The Science Data Processor Overview

Professor Paul Alexander
and
Dr Rosie Bolton

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

SDP project scientist

SKA
SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

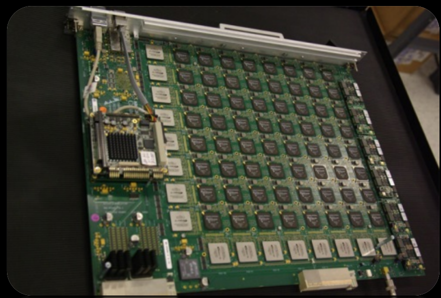
UNIVERSITY OF CAMBRIDGE

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

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)

HPC Processing
2020: 300 PFlop
2028: 30 EFlop

UNIVERSITY OF CAMBRIDGE

Architectural Approach

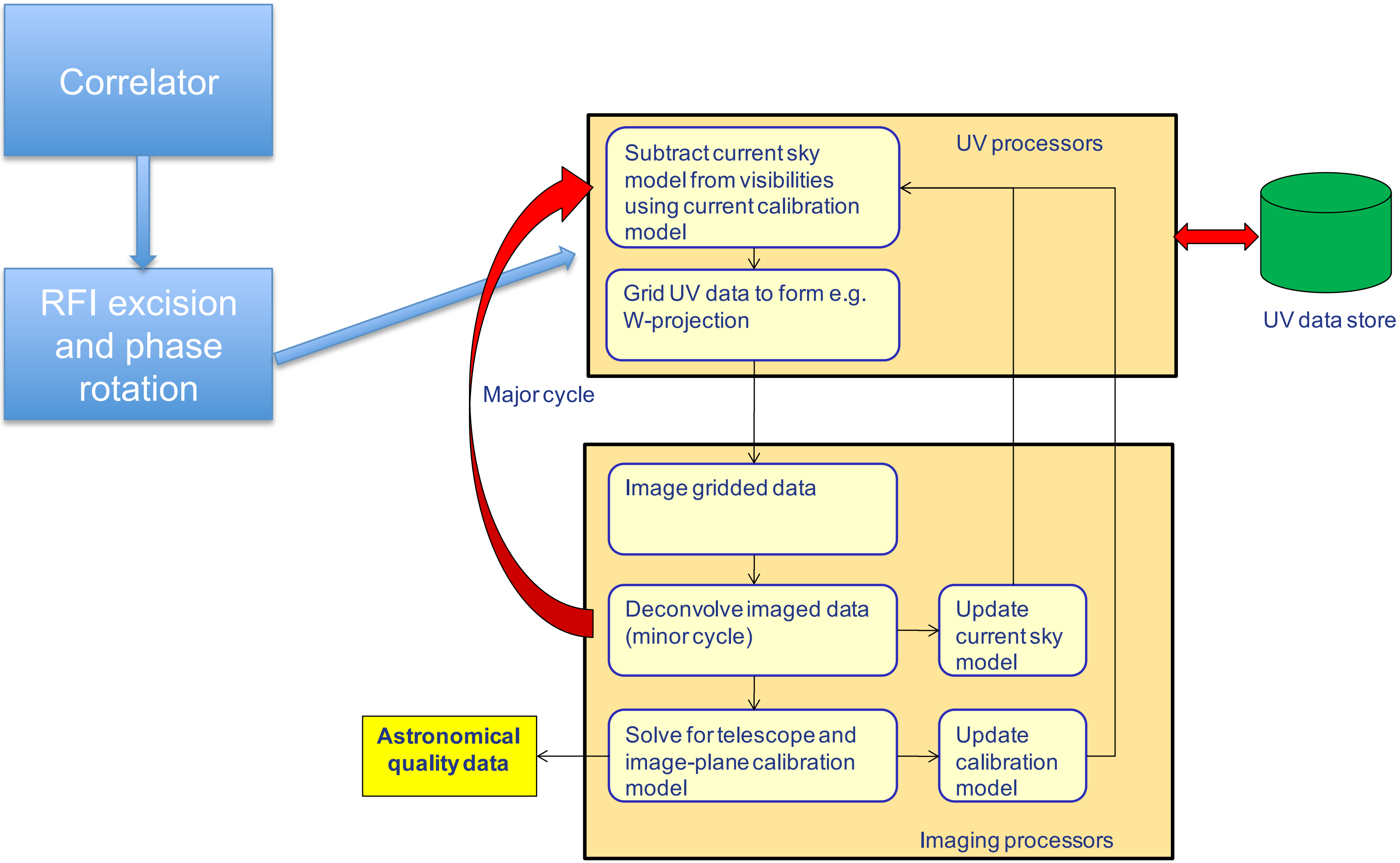
SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

- Main principles
  - Ensure scalability (downwards mostly)
  - Ensure affordability
  - Ensure Maintainability
  - Support current state-of-the-art algorithms
- Exploit data parallelism, frequency & 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

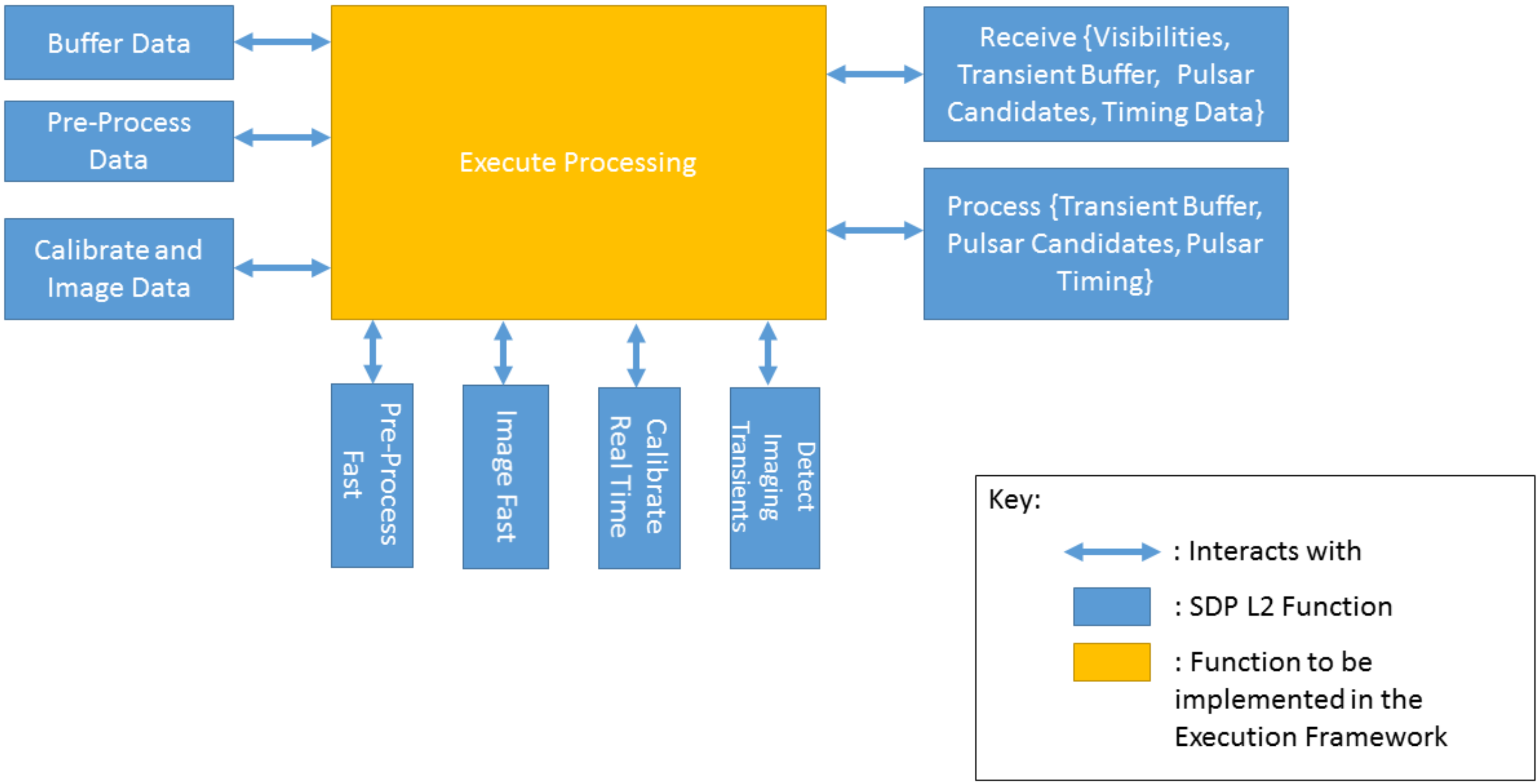
Approach: Co-design of software and physical layer architectures

UNIVERSITY OF CAMBRIDGE

Work Flows: Imaging Processing Model



Execution Framework



UNIVERSITY OF CAMBRIDGE

High Priority Science Objectives

SKA SQUARE KILOMETRE ARRAY | SDP | SCIENCE DATA PROCESSOR

- SKAO has developed a list of HPSO experiments – programmes targeting specific scientific goals and taking long periods (~5000-16000 hours) of telescope time.
- Draft schedule for these taking 5-15 years to complete
- Just an example

Sci Goal	SWG	Science Objective	SKA1 Compone	Hours
1	CD/EoR	EoR - I. Imaging	LOW	5000 hrs
2	CD/EoR	EoR - II. Power spectrum	LOW	10000 hrs
4	Pulsars	Reveal pulsar population	LOW+MID	12750+3200 hrs
5	Pulsars	High precision timing	LOW+MID	4300+3200 hrs
13	HI	Resolved HI out to z~0.8	MID	5000 hrs
14	HI	ISM in the nearby Universe.	MID	2000 hrs
15	HI	ISM in our Galaxy	MID	12600 hrs
18	Transients	Fast Radio Bursts	MID	10000 hrs
22	Cradle of Life	Map dust grain growth	MID	6000 hrs
27	Magnetism	All-Sky magnetic fields	MID	10000 hrs
32	Cosmology	Gravity on super-horizon scales.	MID	10000 hrs
33	Cosmology	Non-Gaussianity and the matter dipole.	MID	10000 hrs
37+38	Continuum	Star formation history of the Universe	MID	16000 hrs

100,000 hours = 11 years

We can use these to generate example SDP use cases and archive growth rates.
This enables **load balancing** because we relax latency requirement of **off-line processing.**

UNIVERSITY OF CAMBRIDGE

High Priority Science Objectives

HPSO	Total (PFLOPS)	Hours of telescope time	Fraction of time
U.HPSO-4a Pulsar Search MID SPF1	~0	800	0.01
U.HPSO-4b Pulsar Search MID SPF2	~0	2400	0.04
U.HPSO-5a Pulsar Timing MID SPF2	~0	1600	0.02
U.HPSO-5b Pulsar Timing MID SPF3	~0	1600	0.02
U.HPSO-13 Hi Kinematics and Morphology	25.6	5000	0.07
U.HPSO-14 Hi MID	32.7	2000	0.03
U.HPSO-15 Studies of the ISM in our Galaxy	26.2	12600	0.19
U.HPSO-18 Transients MID	~0	10000	0.15
U.HPSO-22 Cradle of Life MID Band 5	25.4	6000	0.09
U.HPSO-27 All Sky Magnetism	26.3	10000	0.15
U.HPSO-37a Continuum Survey MID band 2	28.1	2000	0.03
U.HPSO-37b Continuum Survey MID band 2 (deep)	28.1	2000	0.03
U.HPSO-37c Continuum survey, band 2 wide	28.1	10000	0.15
U.HPSO-38a Continuum Survey MID band 5	26.1	1000	0.01
U.HPSO-38b Continuum Survey MID band 5	26.1	1000	0.01
Weighted average FLOPS value for MID HPSOs			20 PFLOPS
Approximate AVERAGE Apparent power requirement[2]			~2.7 MVA

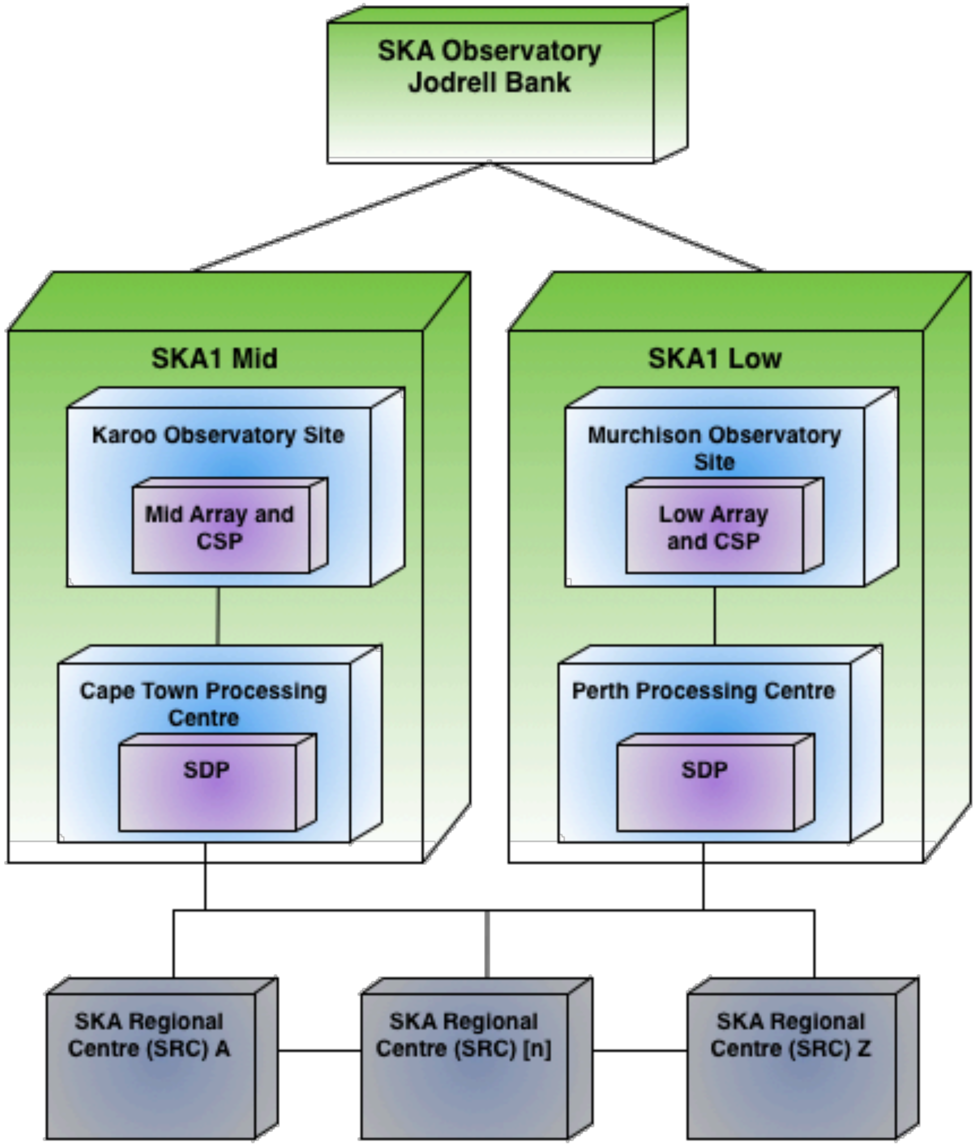
Average FLOPS values 3.5x lower for MID than maximal case.

Relax latency requirement and save both capital cost and power cost

Overall designing to a 250 PFlop peak system (average efficiency ~10% driven by likely memory bandwidth)

UNIVERSITY OF CAMBRIDGE

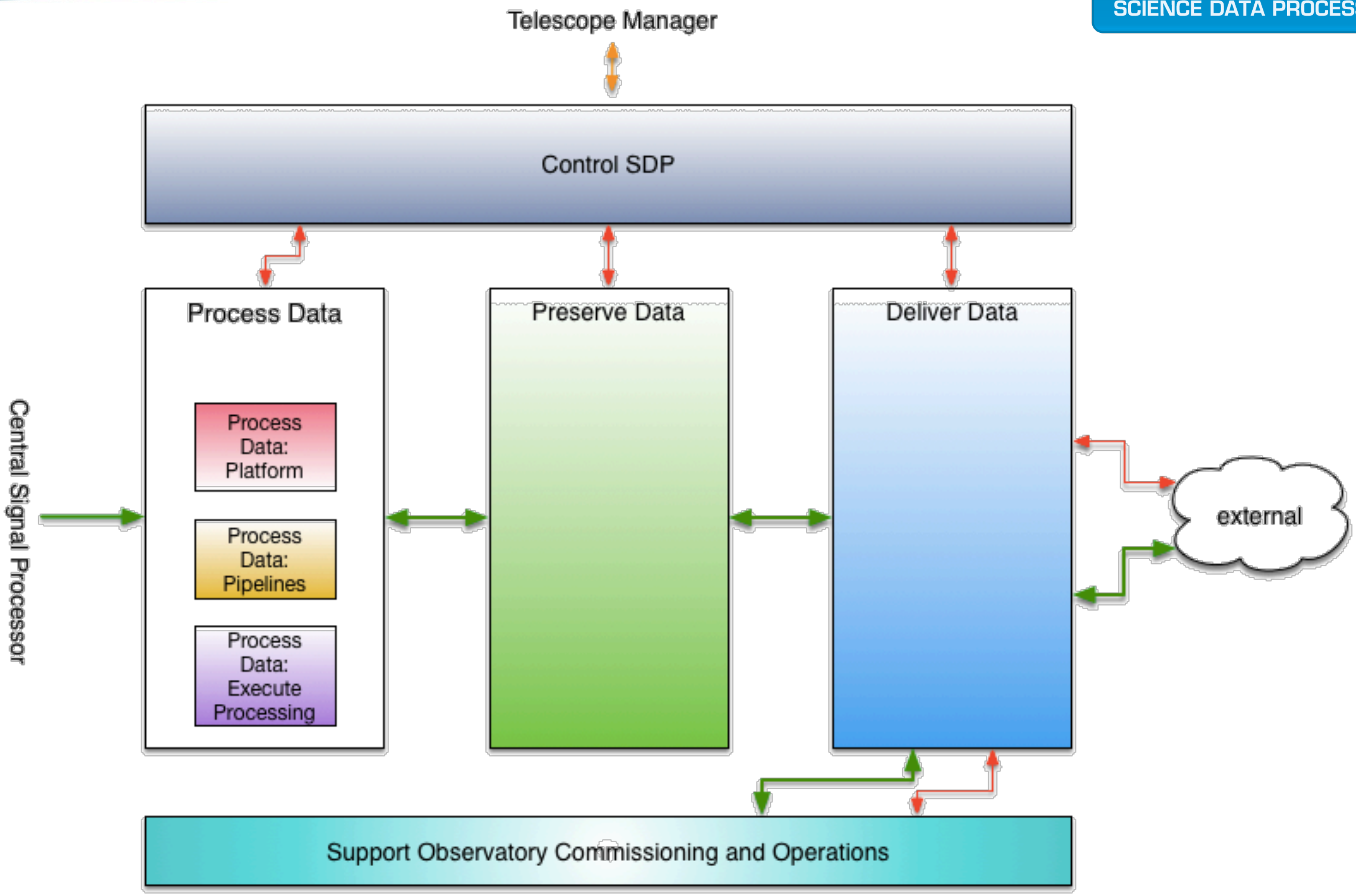
One SDP Two Telescopes



	Ingest (GB/s)
SKA1_Low	400
SKA1_Mid	400

In total need to deploy eventually a system which is around 0.25 EFlop of processing

The SDP System



UNIVERSITY OF CAMBRIDGE

Data rates and processing increase by FACTOR ~100 for SKA2

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

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

UNIVERSITY OF CAMBRIDGE

END