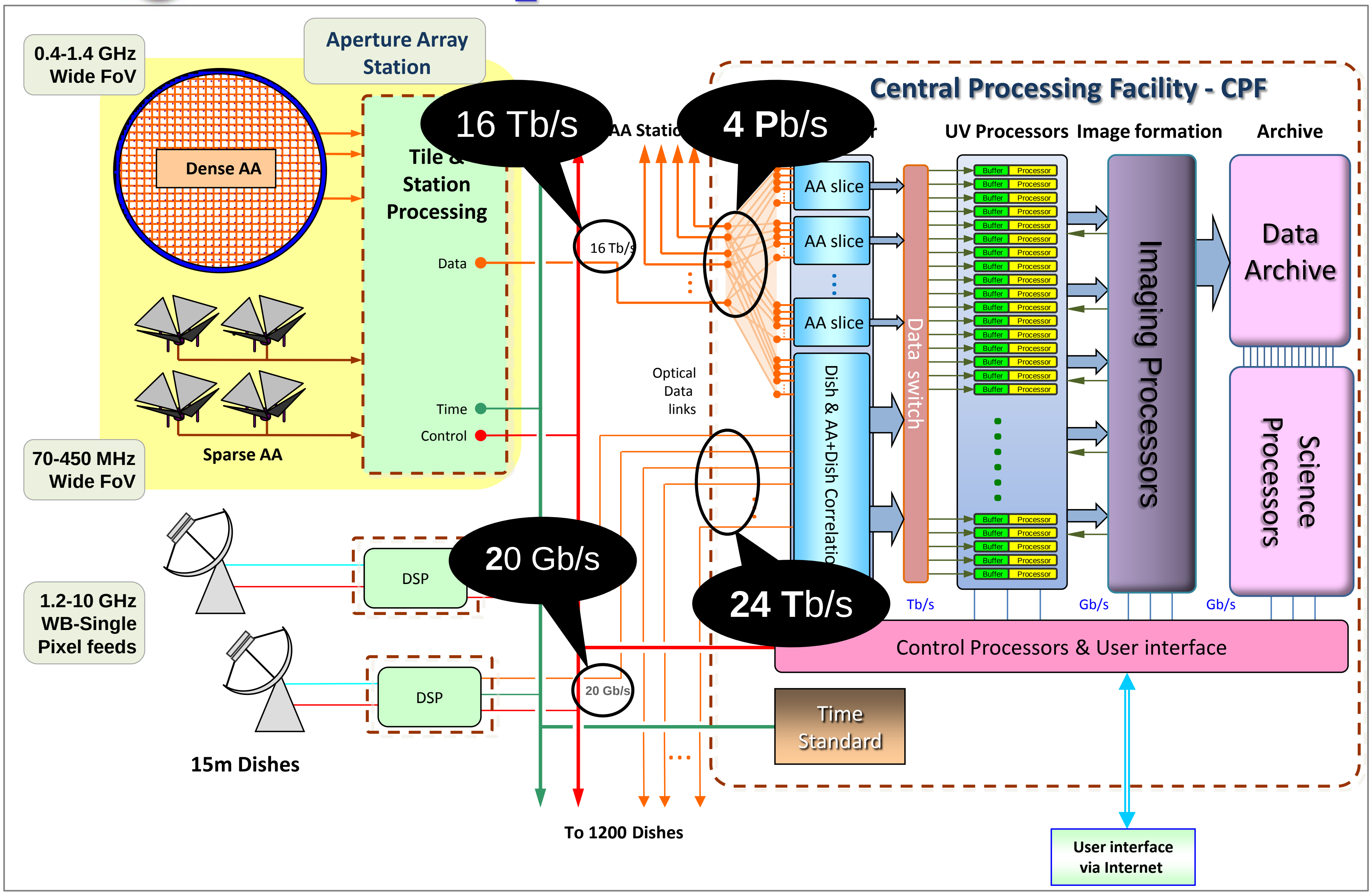
The Imaging Challenge

This illustrates one of our main challenges

- To make effective use of the improved sensitivity we face an immediate problem
- Typically within the field of view of the telescope the noise level will be ~10^6 – 10^7 times less than the peak brightness
- We have to achieve sufficiently good calibration and image fidelity to routinely achieve a "***dyanamic range***" of > 10^7:1
- With very hard work now we can just get to 10^6:1 in some fields

The Processing Challenge

SKA$_2$ wide area data flow



SKA1 Data rates from receptors

- Dishes
  - **Depends on feeds, but illustrate by 2 GHz bandwidth at 8-bits**
  - **G = 64 Gb/s from each dish**
- For Phased Array feeds increased by number of beams (~20)
  - **G ~ 1 Tb/s**
- For Low frequency Aperture Arrays :
  - Bandwidth is 380 MHz
  - Driven by the requirements of Field of View from the science requirements which from DRM is 5 sq-degrees → 20 beams
  - G = 240 Gb/s
- These are from each collector into the correlator or beam former
  - **300 dishes**
  - **285 75-m AA stations**
  - **G(total) ~ 68 Tb/s**

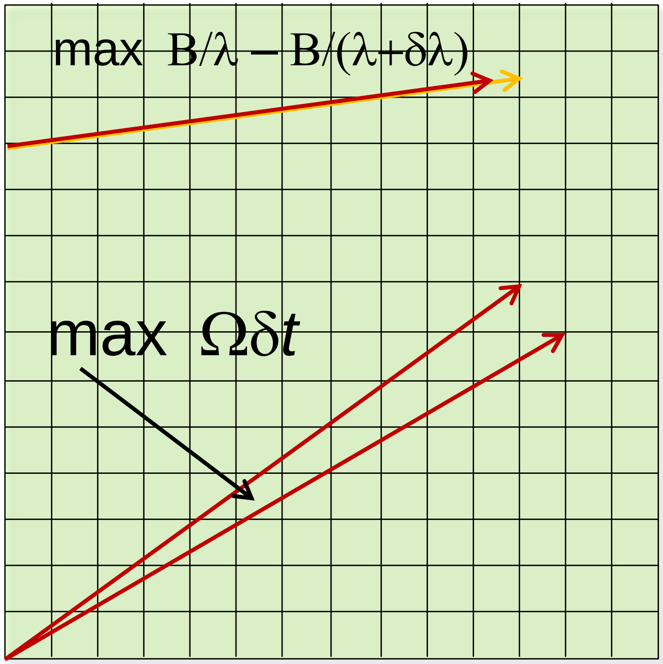
Data Rates

- After correlation the data rate is fixed by straightforward considerations
  - Must sample fast enough (limit on integration time) δt
  - Baseline $\propto$ B/λ
  - UV (Fourier) cell size $\propto$ D/λ

$$\Omega \delta t \frac{B}{\lambda} < \frac{1}{X}\frac{D}{\lambda}$$

  - Must have small-enough channel width to avoid chromatic aberration

$$\delta\left(\frac{B}{\lambda}\right) < \frac{1}{X}\frac{D}{\lambda}$$



Data rates from the correlator

- Standard results for integration/dump time and channel width

$$\frac{\delta t}{s} = a_t \frac{D}{B} \sim 1200 \frac{D}{B} \qquad \frac{\delta f}{f} = a_f \frac{D}{B} \sim \frac{1}{10}\frac{D}{B}$$

- Data rate then given by

$$G = g(B)\frac{1}{2}N^2 N_p^2 N_b \frac{1}{\delta t}\frac{\Delta f}{\delta f} 2N_w \qquad G = g(B) N^2 N_w N_p^2 N_b \frac{1}{a_t a_f}\frac{\Delta f}{f}\left(\frac{B}{D}\right)^2$$

antennas   # polarizations   # beams   word-length

- Can reduce this using baseline-dependent integration times and channel widths

Example correlator data rates and products SKA1

- Aperture Array Line experiment (e.g. EoR)
  - 5 sq degrees; 170000 channels over 250 MHz bandwidth
    - ~ 30 GB/s reducing quickly to ~ 1GB/s
    - Up to 500 TB UV (Fourier) data; Images (3D) ~ 1.5 TB
- Continuum experiment with long baselines with the AA
  - 100 km baseline with the low-frequency AA
    - 1.2 TB/s reducing to ~ 12.5 GB/s
    - Up to 250 TB/day to archive if we archive raw UV data
- Spectral-line imaging with dishes
    - Data rates ~ 50 GB/s; Images (3D) ~ 27 TB

Example beam-formed data rates SKA1

- **Pulsar search**
  - **Galactic-plane survey for pulsars**
    - ➢ **~ 400 GB/s to de-disperser (hardware?)**
    - ➢ **Data products are of small volume as all analysis is done in pseudo real-time.**

Example Data rates SKA2

Experiment				3000 Dishes + SPF		1630 Dishes + PAFS		250 AA stations	
Description	B_{max} (km)	Δf (MHz)	f_{max} (MHz)	Achieved FoV[1]	Data rate (Tb/s)	Achieved FoV[1]	Data rate (Tb/s)	Achieved FoV[1]	Data rate (Tb/s)
Survey: High surface brightness continuum	5	700	1400	0.78	0.055	15	0.11	108	0.03
Survey: Nearby HI high res. 32000 channels	5	700	1400	0.78	1.0	15	2.0	108	2.6
Survey: Medium spectral resolution; resolved imaging (8000)	30	700	1400	0.78	1.2	15	2.4	108	5.4
Survey: Medium resolution continuum	180	700	1400	0.78	33.1	15	66	108	14.1
Pointed: Medium resolution continuum deep observation	180	700	1400	0.78	33.1			0.78	0.15
High resolution with station beam forming[2]	1000	2000	8000	0.0015	33.4				
High resolution with station beam forming[3]	1000	2000	8000	0.0015	429				
Highest resolution for deep imaging[2]	3000	4000	10000	0.001	391				

Notes

1. Achieved FoV is at f_{max} and has units of degrees squared. For the AA and PAFs we calculate the data rate assuming it is constant across the band.
2. Assuming that for the dynamic range the FoV of the station only has to be imaged
3. Assuming that for the dynamic range the FoV of the dish must be imaged

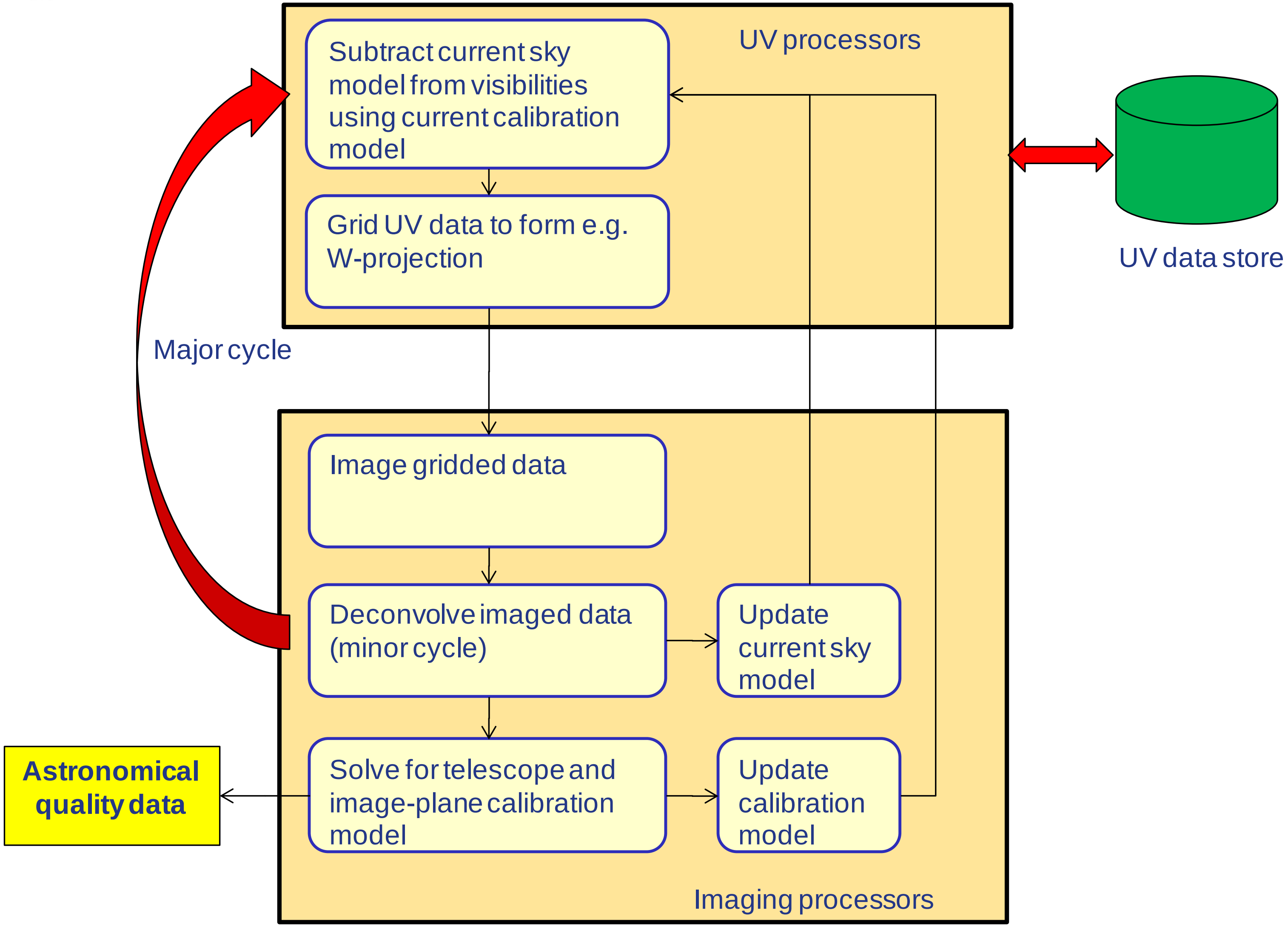
Data Products

Experiment	T_{obs}	B/km	D/m	N_b	N_{ch}	N_v	Size / TB
High resolution spectral line	3600	200	15	1	32000	$5\ 10^{13}$	200
Survey spectral line medium resolution	3600	30	56	1000	32000	$8\ 10^{13}$	330
Snapshot continuum – some spectral information	60	180	56	1200	32	$7\ 10^{12}$	30
High resolution long baseline	3600	3000	60	1	4	$7\ 10^{14}$	360

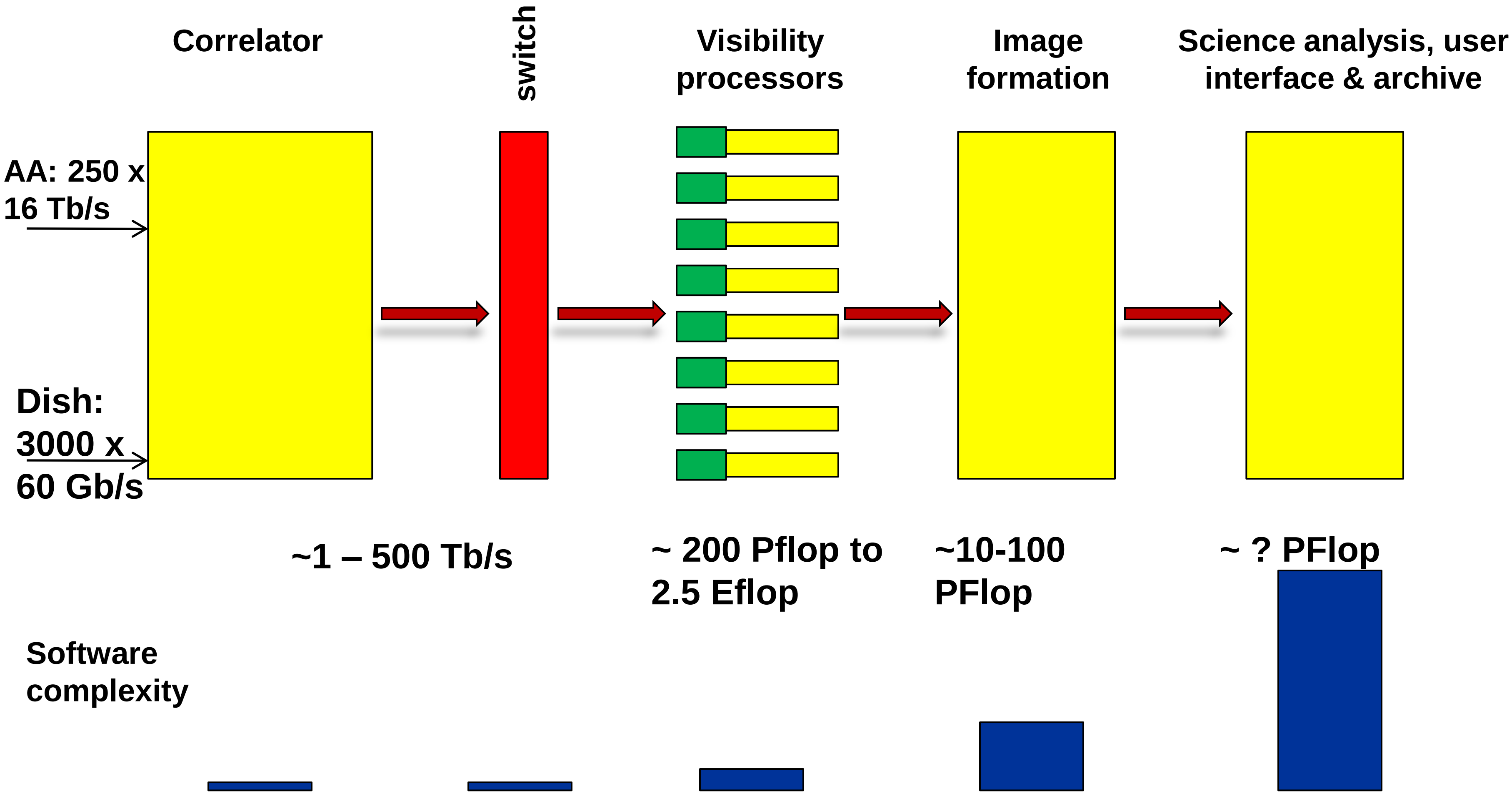
- **~0.5 – 10 PB/day of image data**
- **Source count ~10^6 sources per square degree**
- **~10^{10} sources in the accessible SKA sky, 10^4 numbers/record**
- **~1 PB for the catalogued data**

100 Pbytes – 3 EBytes / year of fully processed data

Processing model



The SKA Processing Challenge



Conclusions

The next generation radio telescopes offer the possibility of transformational science, but at a cost

A major processing challenge

- Need to analyse very large amounts of streaming data
    - Current algorithms iterative – need to buffer data
- Problem too large to, for example, use a direct Bayesian approach
- Are our (approximate) algorithms good enough to take into account all error effects that need to be modelled?
- Only recently have we had to consider most of the effects – what have we forgotten?
- Phased approach to SKA is very good for the processing – performance increasing and critically we can continually learn