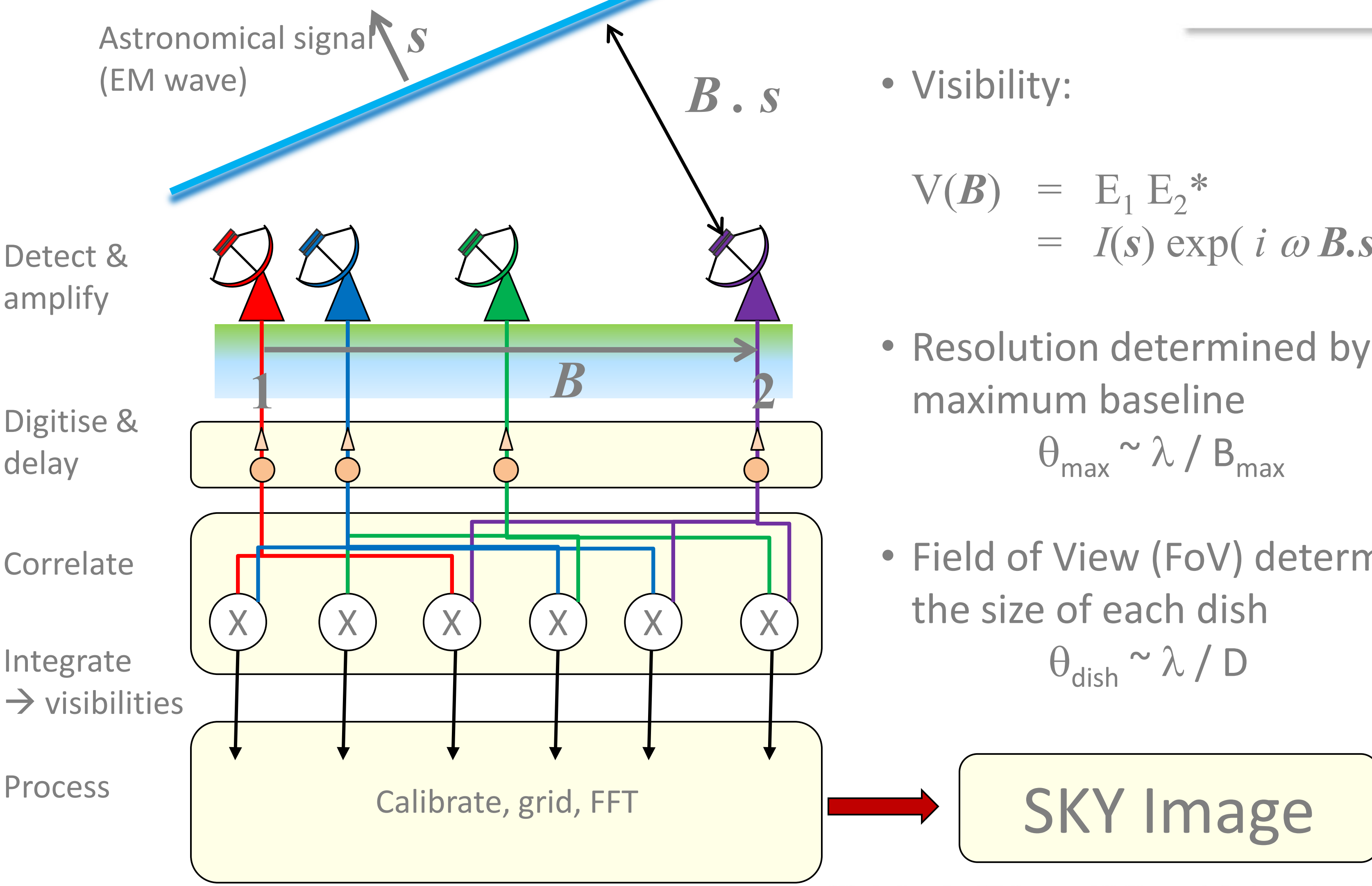
Flow of Data from Receptors to Archive

Standard interferometer



- Visibility:

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

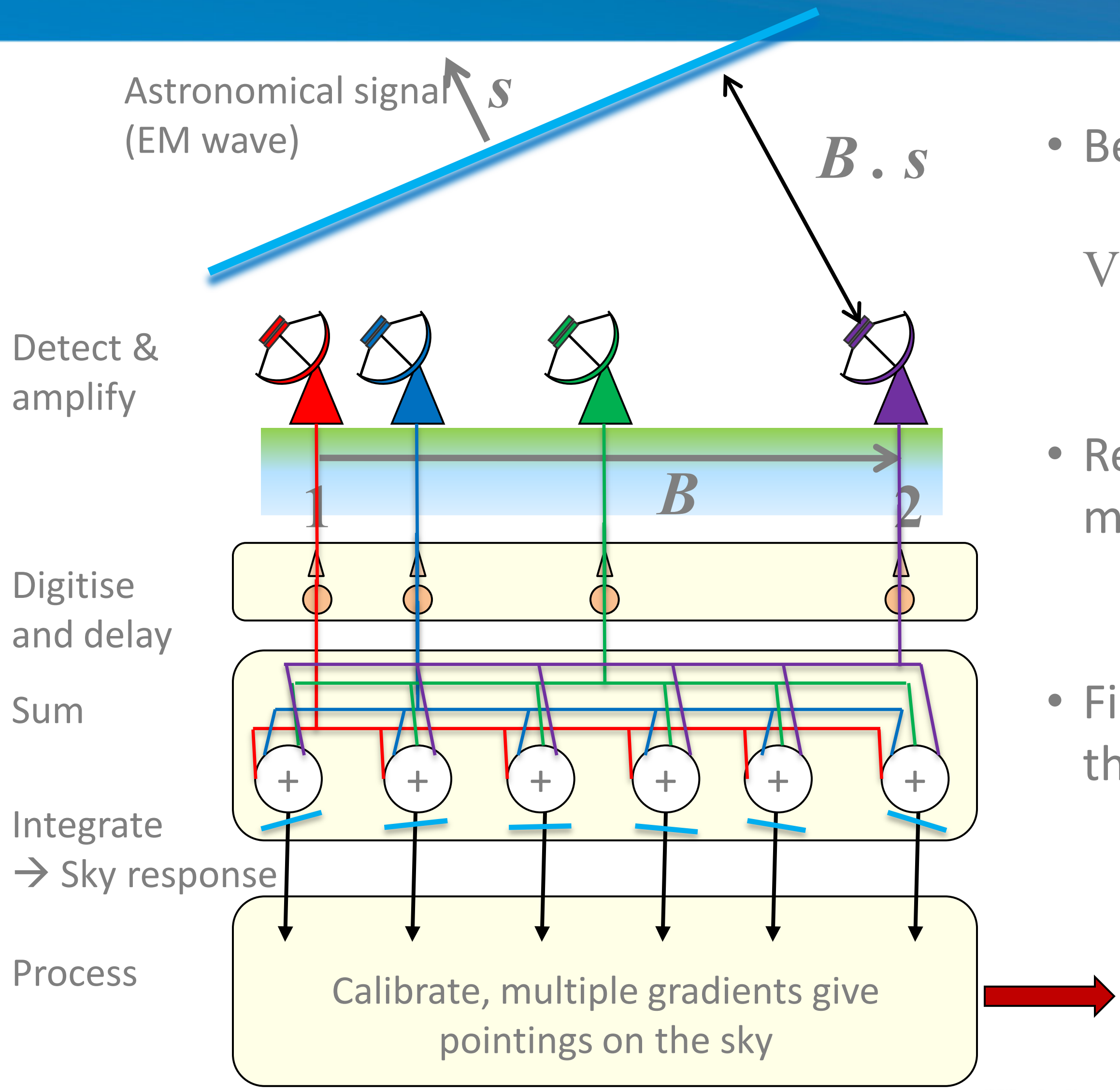
- Resolution determined by maximum baseline

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

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

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

Beam forming



- Beamform:

  $$V(\boldsymbol{B}) = E_1 + E_2 \exp(\, i\, \omega\, \boldsymbol{B_2}.\boldsymbol{s}/c\,) + \ldots$$

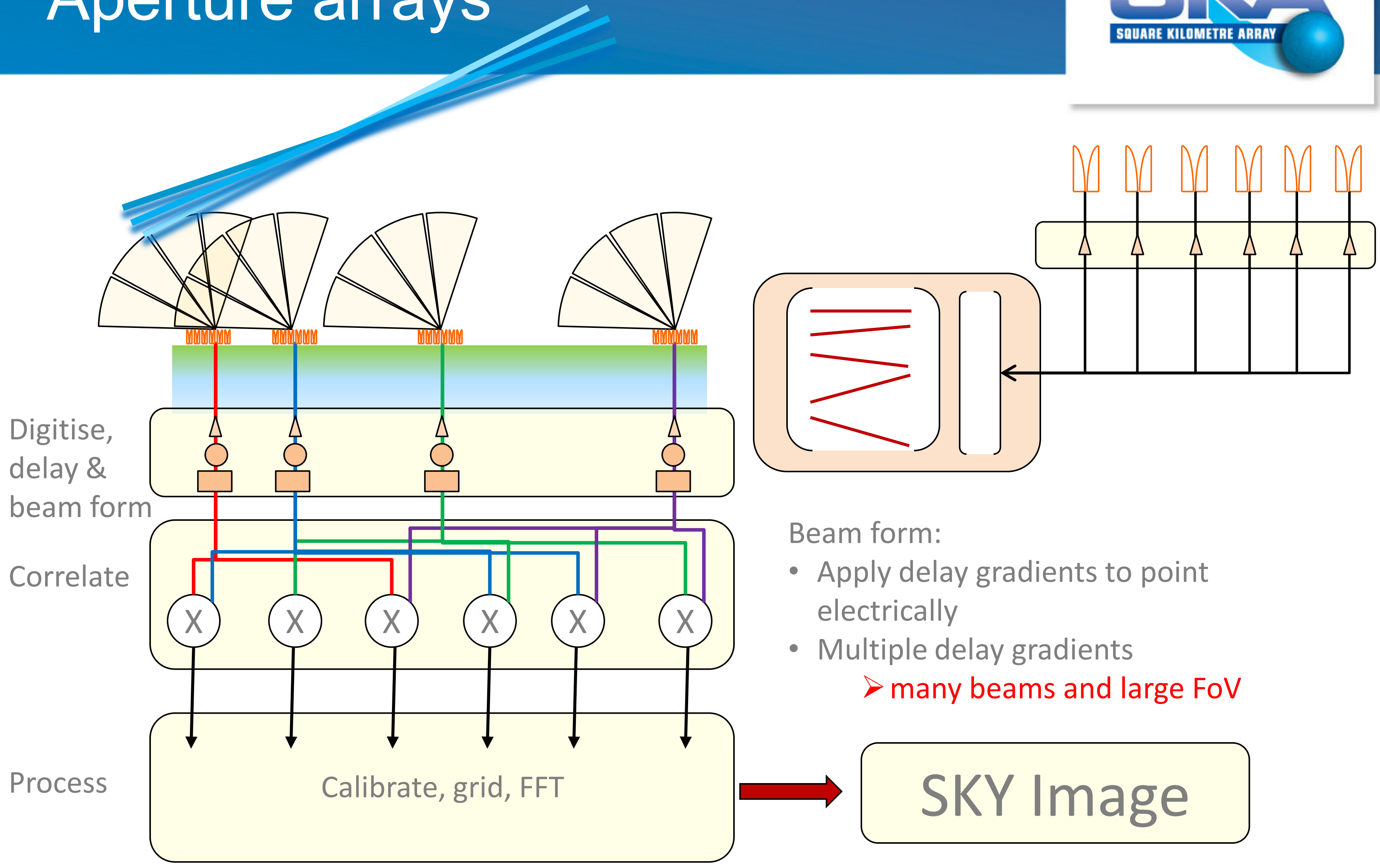
- Resolution determined by maximum baseline

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

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

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

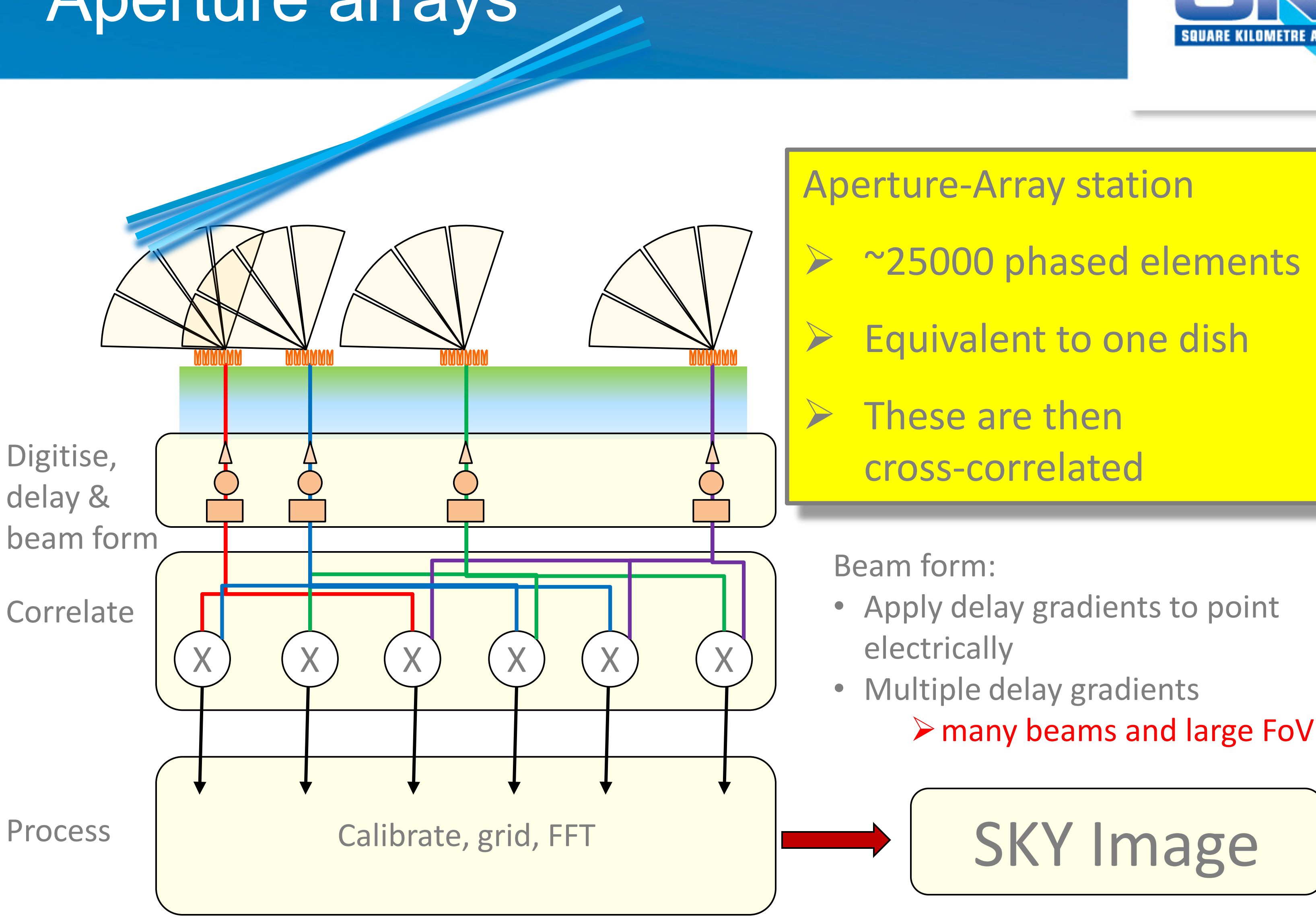
Aperture arrays



Beam form:

- Apply delay gradients to point electrically
- Multiple delay gradients
  - ➢ many beams and large FoV

Aperture arrays



Aperture-Array station

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

Beam form:

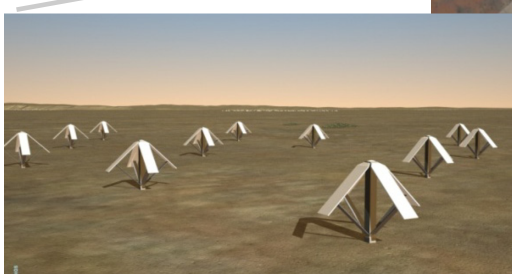
- Apply delay gradients to point electrically
- Multiple delay gradients
  - many beams and large FoV

SKA2

~ 250 Dense Aperture Array
Stations 300-1400MHz

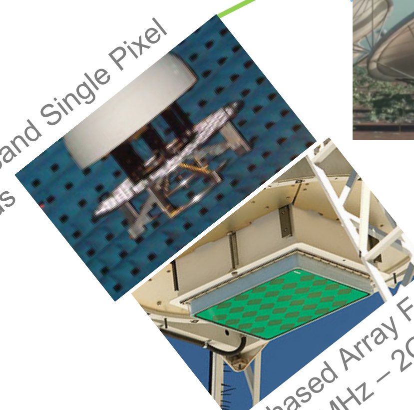
~ 2700 Dishes

3-Core
Central
Region



~250 Sparse Aperture Array
Stations 70-450MHz

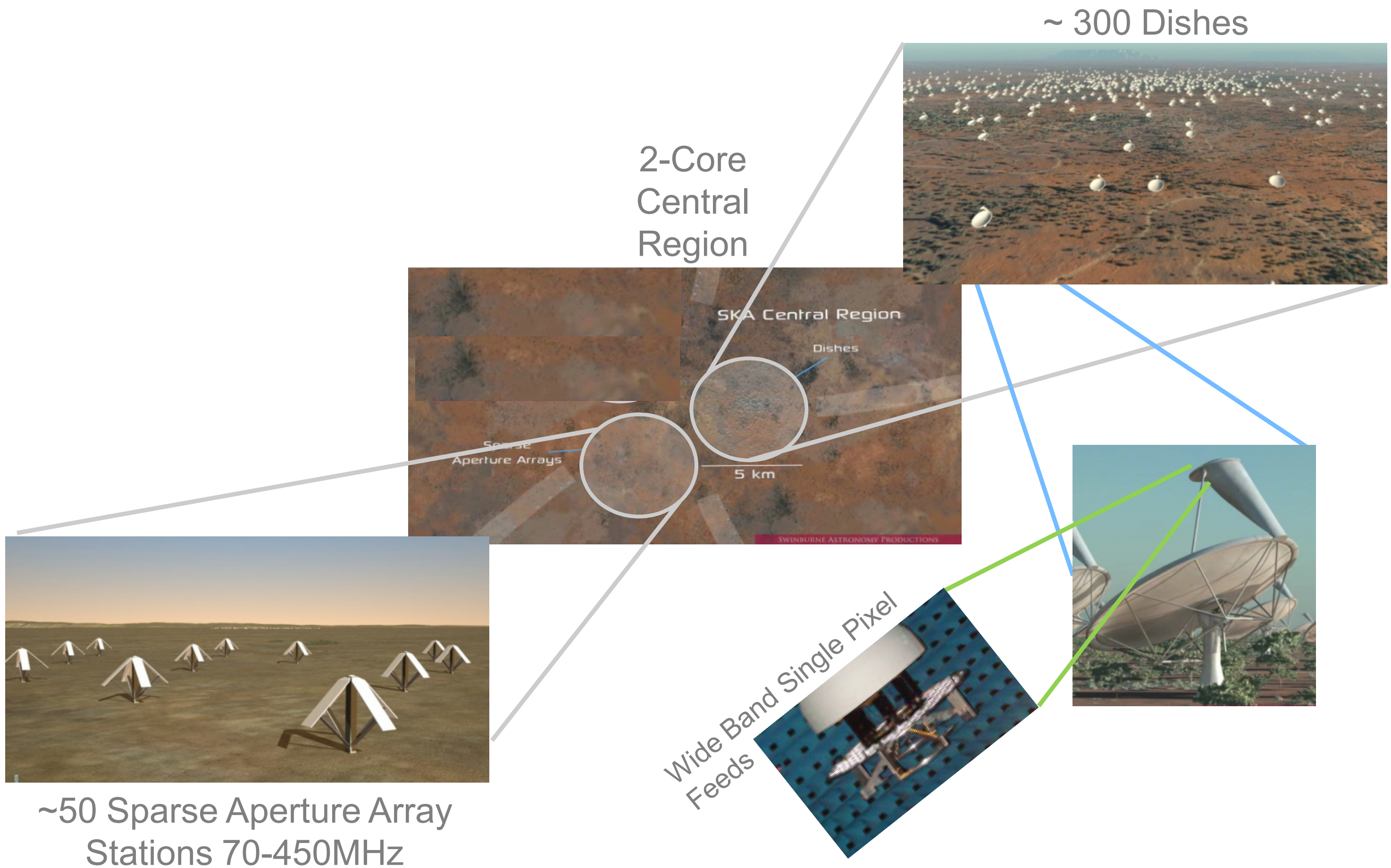
Wide Band Single Pixel
Feeds



Phased Array Feeds
800 MHz – 2GHz

Artist renditions from Swinburne Astronomy Productions

SKA1



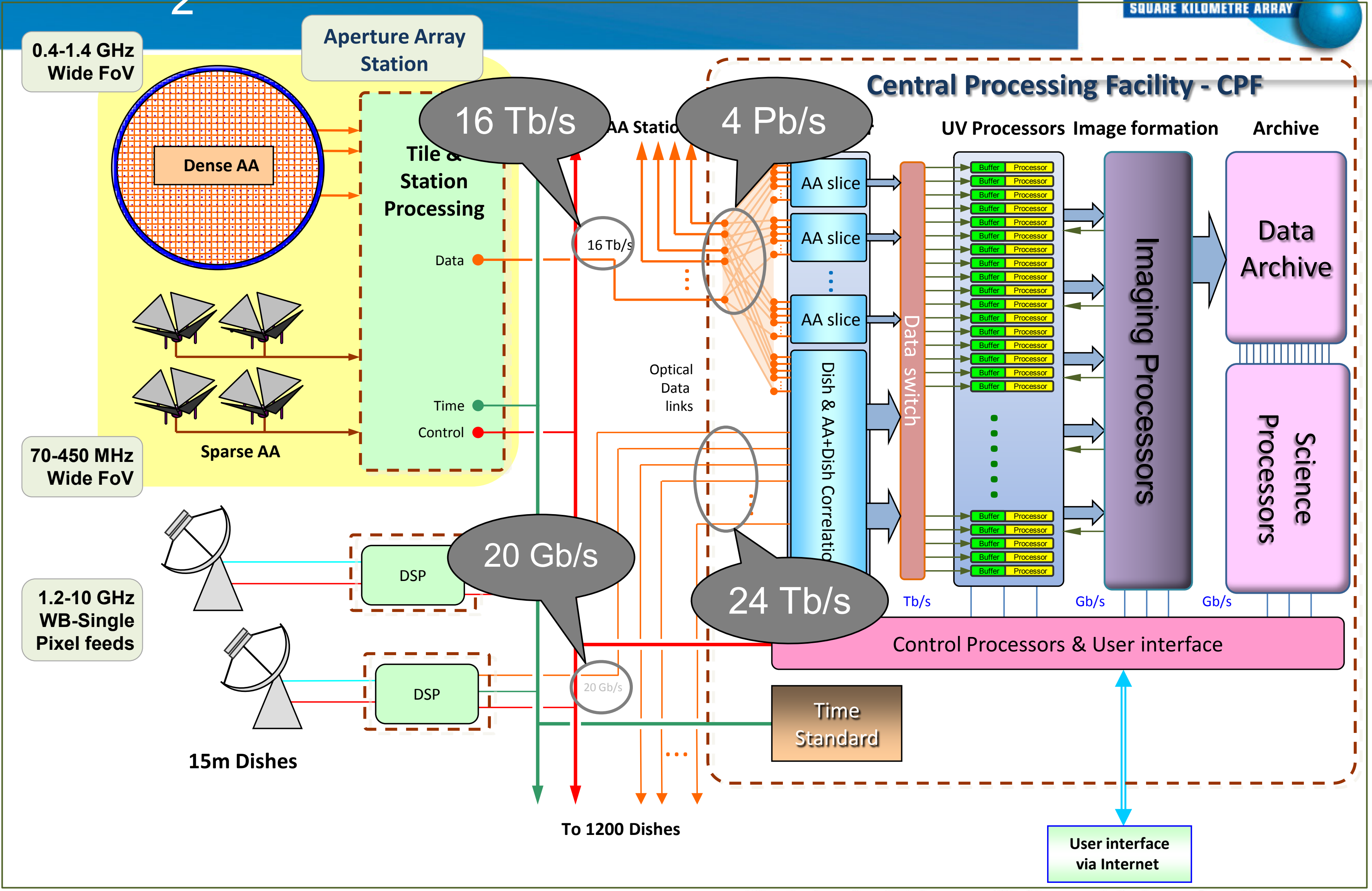
Artist renditions from Swinburne Astronomy Productions

Data rates from receptors

- For all receptors
  - Assume we Nyquist sample each polarization
- For dishes with single pixel feeds data rate is just determined by the bandwidth
- For Aperture Arrays and Dishes with Phased Array Feeds the data rate is determined by:
  - Bandwidth times beams across the band
  - Driven by the requirements of Field of View from the science requirements
  - Beam form at the AA station, but cannot time average as we need to keep data coherent
  - Reduce data rate by producing fewer beams than individual antenna elements
- Can (in principle) trade bandwidth for field of view so the product is the fixed quantity

$$G_1 = 2N_p \Delta f N_{bit} N_b = 4 \Delta f N_{bit} N_b$$

SKA$_2$ wide area data flow



Exploring the Universe with the world's largest radio telescope

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**

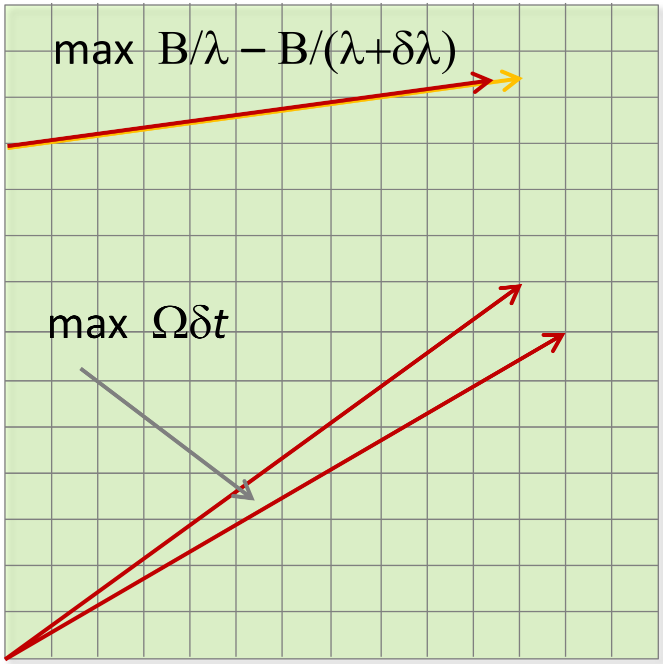
Data rates from the correlator

- 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

- Adopted 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 correlator 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

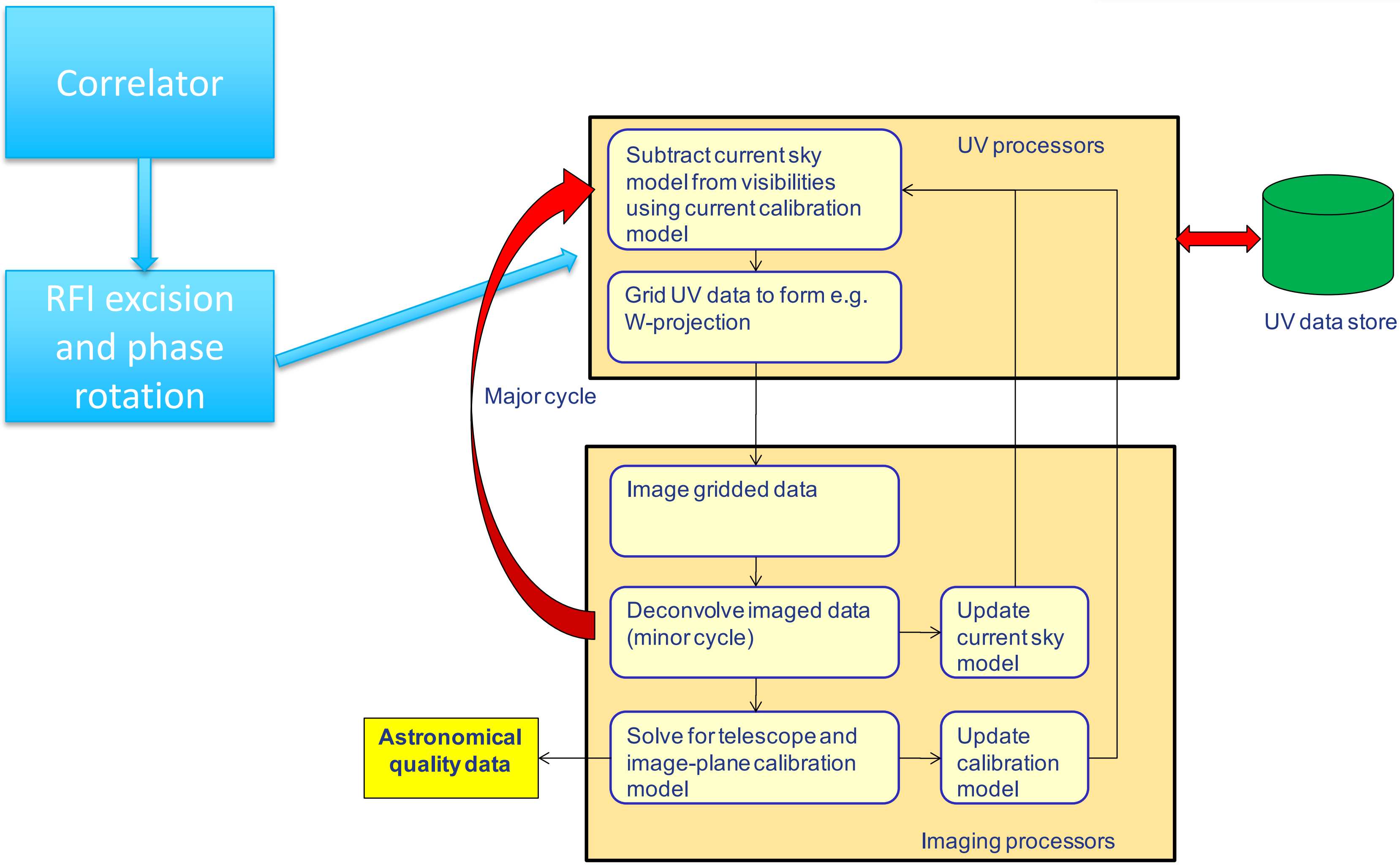
Example Data Products SKA2

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: imaging



Where does the data rate drop?

- In current interferometers the data rate drops dramatically out of the correlator
- This is true for SKA1 but not SKA2
- RFI excision needs to be done at high spectral and temporal resolution
- Pulsar search likewise
- In the imaging pipeline data can be averaged after RFI excision

Where does the data rate drop?

For SKA_2

Data rate out of correlator exceeds input data rate for 15-m dishes for baselines exceeding ~130km (36km if single integration time)

At best for dishes output data rate ~ input; AA's reduction by ~10^4

- Image size: $a^2 N_{ch} (B/D)^2 N_b$ Ratio UV to "image" data

$$\sim 0.06\, T_{obs}\, N^2 g(B) \frac{\Delta f}{f} \frac{1}{a_t a_f} \frac{1}{a^2} \frac{N_p^2}{N_{ch}} \quad \sim 210 \left(\frac{T_{obs}}{1\mathrm{min}}\right) \left(\frac{N}{1000}\right)^2 \left(\frac{N_{ch}}{32000}\right)^{-1}$$

Major reduction in data rate occurs between UV data and image data

Final comments

- SKA1 Challenging, but numbers do not scare industry only interest them
- SKA2 will require significant developments

Data flow through correlator
Example: Software correlator