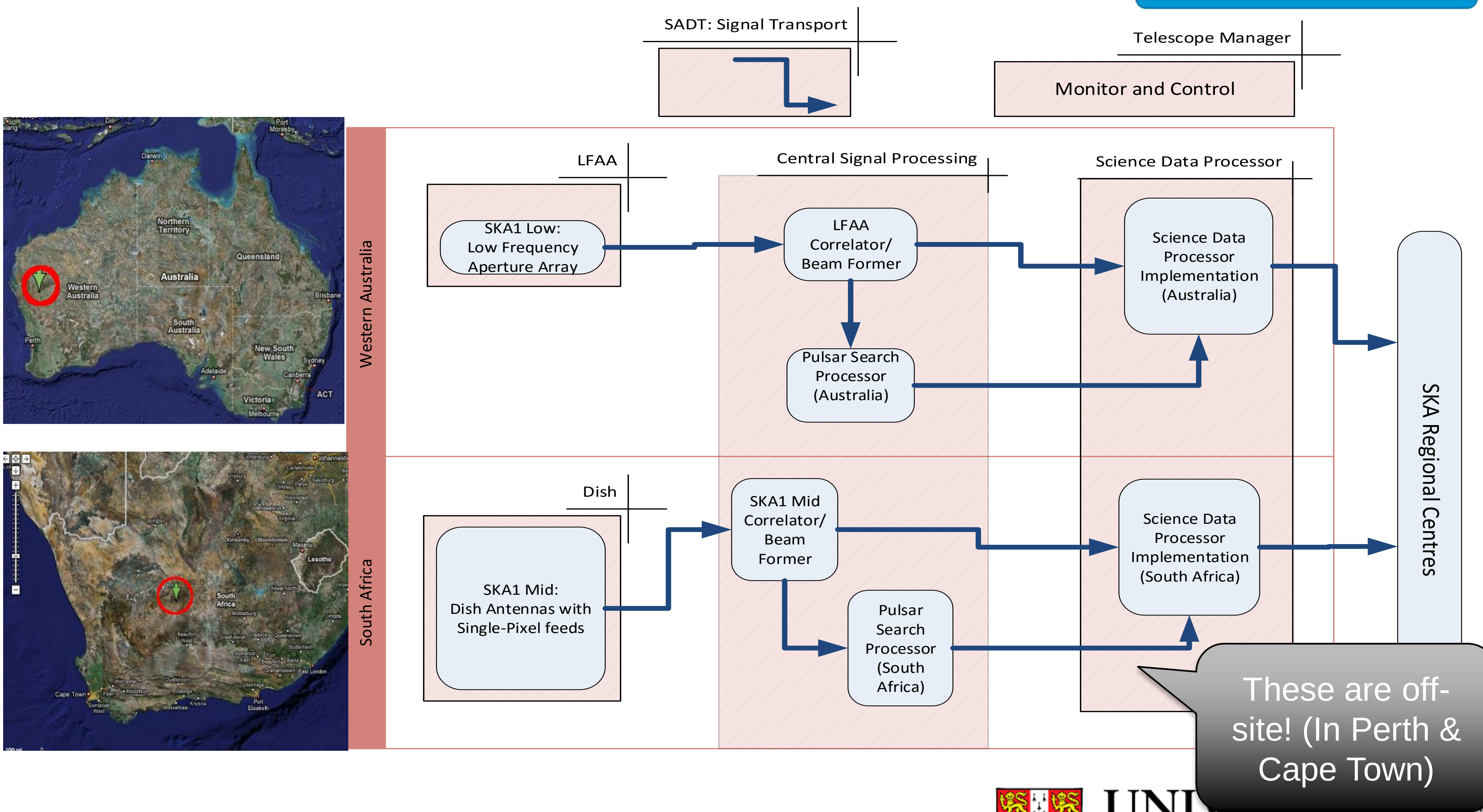
Delivering SKA Science:

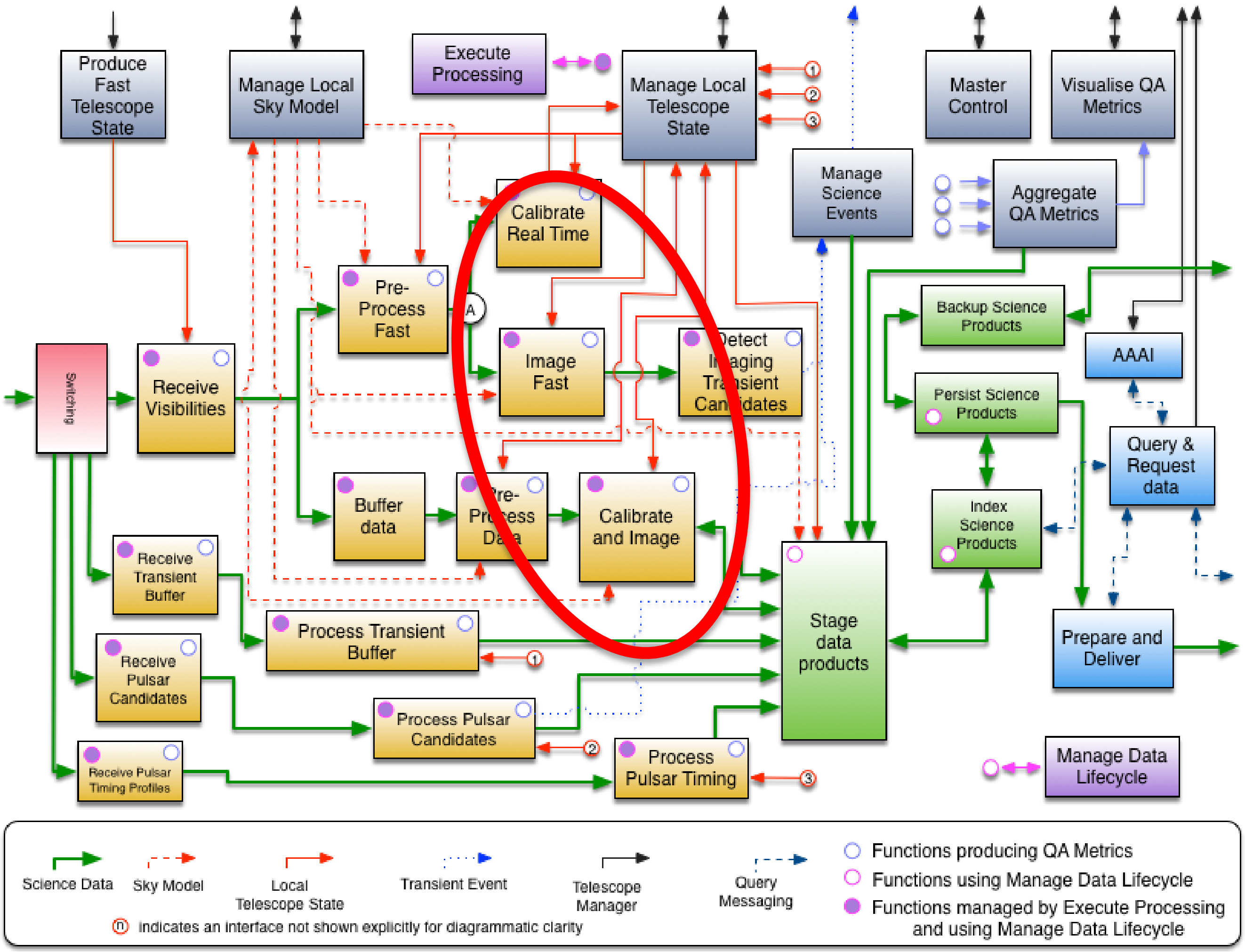
THE SKA IMAGING AND CALIBRATION CHALLENGE

Bojan Nikolic
SKA Science Data Processor Project Engineer
Principal Research Associate
Astrophysics Group, Cavendish Laboratory
University of Cambridge

SKA Context Diagram



Imaging and Calibration Context



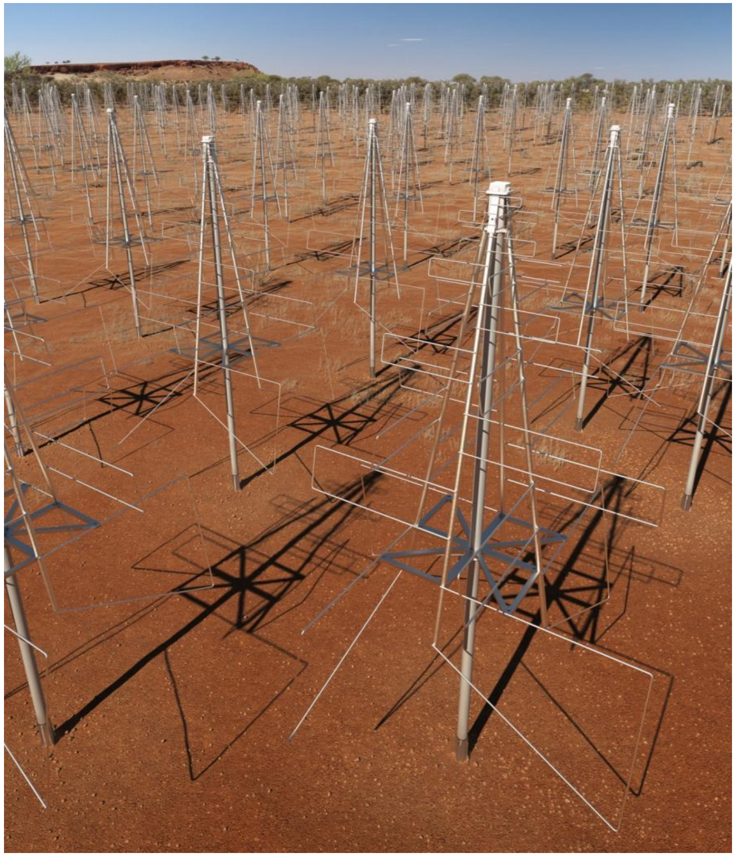
IMAGING AND CALIBRATION PROCESSING IS A MAJOR PART OF THE SKA BY DESIGN

Large “D” – vs – Large “N”

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

GBT 100-m diameter telescope

SKA LFAA prototype array



No 1 aim: collect as many photons as possible -> high sensitivity
No 2 aim: collect radiation from different directions -> high survey speed
No 3 aim: maximum separation of collectors -> high angular resolution

UNIVERSITY OF CAMBRIDGE

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

Very high data rate in

- **Unfeasible to permanently store**
- **Unfeasible to move off-continent**
- **Expensive to store even temporarily**

<u>Optimal</u> processing strategy, algorithms and parameters <u>unknown</u>:

- **<u>Will not</u> be known until the telescope begins operations**
- **Will depend in part on science goals and demands of individual projects**

High computational requirements to process

- **Capital and operational expense**
- **Hardware/software failures rare for individual computers become frequent**

UNIVERSITY OF CAMBRIDGE

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

Very high data rate in

- **Unfeasible to permanently store**
- **Unfeasible to move off-continent**
- **Expensive to store even temporarily**

Optimal processing strategy, algorithms and parameters unknown:

- Will not be known until the telescope begins operations
- Will depend in part on science goals and demands of individual projects

High computational requirements to process

- Capital and operational expense
- Hardware/software failures rare for individual computers become frequent

UNIVERSITY OF CAMBRIDGE

High data rate for imaging

Direct consequence of:

- Large field of view, fast survey speed (small D)
- High angular resolution (long B)
- High continuum sensitivity (large bandwidth)
- Good sampling (large N)
- Mechanical engineering constraints (SKA1-mid)

-> 0.5 TB/s for each of the telescopes

100000x ALMA sustained data rate

10000x ALMA maximum data rate

1000x JVLA maximum data rate

SDP Design Phase approach

- Receive, temporarily store incoming data
- Fairly demanding network but in principle can be done today
- Key challenge is:
  - Where to put the data, how to organise it
  - How to process the data

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

Very high data rate in

- Unfeasible to permanently store
- Unfeasible to move off-continent
- Expensive to store even temporarily

Optimal processing strategy, algorithms and parameters unknown:

- **Will not be known until the telescope begins operations**
- **Will depend in part on science goals and demands of individual projects**

High computational requirements to process

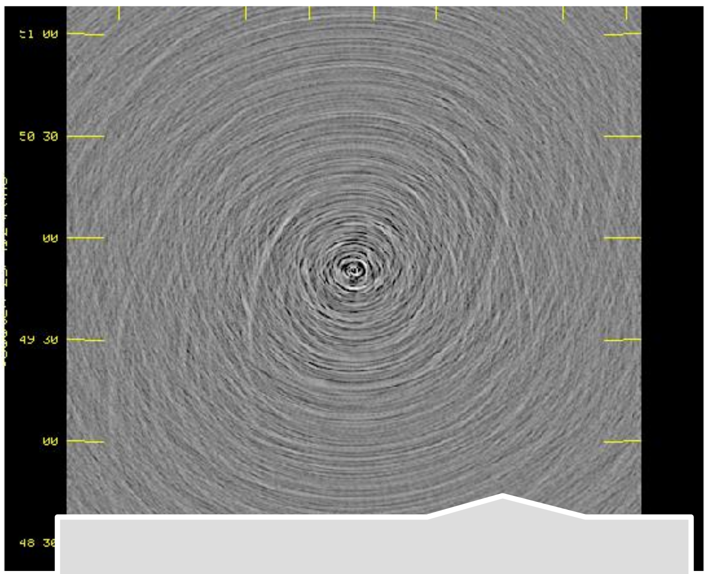
- Capital and operational expense
- Hardware/software failures rare for individual computers become frequent

UNIVERSITY OF CAMBRIDGE

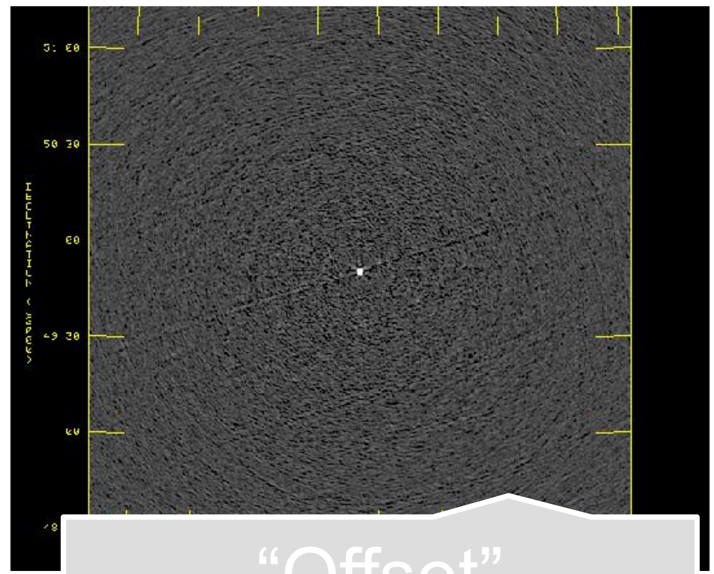
Imaging and calibration algorithmic requirements

- Time- and frequency-variable corruption of the incoming signal:
  - Atmospheric, mechanical & electronic causes
  - Requires iteratively solving for Sky and the corrupting effects – “Self Calibration”
- Irregular, non-uniform sampling of measurements
  - Requires (typically iterative) de-convolution – CLEAN, Wavelets, compressed sensing, etc
- Non-planar distribution of measurements
  - Approximate correction to the plane required if want to use 2D FFTs

Measurements are imperfect – corrupted by slowly changing mechanical, electrical & atmospheric effects



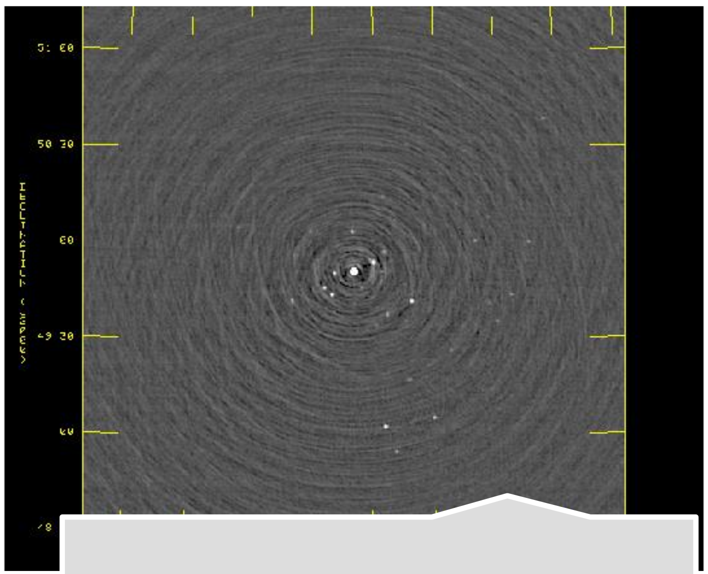
Uncalibrated



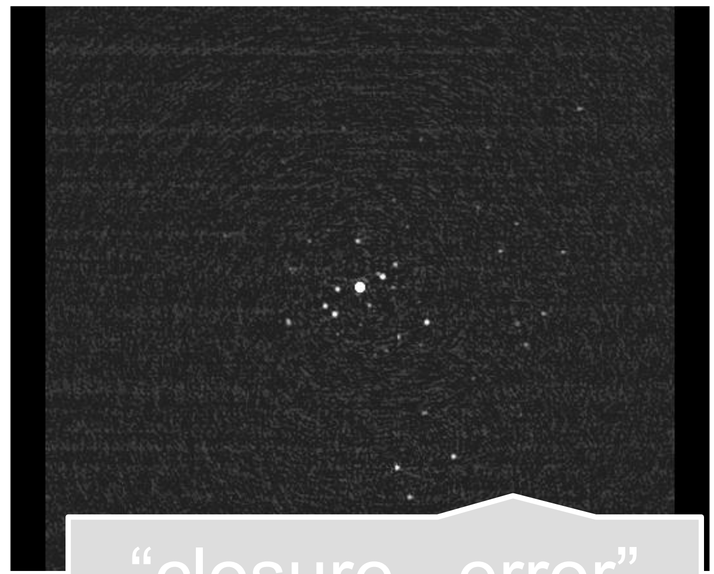
"Offset" Calibration

Rick Perley & Oleg Smirnov: "High Dynamic Range Imaging", www.astron.nl/gerfeest/presentations/perley.pdf

<u>Iterative & joint</u> solving for the image of the Sky & Calibration



“Self-Calibration”



“closure –error” calibration

Rick Perley & Oleg Smirnov: “High Dynamic Range Imaging”, www.astron.nl/gerfeest/presentations/perley.pdf

SKA/SDP Design:

- To support current best-practice algorithms:
  - Multi-frequency multi-scale CLEAN
  - Self-calibration
  - Direction dependent correction using “A” terms
- Flexibility to update and improve in future

Important role for ongoing current research and future optimisation and commissioning

SKA/SDP Approach

SCIENCE DATA PROCESSOR

SKA/SDP Design:

- To support curre
  - Multi-frequency
  - Self-calibration
  - Direction dependent correction using “A” terms
- Flexibility to update and

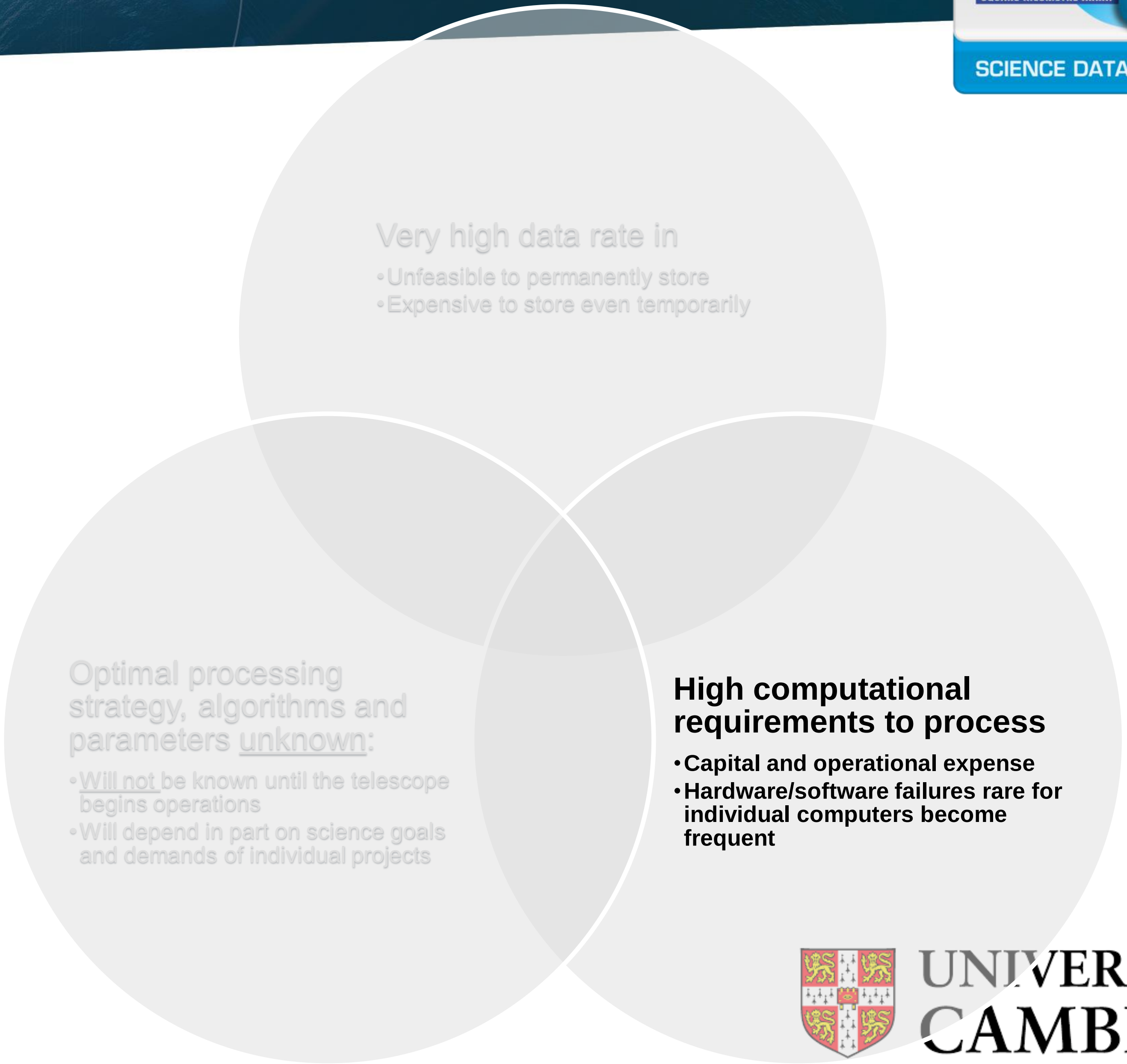
Challenge: Can these algorithms be expressed scalably?
Need >1000x improvement from current proven scales

Challenge: too much flexibility – nothing ever works

Important role for ongoing current research and future optimisation and commissioning

UNIVERSITY OF CAMBRIDGE

Factors driving the SKA challenge



SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

UNIVERSITY OF CAMBRIDGE

Illustrative Computing Requirements

- ~100 PetaFLOPS total achieved
- ~200 PetaByte/s aggregate BW to fast working memory
- ~50 PetaByte Storage
- ~1 TeraByte/s sustained write to storage
- ~10 TeraByte/s sustained read from storage
  - ~~ 10000 FLOPS/byte read from storage

UNIVERSITY OF CAMBRIDGE

Illustrative Computing Requirements

- ~100 PetaFLOPS total achieve[d]
- ~200 PetaByte/s aggregate BW working memory
- ~50 PetaByte Storage
- ~1 TeraByte/s sustained write t[o storage]
- ~10 TeraByte/s sustained read from storage
  - ~~ 10000 FLOPS/byte read from storage

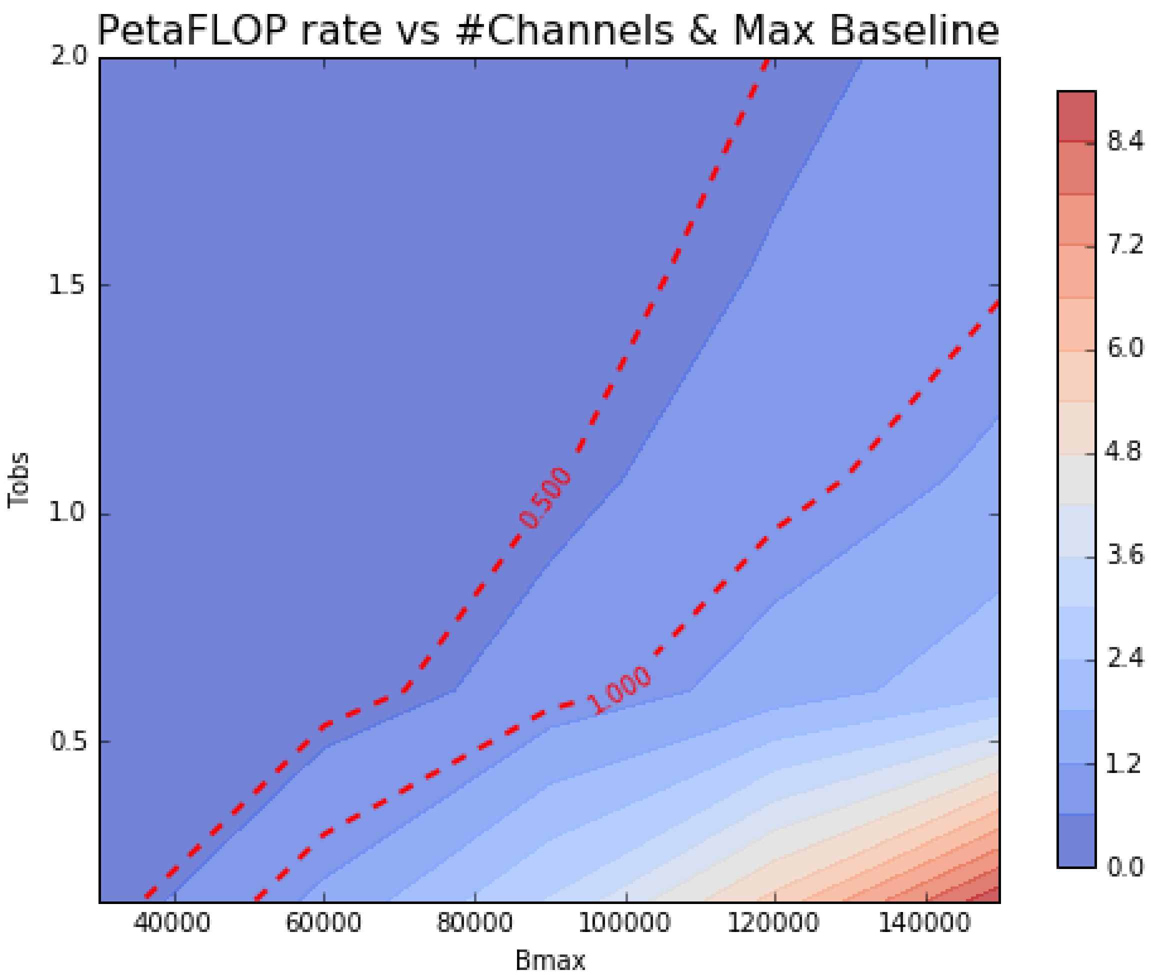
Likely to be achievable ~ 2020

One of the big challenges

Also likely to be achievable well ahead of SDP roll out

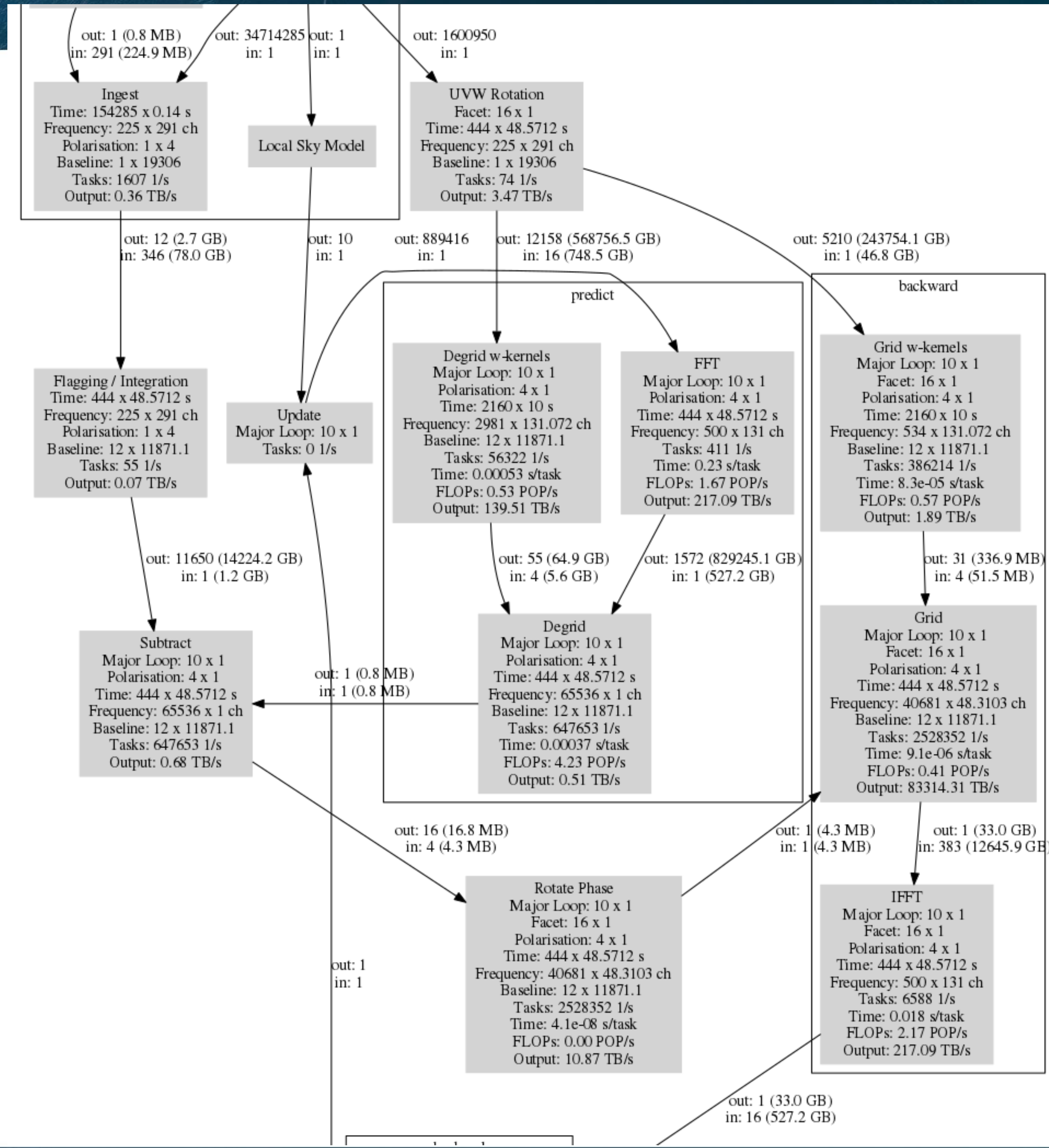
One of the big challenges

Parametric Model Example



Computational cost of a transient survey as a function of integration of each pointing and maximum baseline length that is used

Computational requirements breakdown



SDP Design-Phase Approach

- Document the computational requirements, their relationship to the SDP requirements
- Document the roadmap for likely evolution of computing systems
- Ensure the SDP software architecture can make reasonably efficient use of likely future computing system
- Ensure the maintenance of software is tractable, especially across changes in future computing system architectures
- Prototyping to provide evidential support to the above, demonstrate appropriate technical readiness of potential solutions

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP
SCIENCE DATA PROCESSOR

Critical learning period during commissioning and early operations

Very high data rate in

- Unfeasible to permanently store
- Expensive to store even temporarily

High Degree of Parallelism, automatic unsupervised pipelines

SDP

Optimal processing strategy, algorithms and parameters <u>unknown</u>:

- Will not be known until the telescope begins operations
- Will depend in part on science goals and demands of individual projects

High computational requirements to process

- Capital and operational expense
- Hardware/software failures rare for individual computers become frequent

Good Models, Simulations essential, Early science planning

UNIVERSITY OF CAMBRIDGE

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

Critical learning period during commissioning and early operations

Very high data rate in

- Unfeasible to permanently store
- Expensive to store even temporarily

High Degree of Parallelism, automatic unsupervised pipelines

SDP

Optimal processing strategy, algorithms and parameters <u>unknown</u>:

- Will not be known until the telescope begins operations
- Will depend in part on science goals and demands of individual projects

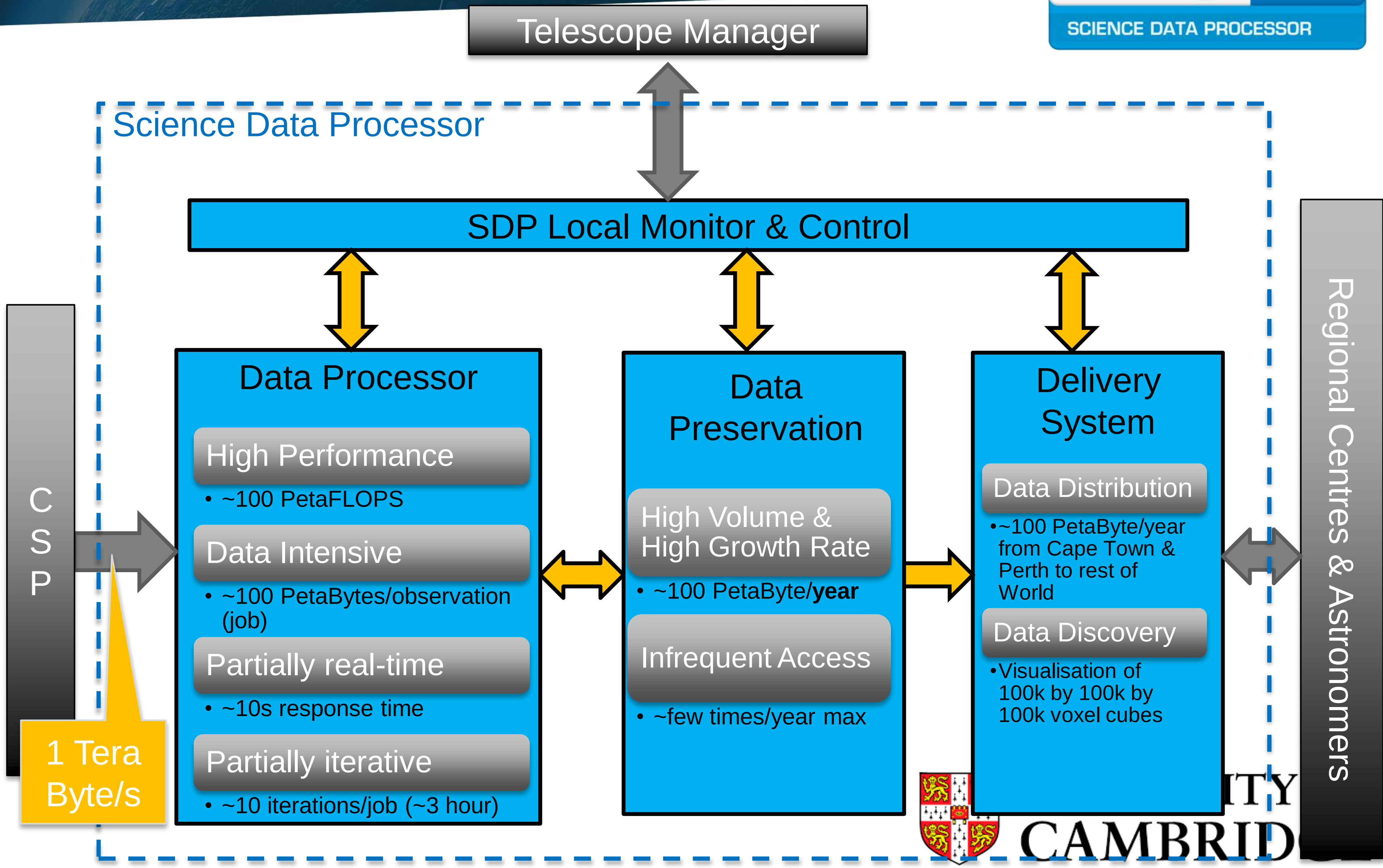
High computational requirements to process

- Capital and operational expense
- Hardware/software failures rare for individual computers become frequent

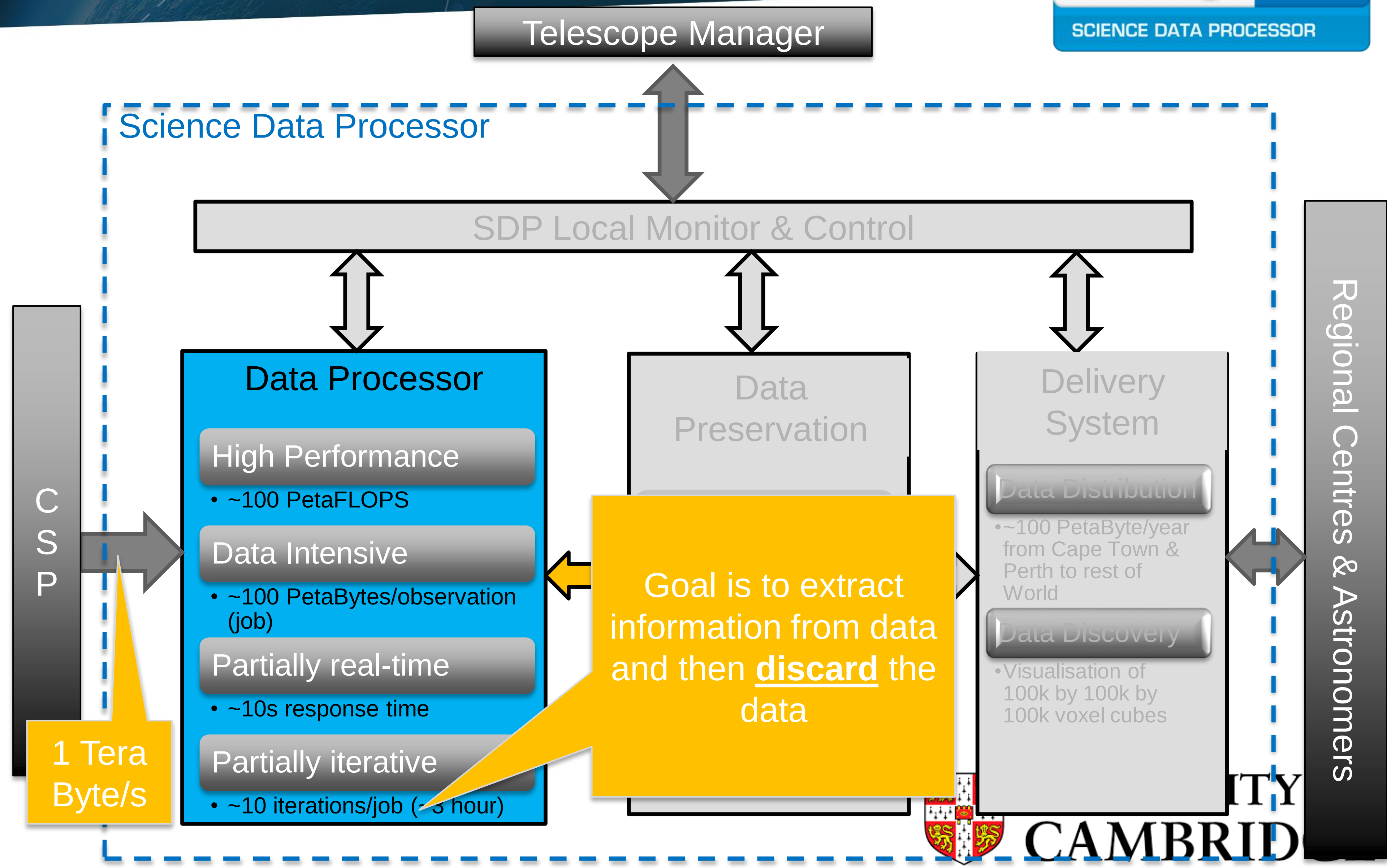
Good Models, Simulations essential, Early science planning

UNIVERSITY OF CAMBRIDGE

SDP Top-level Components & Key Performance Requirements -- **SKA Phase 1**



SDP Top-level Components & Key Performance Requirements -- **SKA Phase 1**



Programming model

- Hybrid programming model:
  - Dataflow at **coarse-grained** level:
    - About 1 million tasks/s max over the whole processor (-> ~10s – 100s milli second tasks), consuming ~100 MegaByte each
    - **Static scheduling** at coarsest-level (down to “data-island”)
      - Static partitioning of the large-volume input data
    - **Dynamic scheduling** within data island:
      - Failure recovery, dynamic load-balancing
    - *Data driven* (all data will be used)
  - Shared memory model at **fine-grained** level e.g.: threads/OpenMP/SIMT-like
    - ~100s active threads per shared memory space
    - Allows manageable working memory size, computational efficiency

Challenge: Unsupervised pipelines and processing

- Extremely challenging to deliver early in operations
- Very challenging to deliver for a diverse set of science programmes and goals
- Unsatisfactory performance will lead to low observatory efficiencies

Factors driving the SKA challenge

SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

Critical learning period during commissioning and early operations

Very high data rate in

- Unfeasible to permanently store
- Expensive to store even temporarily

High Degree of Parallelism, automat unsupervised pipelin

SDP Design Phase

SDP

Optimal processing strategy, algorithms and parameters unknown:

- Will not be known until the telescope begins operations
- Will depend in part on science goals and demands of individual projects

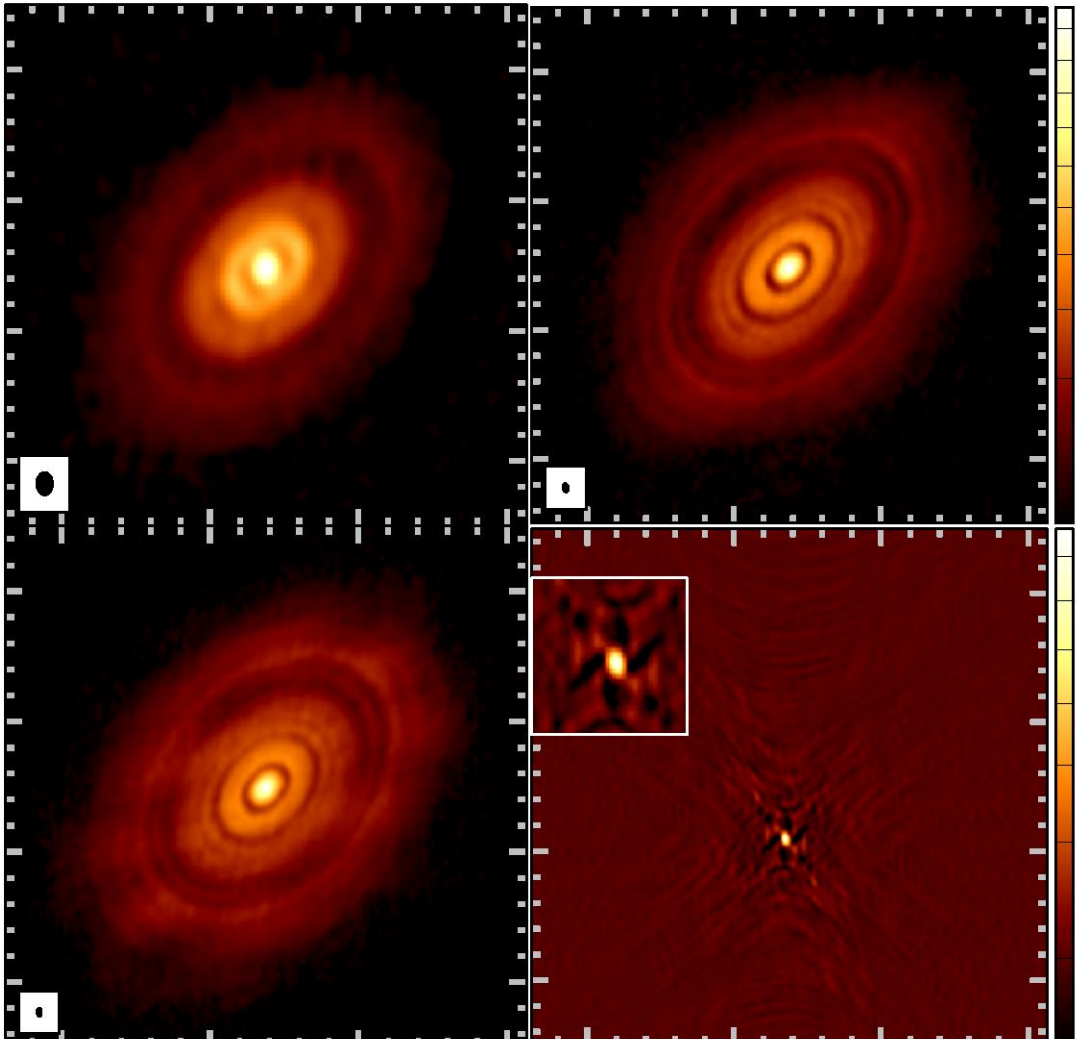
High computational requirements to process

- Capital and operational expense
- Hardware/software failures rare for individual computers become frequent

Good Models, Simulations essential Early Science Planning

UNIVERSITY OF CAMBRIDGE

Long term commissioning and optimisation
Example from ALMA



First Results from High Angular Resolution ALMA Observations Toward the HL Tau Region,
ALMA Partnership,
2015ApJ...808L...3A
Result of collaboration of observatory staff, institutes and universities to characterise and commission ALMA long baselines

END