Square Kilometre Array eInfrastructure: Requirements, Planning, Future Directions

Duncan Hall

SPDO Software and Computing

ESFRI @ EGEE 2009

Outline

SKA – in a nutshell
How do radio telescope arrays work?
What are the SKA's prime characteristics?
Real-time data: pushing the HPC envelope
How much data do we need to store?
Where are we at?
What is the future direction?
Summary

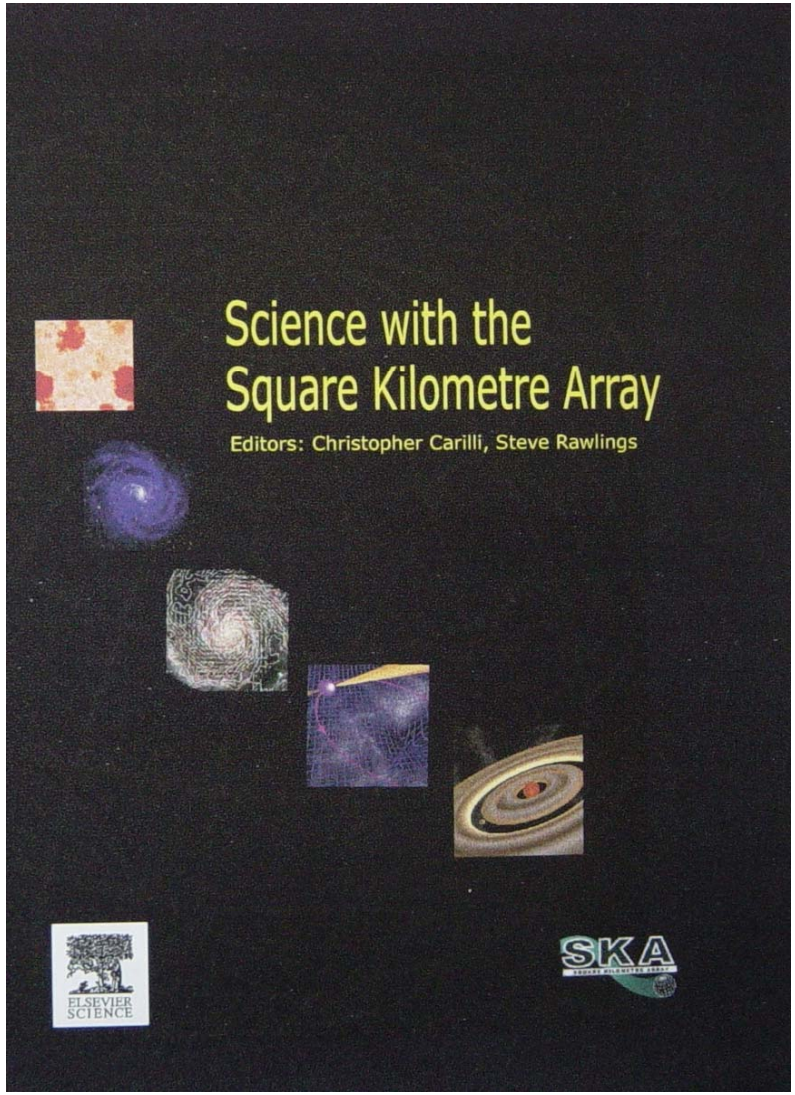
A global project:

75+ institutes in 19 countries

Similarities to CERN?

Science drivers

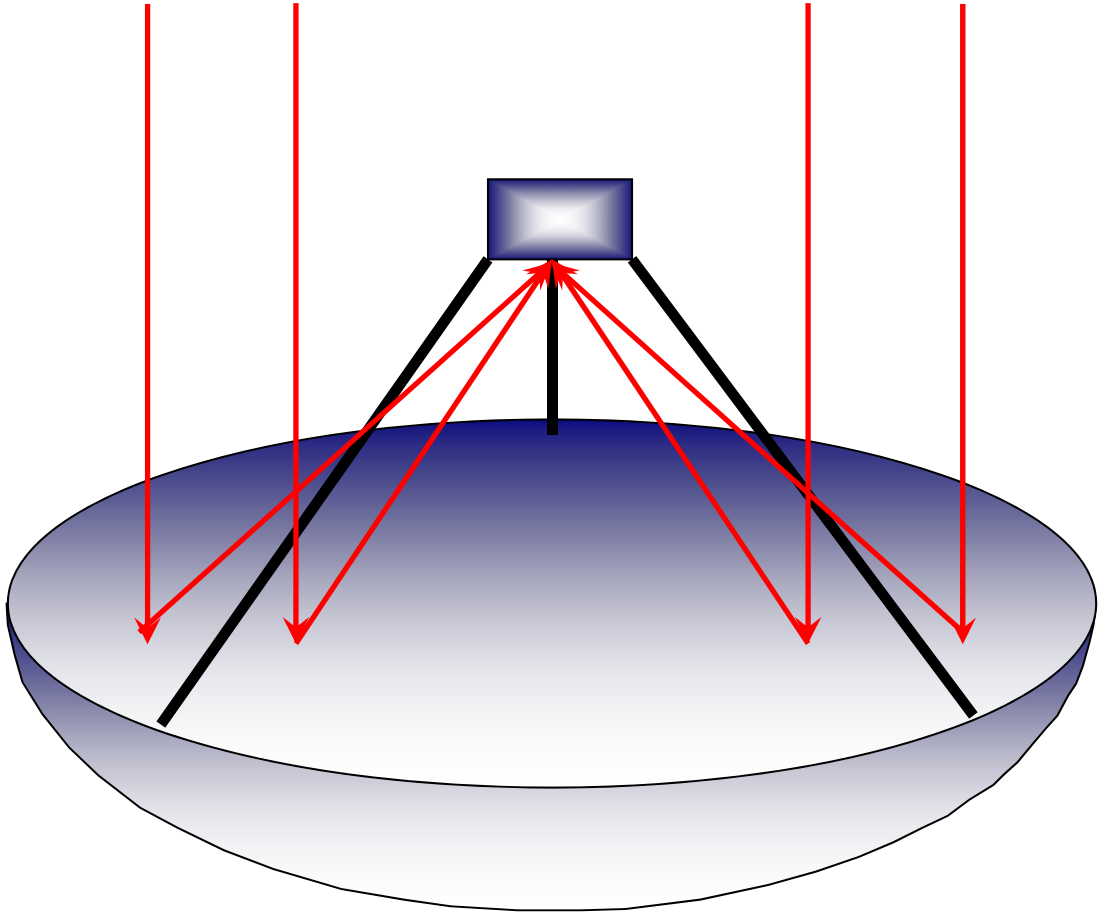
- Origins
  - Cosmology and galaxy evolution
  - Galaxies, dark matter and dark energy
  - Probing the “Dark Ages”
  - Formation of the first stars
  - Cradle of life
  - Search for signs of life
- Fundamental Forces
  - Strong-field tests of general relativity
  - Was Einstein correct?
  - Origin and evolution of cosmic magnetism
  - Where does magnetism come from?
- Exploration of the Unknown



Adapted from R. Schilizzi

How do radio telescope arrays work?

How telescopes work

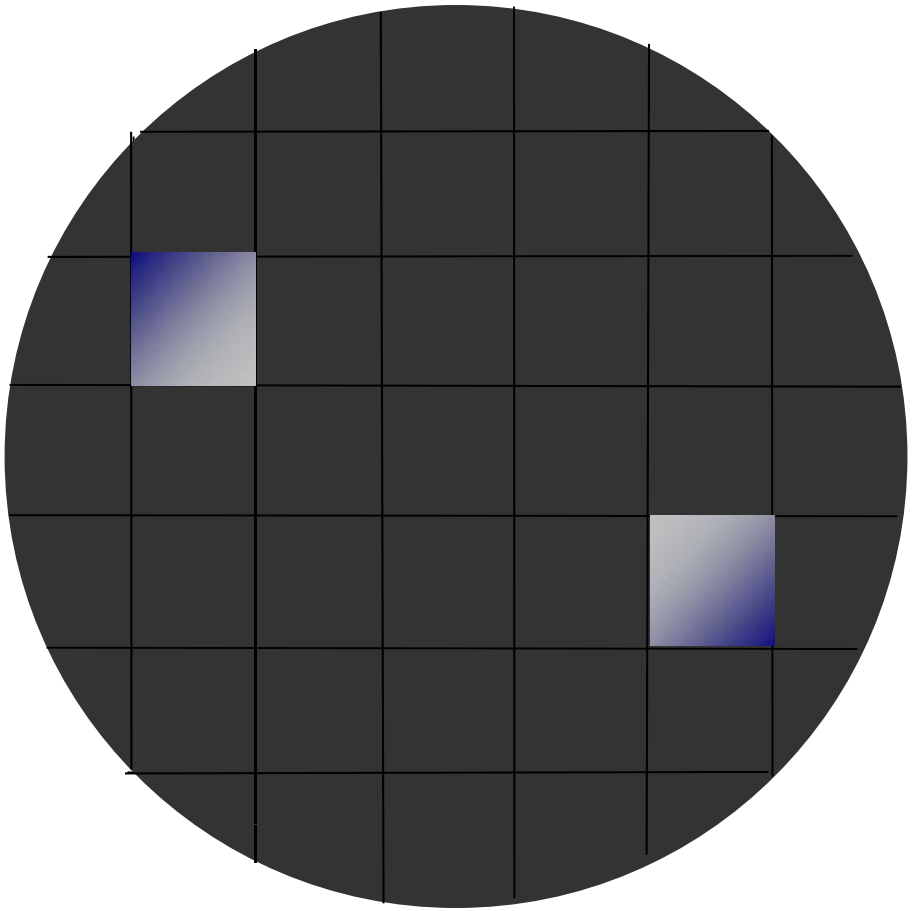


Conventional telescopes reflect the light of a distant object from a parabolic surface to a focus

M. Longair via R. Schilizzi

76 m: 3,000+ tonnes

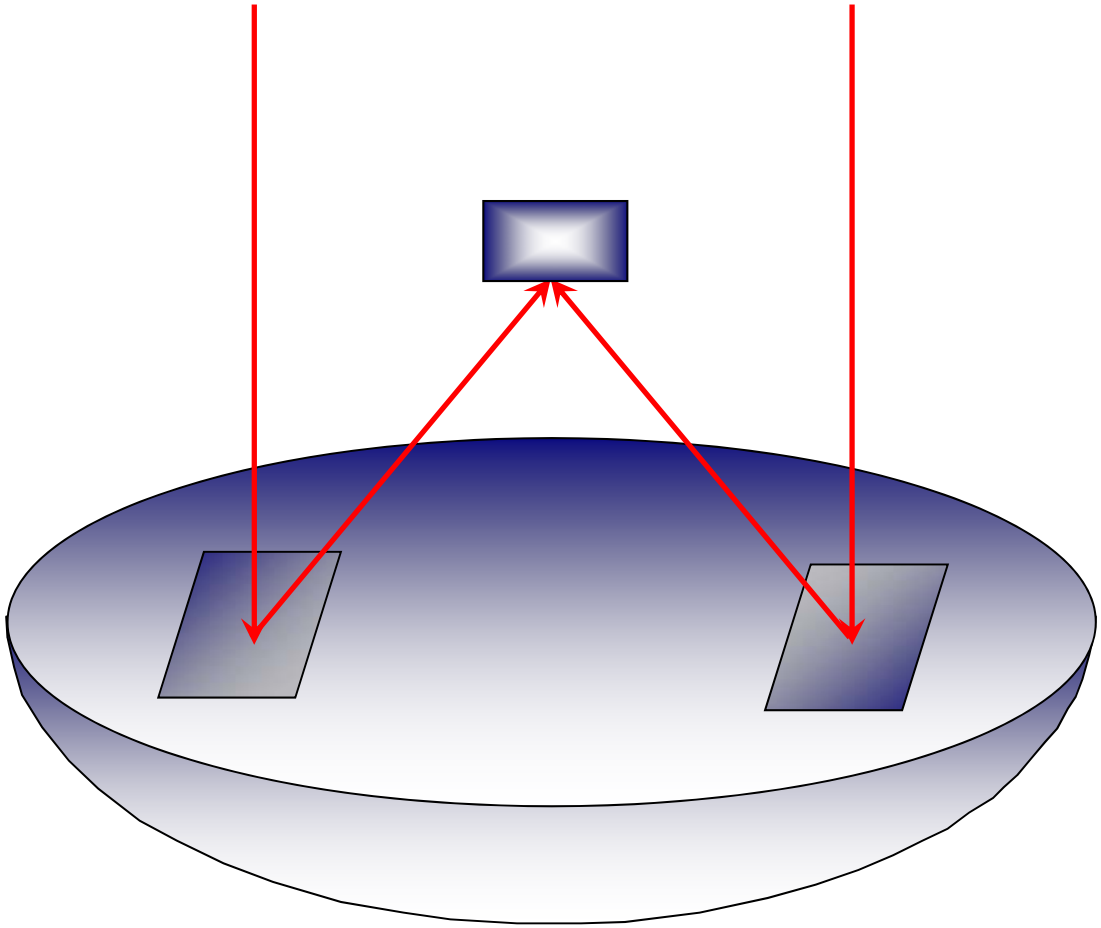
A partially filled aperture ...



But the reflecting surfaces do not need to be part of the same surface

Suppose we cover up most of the surface of the mirror

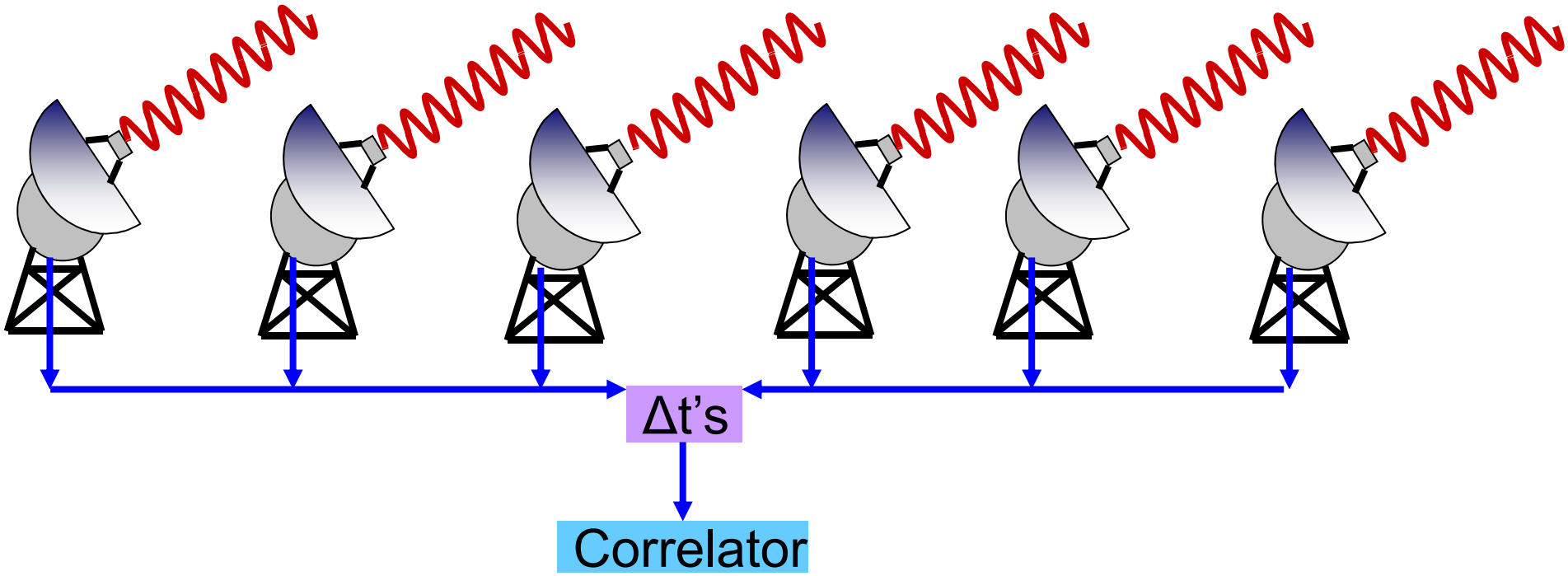
... can produce images



We can still combine the radiation from the uncovered sections to create an image of the distant object, if we arrange the path lengths to the focus to be the same.

Radio interferometry

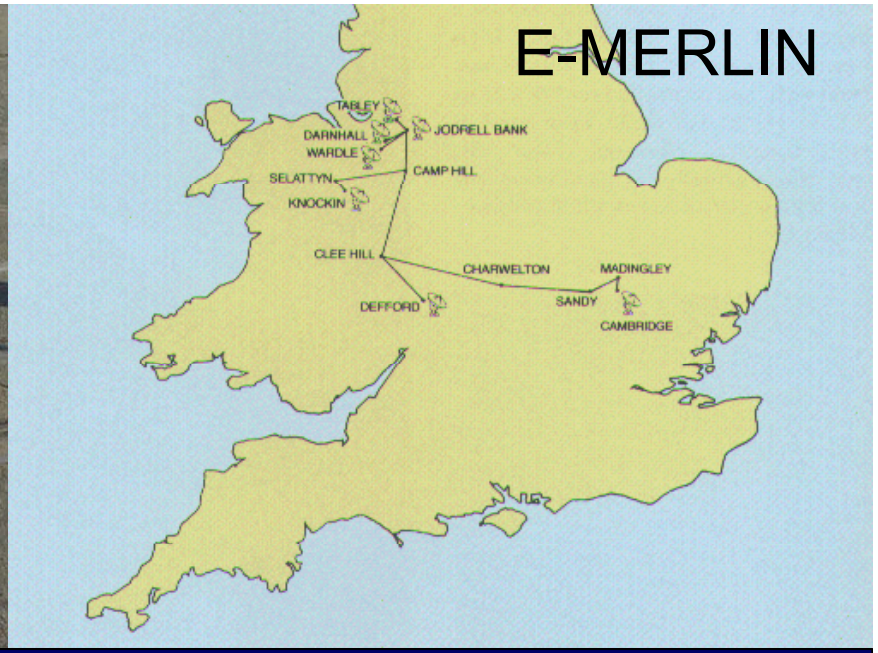
- Each pair of antennas is called a baseline
- More different baselines → more detailed the image
- Short baselines - antennas are close to each other - provide coarse structure
- Long baselines provide the fine detail: the longer → the finer the detail

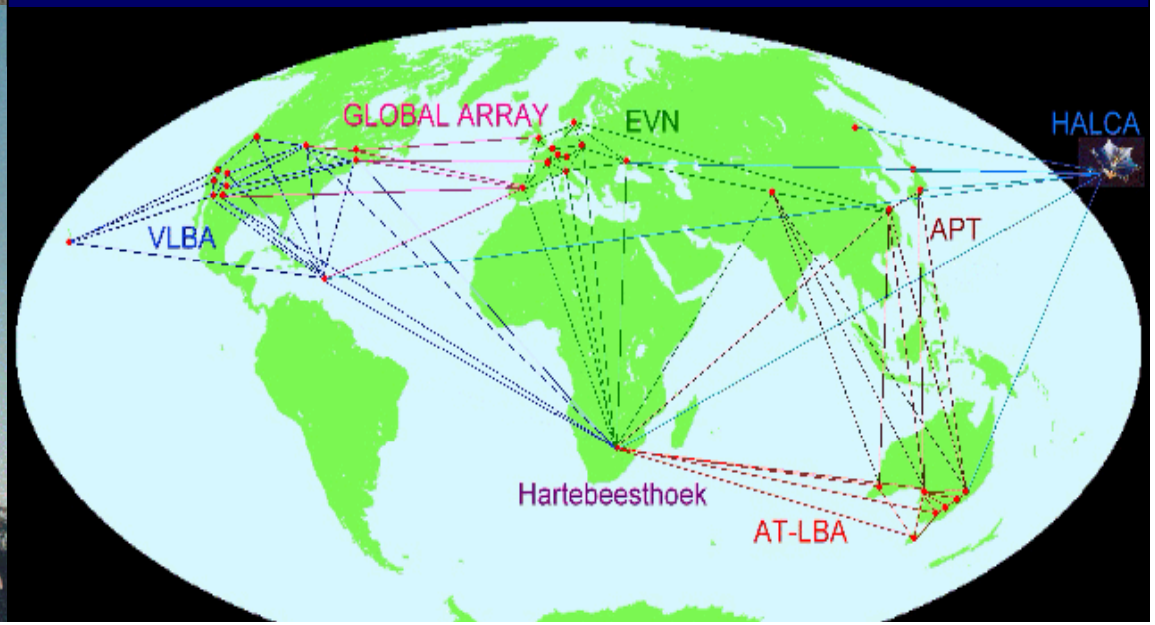


A 6 antenna interferometer has 15 baselines

M. Longair via R. Schilizzi

Array Telescopes





What are the SKA's prime characteristics?

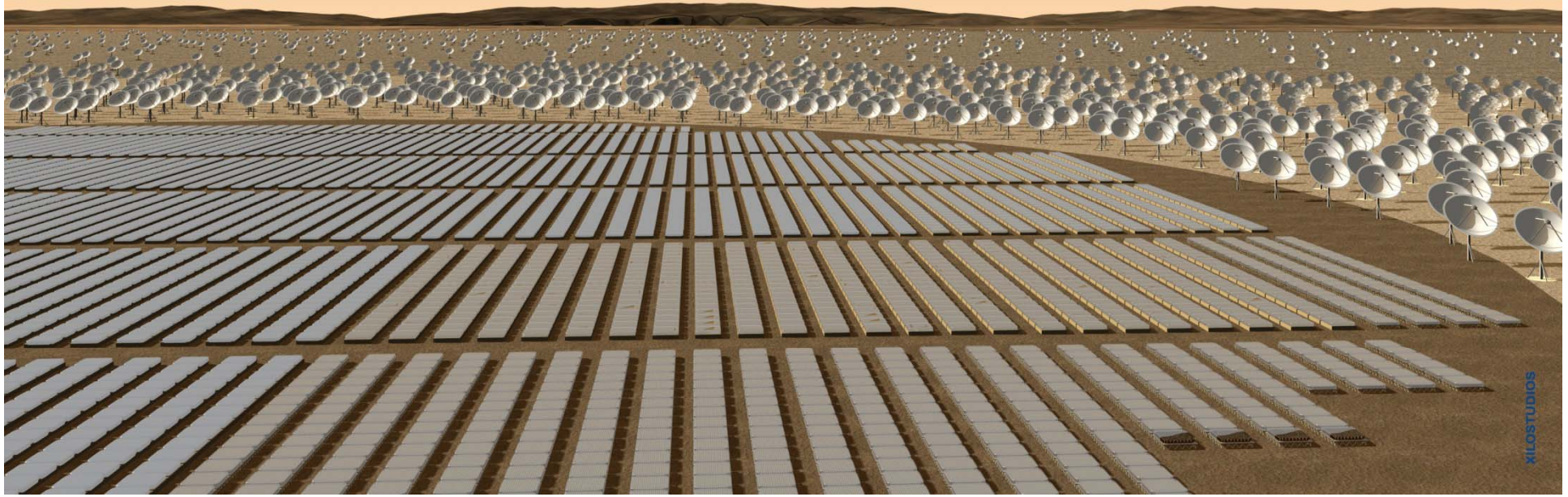
SKA: prime characteristics

1. More collecting area: ~1km^2
   - Detect and image hydrogen in the early universe
   - Sensitivity ~ 50 x EVLA, LOFAR
2. Bigger field of view
   - Fast surveying capability over the whole sky
   - Survey speed ~ 1,000,000 x EVLA
3. Wide ranges of frequencies
   - Low : 70-300 MHz
   - Mid: 300 MHz-10 GHz
   - High: 10-25+ GHz
4. Large physical extent : ~3,000+ km
   - Detailed imaging of compact objects
   - Astrometry with ~0.001 arc second angular resolution

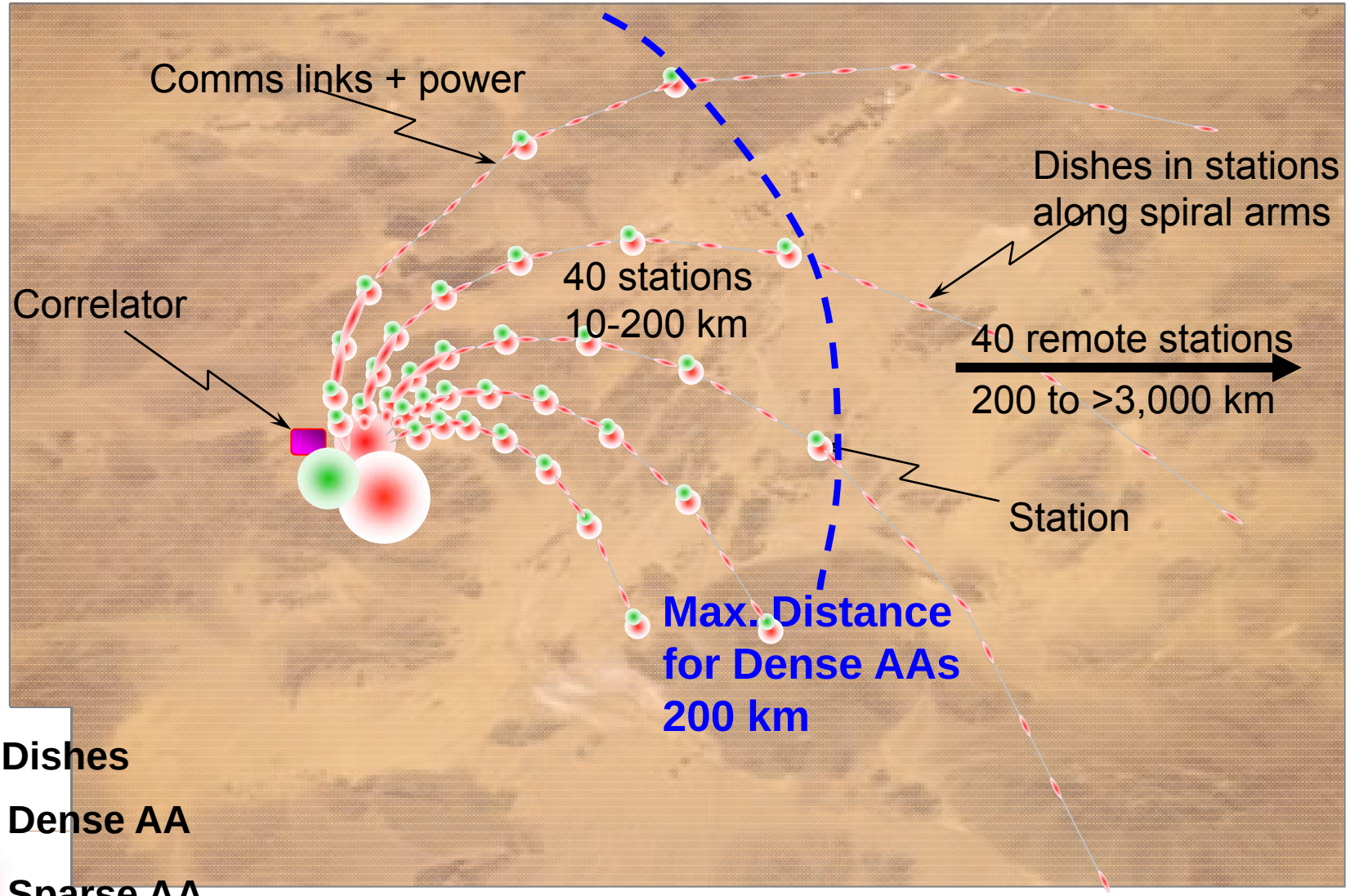
Adapted from R. Schilizzi

Artist's impression

- 1,500 dishes (~15m diameter) in a 5 km core
- Additional 1,500 dishes from 5 km to ~3,000+ km
- Aperture arrays (AA) in a core
- Signal processing: (1) beam forming
- Optical fibre connection to (2) correlator
- Optical fibre to remote (3) High Performance Computer

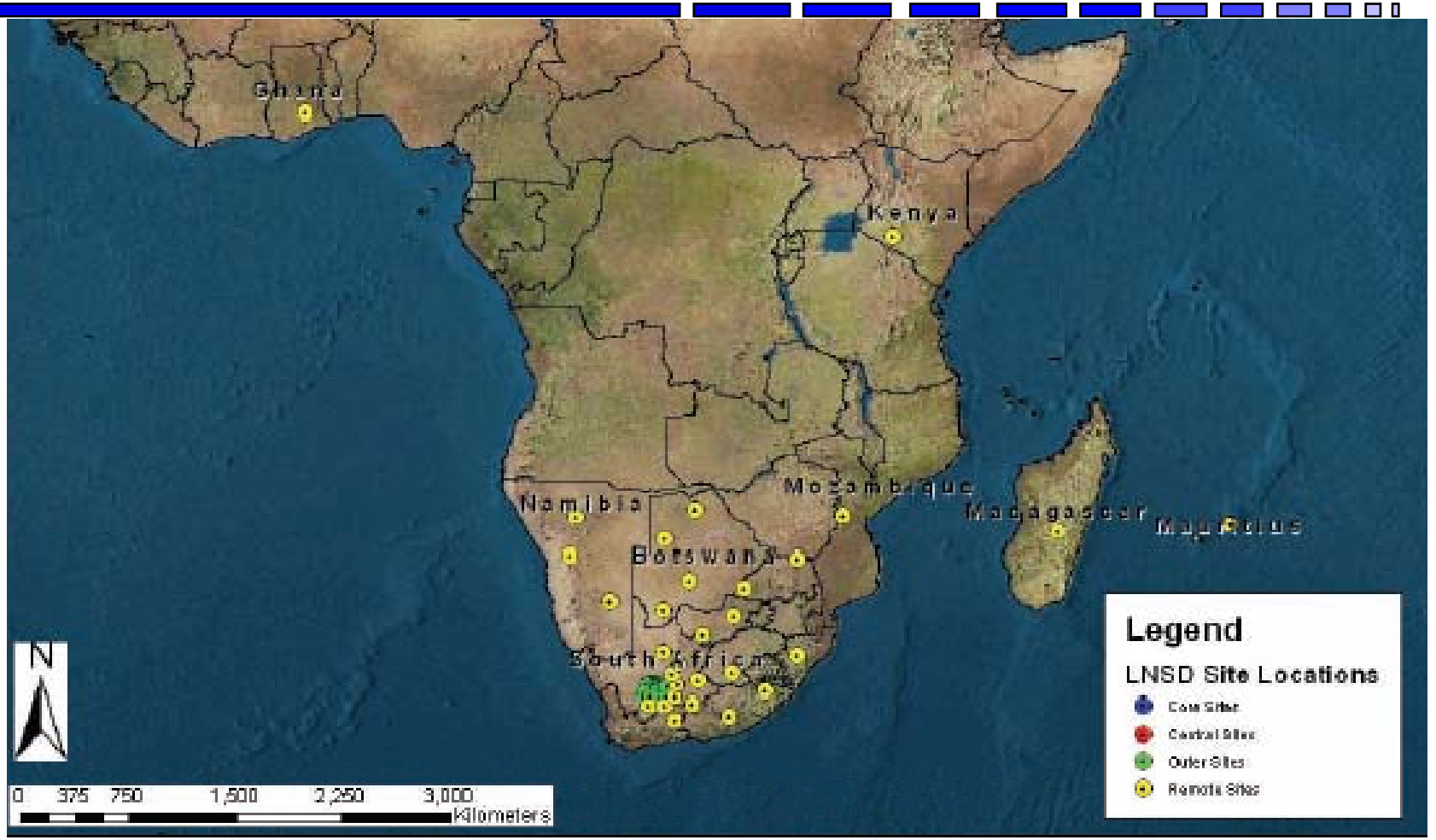


One possible configuration



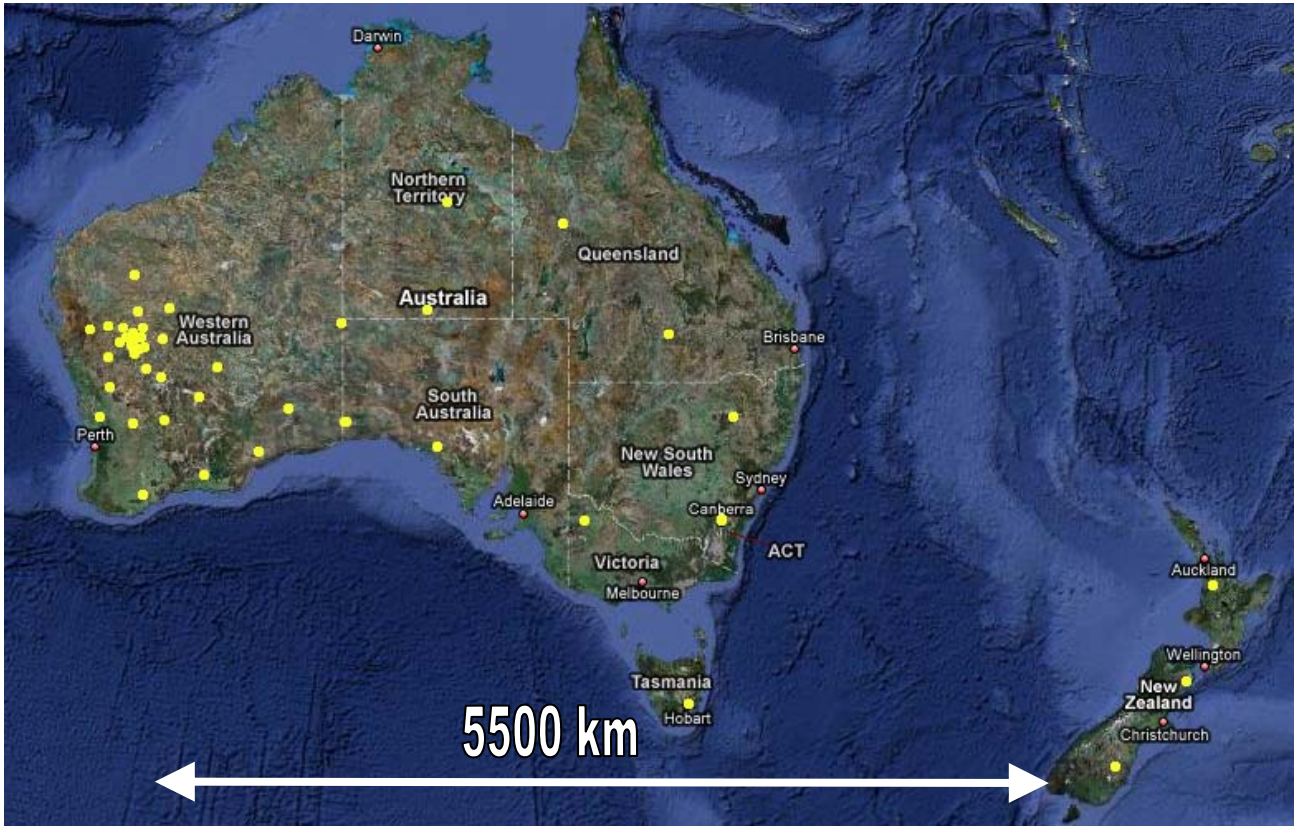
Adapted from A. Faulkner

South Africa + 7 countries



Adapted from R. Schilizzi

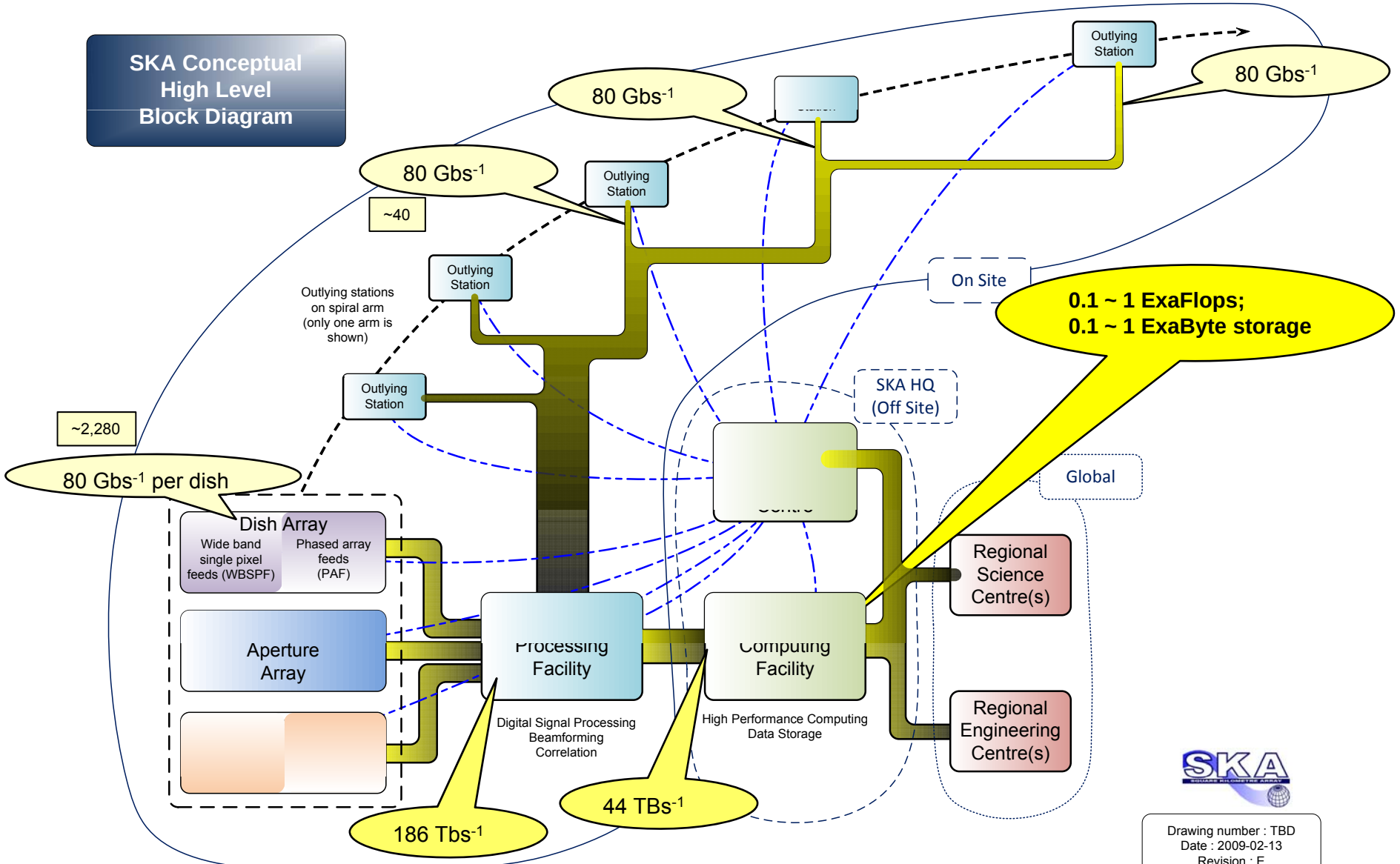
Australia + New Zealand



Adapted from P. Hall

Real-time data:
pushing the HPC envelope ...

Φ2 real-time data from dishes

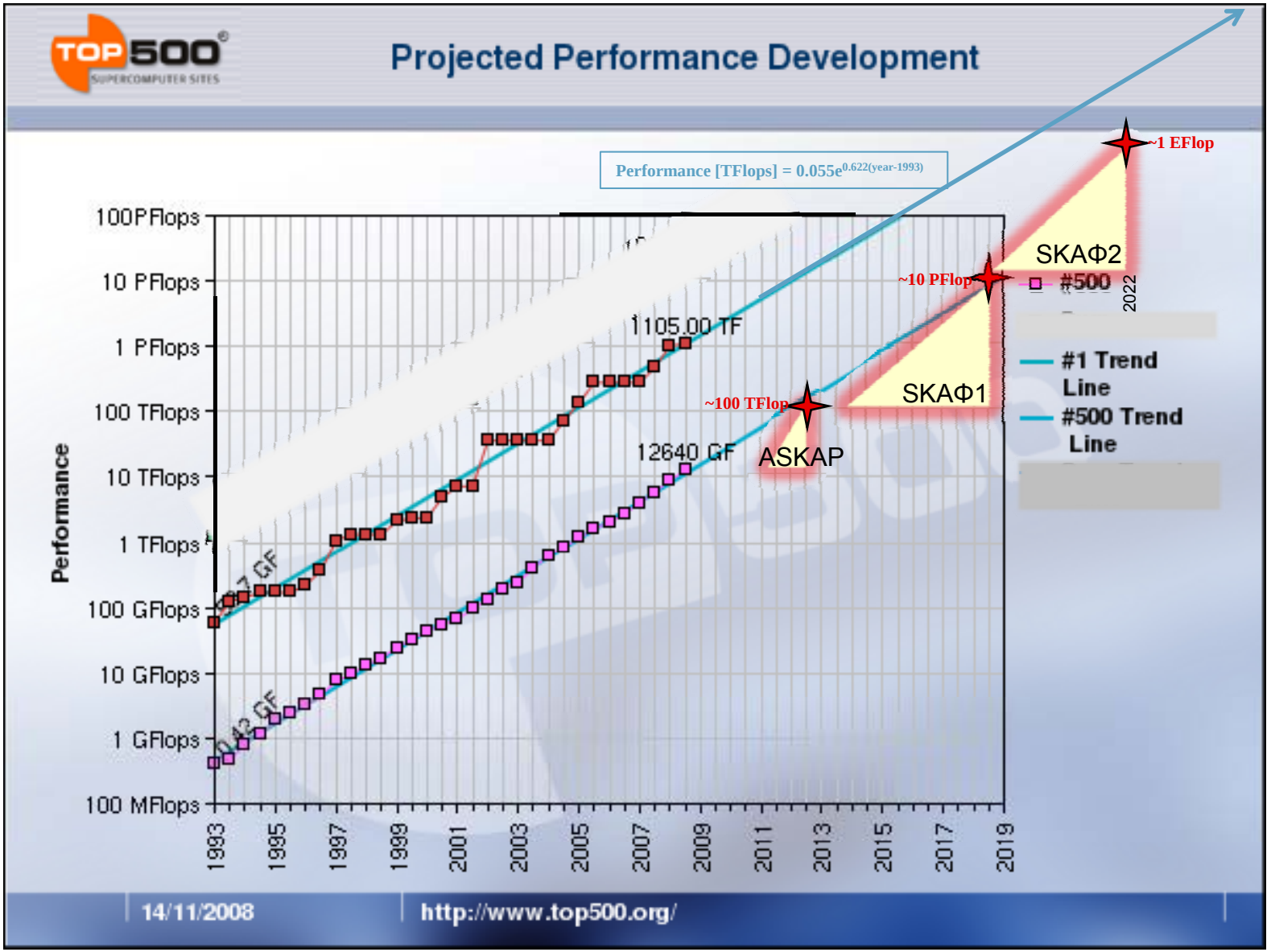


Drawing number : TBD
Date : 2009-02-13
Revision : E

P. Dewdney et al. "The Square Kilometre Array"; Proceedings of the IEEE; Vol. 97, No. 8; pp 1482 – 1496; August 2009

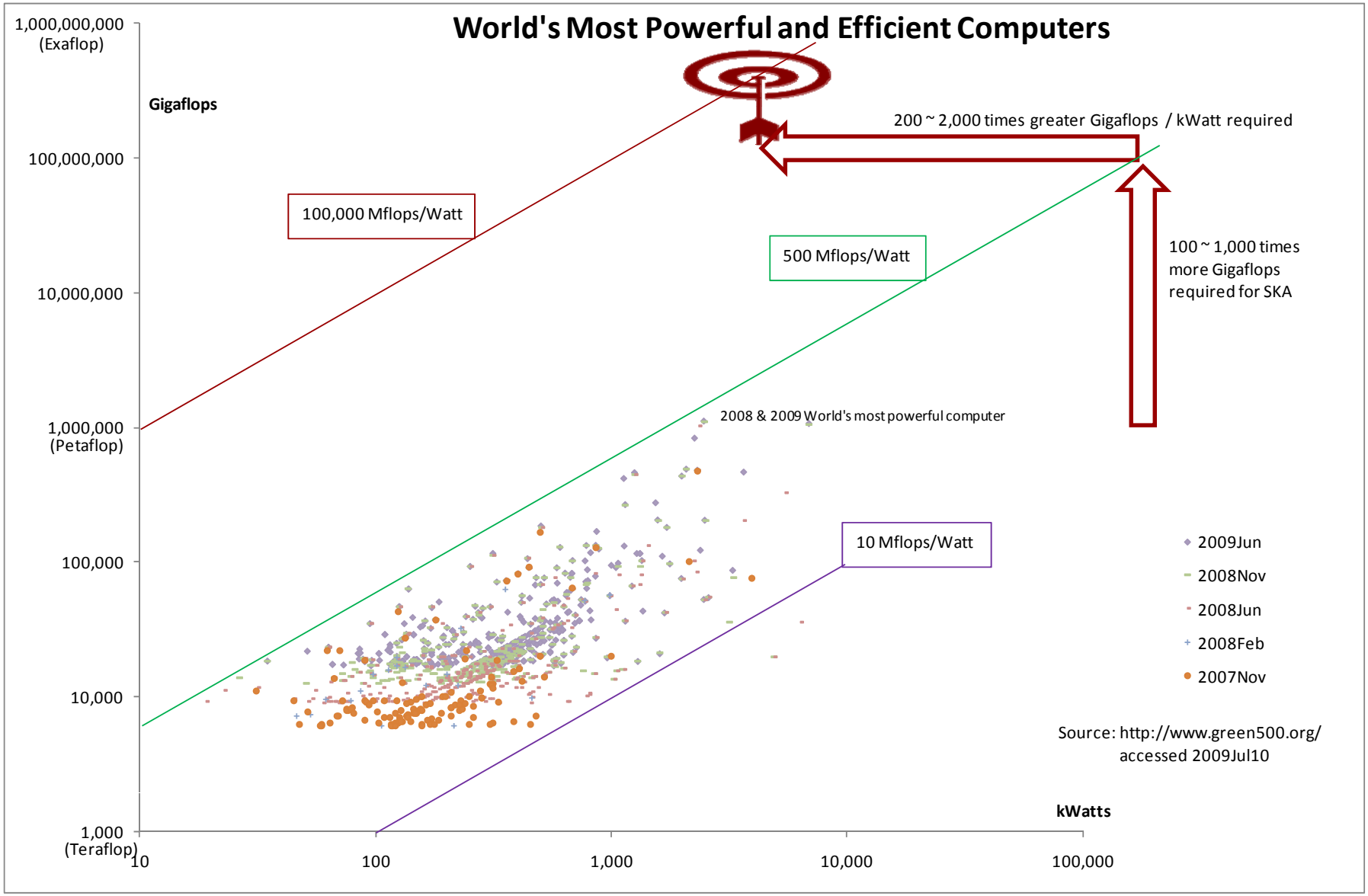
J. Cordes "The Square Kilometre Array – Astro2010 RFI #2 Ground Response" 27 July 2009; Table 1, pp 9 - 10

Pushing the Flops envelope



Cornwell and van Diepen “Scaling Mount Exaflop: from the pathfinders to the Square Kilometre Array” http://www.atnf.csiro.au/people/Tim.Cornwell/MountExaflop.pdf

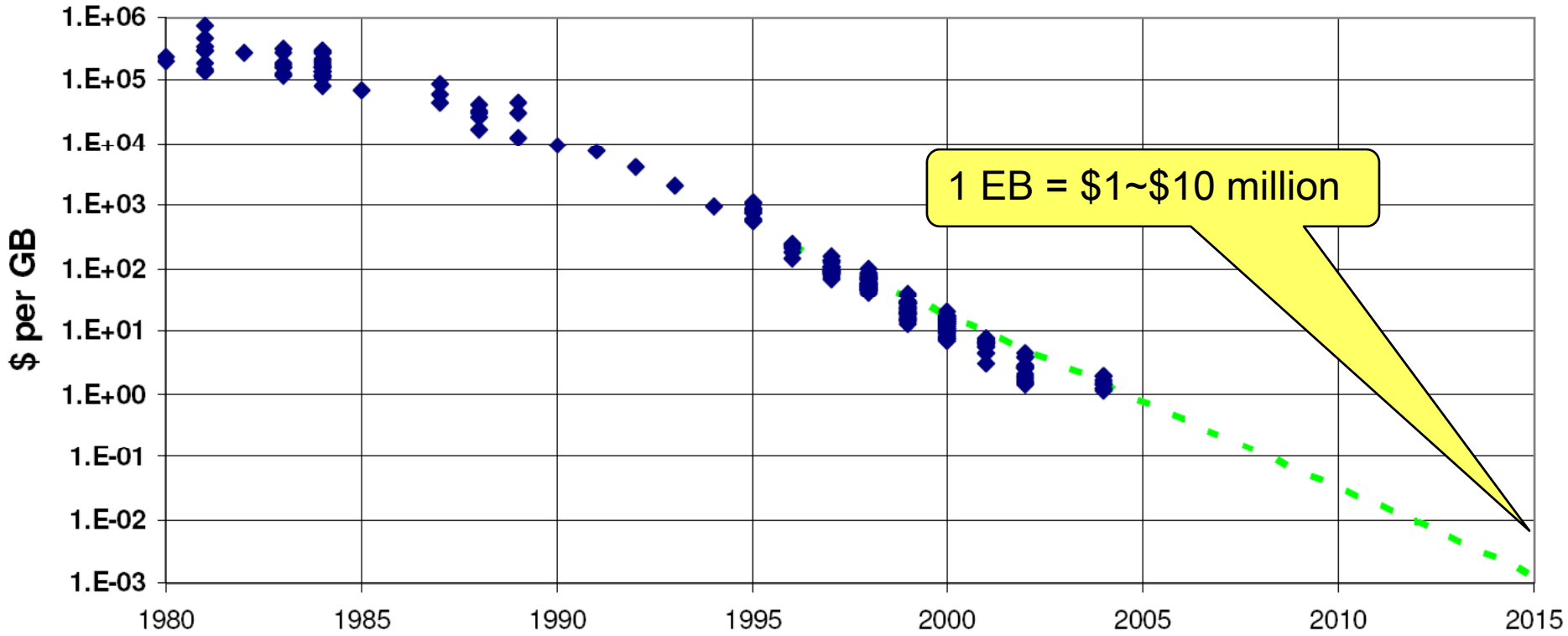
Power efficiency challenges:



http://www.green500.org accessed July 2009

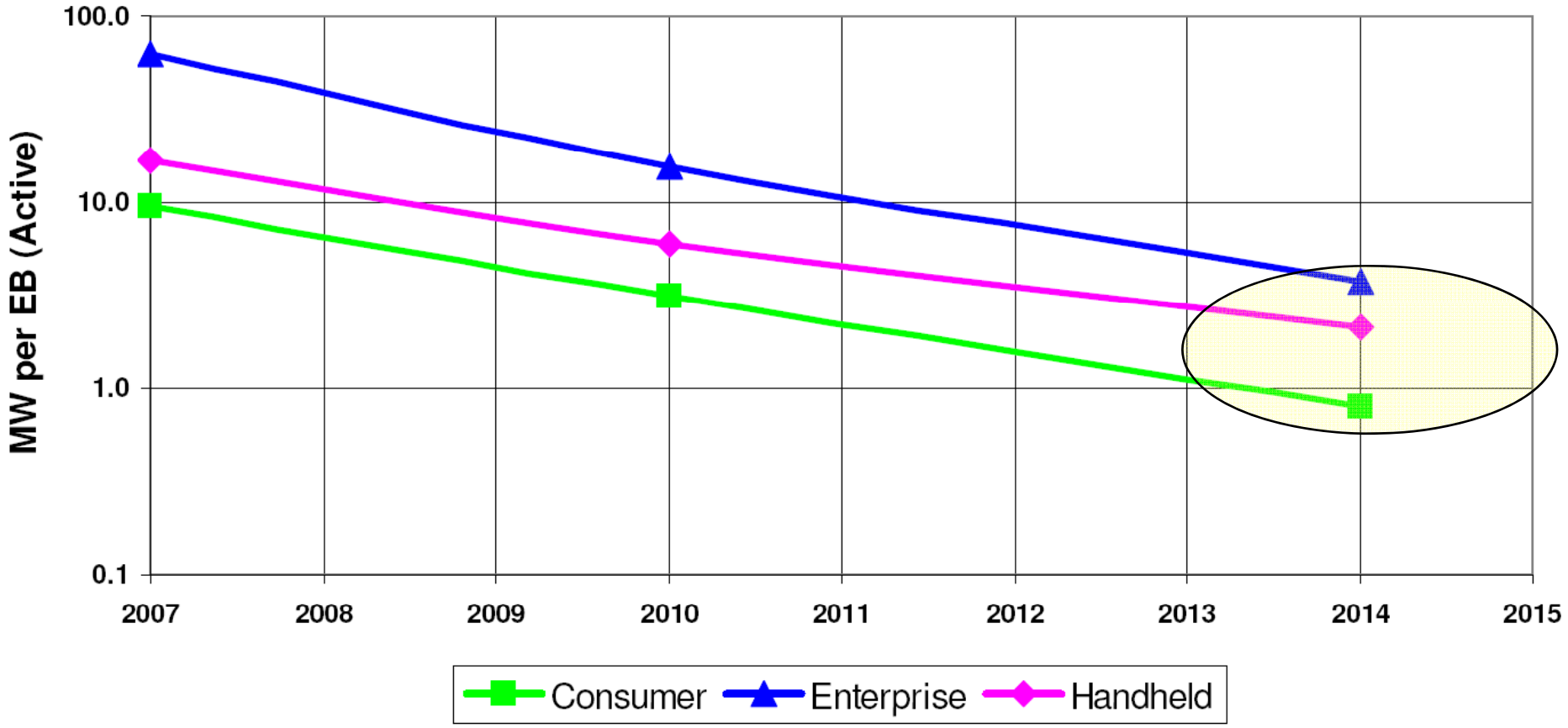
How much data do we need to store?

Disk storage: annual 50% cost reduction



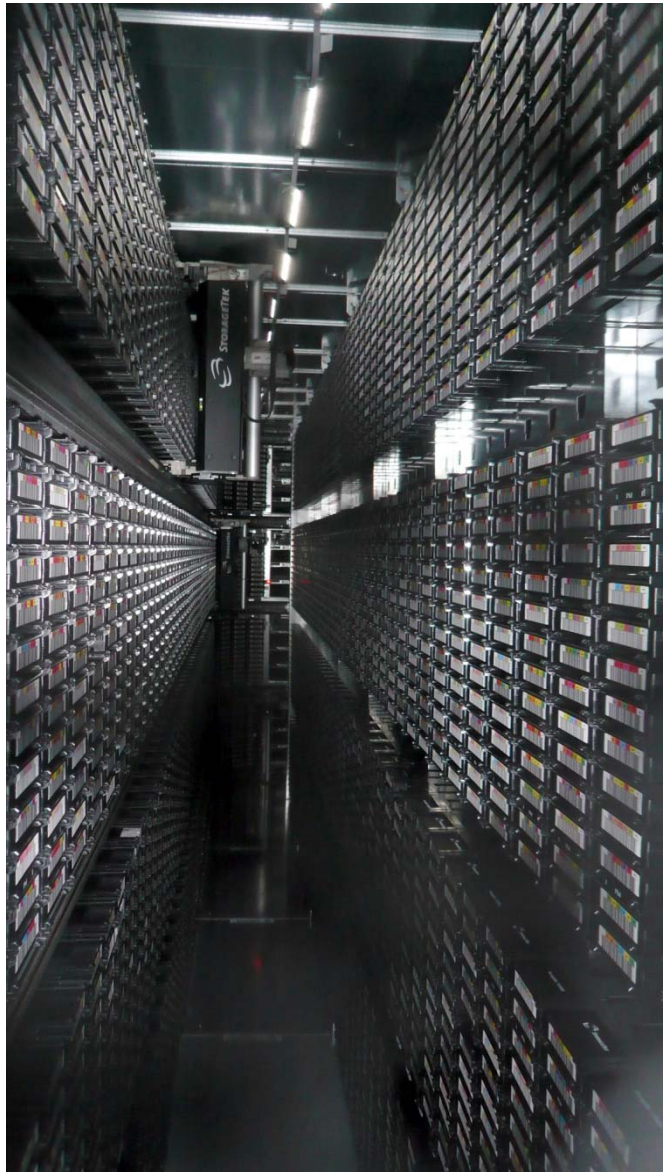
P. Kogge et alia “ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems”; TR-2008-13, DARPA ExaScale Computing Study, 2008 Sep 28, page 125
Note: neither RAID, controllers, nor interconnect cables are included in these estimates

Power for EB-size disk looks reasonable



P. Kogge et alia "ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems"; TR-2008-13, DARPA ExaScale Computing Study, 2008 Sep 28, page 124
Note: power numbers here are for the drives only; any electronics associated with drive controllers [e.g. ECC, RAID] needs to be counted separately

10 PetaByte tape robot at CERN



500-GB tapes switched to 1-TB models – an upgrade that took a year of continuous load/read/load/write/discard operations, running in the interstices between the data centre's higher-priority tasks

http://www.flickr.com/photos/doctorow/2711081044/sizes/o/in/set-72157606675048531/
C. Doctorow "Welcome to the Petacentre"; Nature, Vol. 455, $ September 2008, pp. 17-21

Where are we at?

Current development

PrepSKA 2008-2011

The Preparatory Phase for the SKA is being funded by the European Commission's 7th Framework Program

€5.5M EC funding for 3 years + €17M contributed funding from partners (still growing)

€150M SKA-related R&D around the world

Coordinated by the Science and Technology Facilities Council (UK)

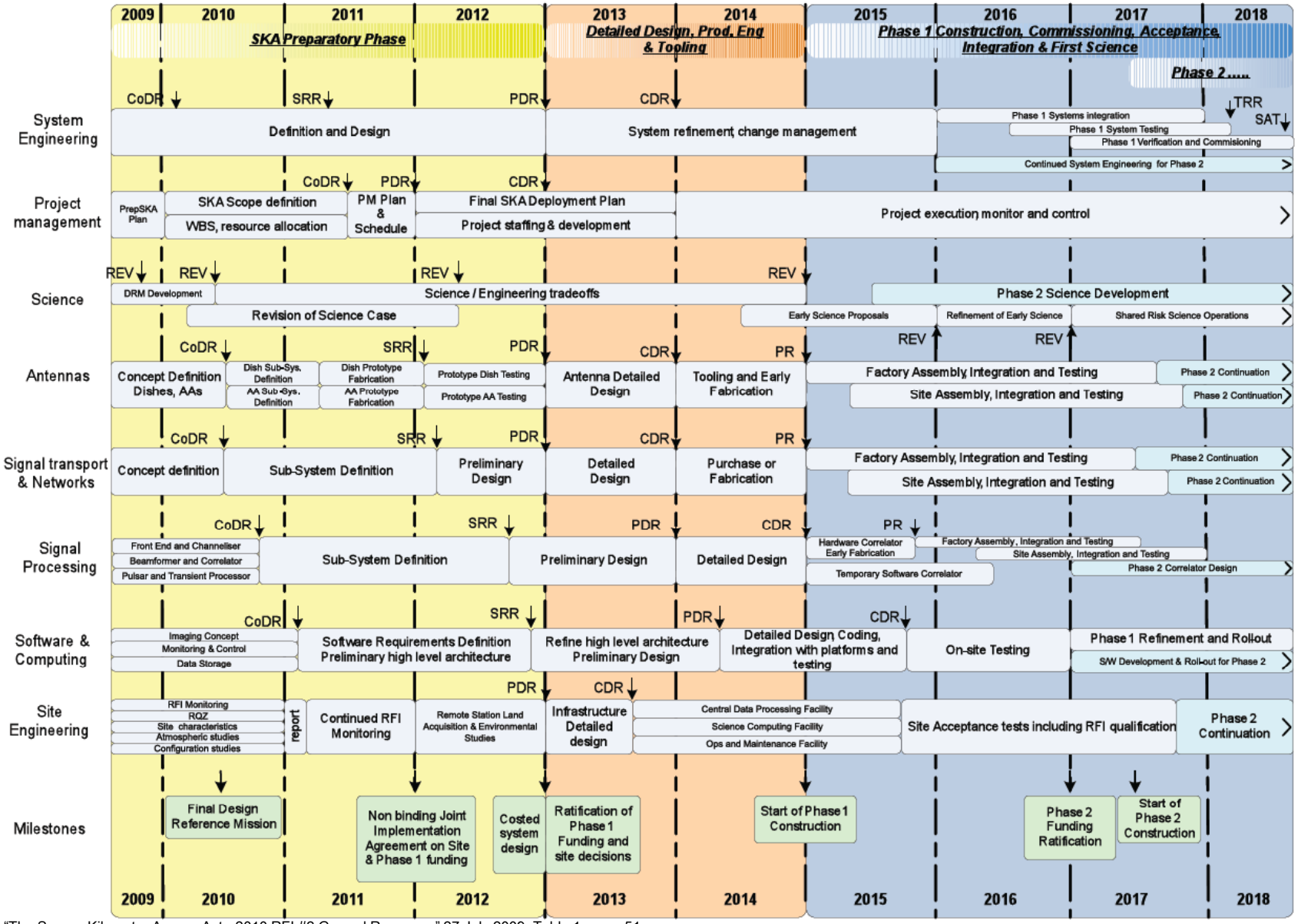
WP2: Design + Cost

Coordinated by the SKA Program Development Office in Manchester

- System Definition
- Dishes, feeds, receivers
- Aperture arrays
- Signal transport
- Signal processing
- Software
- High performance computers
- Data storage
- Power requirements

What is the future direction?

Schedule to 2018



J. Cordes "The Square Kilometre Array – Astro2010 RFI #2 Ground Response" 27 July 2009; Table 1, page 51

Phased construction

- Phase 1:
  - 2013 – 2018 construction
  - 10-20% of the collecting area
- Phase 2:
  - 2018 – 2022 construction
  - Full array at low and mid frequencies
- Phase 3:
  - 2023+ construction
  - High frequencies

Summary

- **Required eInfrastructure in host region**
  - Distributed 100's Gb fibre optic signal transport and networks
  - Distributed 10's MW power for collecting dishes and stations
  - Central Digital Signal Processing Facility: specialised hardware
  - Fibre to Science Processing Facility
- **Other eInfrastructure: could be anywhere**
  - Science Processing Facility – ExaFlop "production" environment
  - Tiered hierarchy (1-3 ?) of ExaByte storage, archive and distribution
  - Software applications development and maintenance environments
  - Power for computing and (-H)VAC: 10's MW
- **Current state of planning**
  - Preparatory costing and high level design / architecture development