SKA Science Data Processor

Current status and preliminary design

Chris Broekema (Compute platform lead)

Architectural Principles

SKA SQUARE KILOMETRE ARRAY
SDP
SCIENCE DATA PROCESSOR

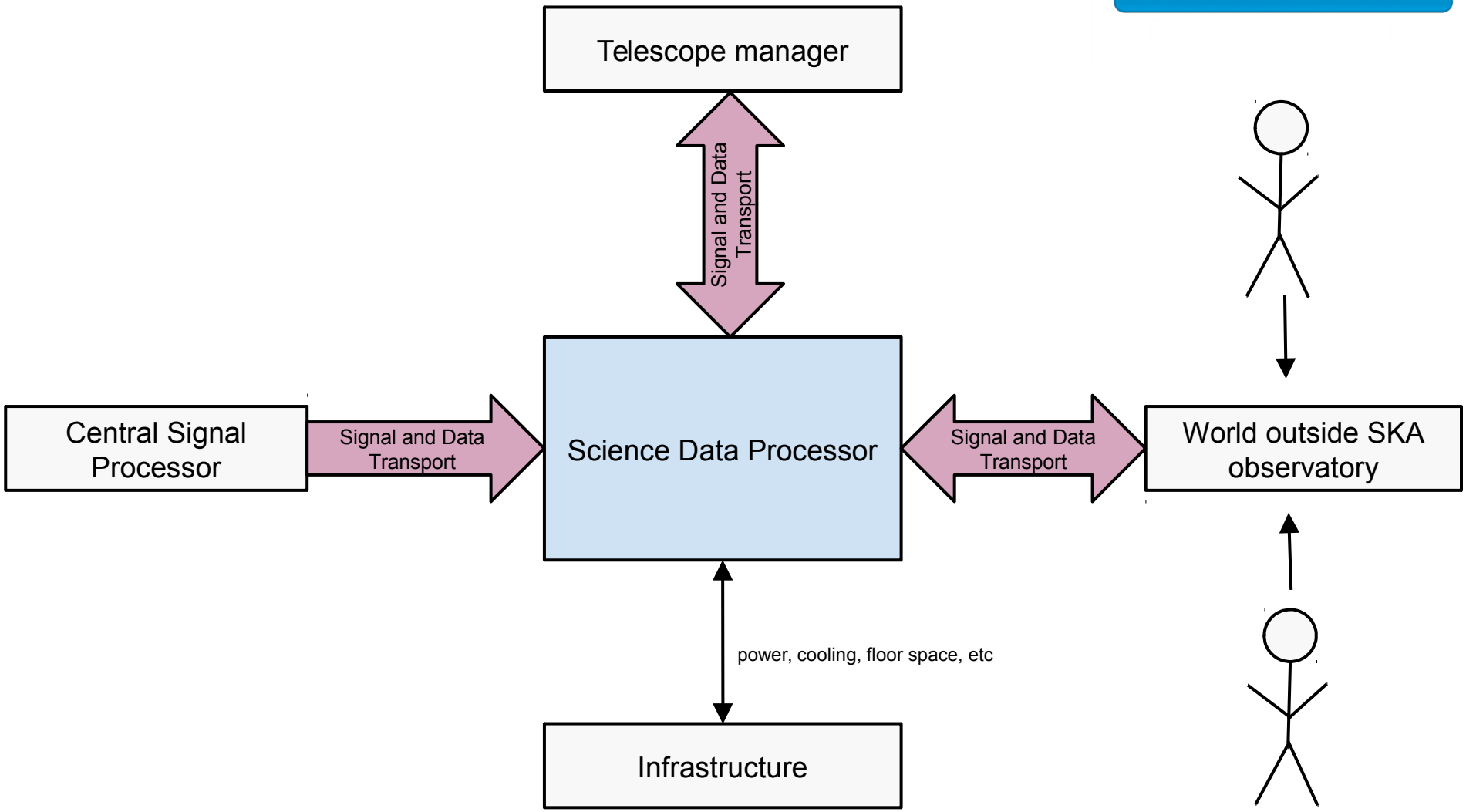
- Main principles:
  - **Ensure scalability**
  - **Ensure affordability**
  - **Ensure maintainability**
  - **Support current state-of-the-art algorithms**

- Exploit data parallelism, not just in frequency but also other dimensions
  - We have only two fundamental/bulk data structures
  - Raster grids and key-value-value stream records [e.g. u,v,w, -> visibility]

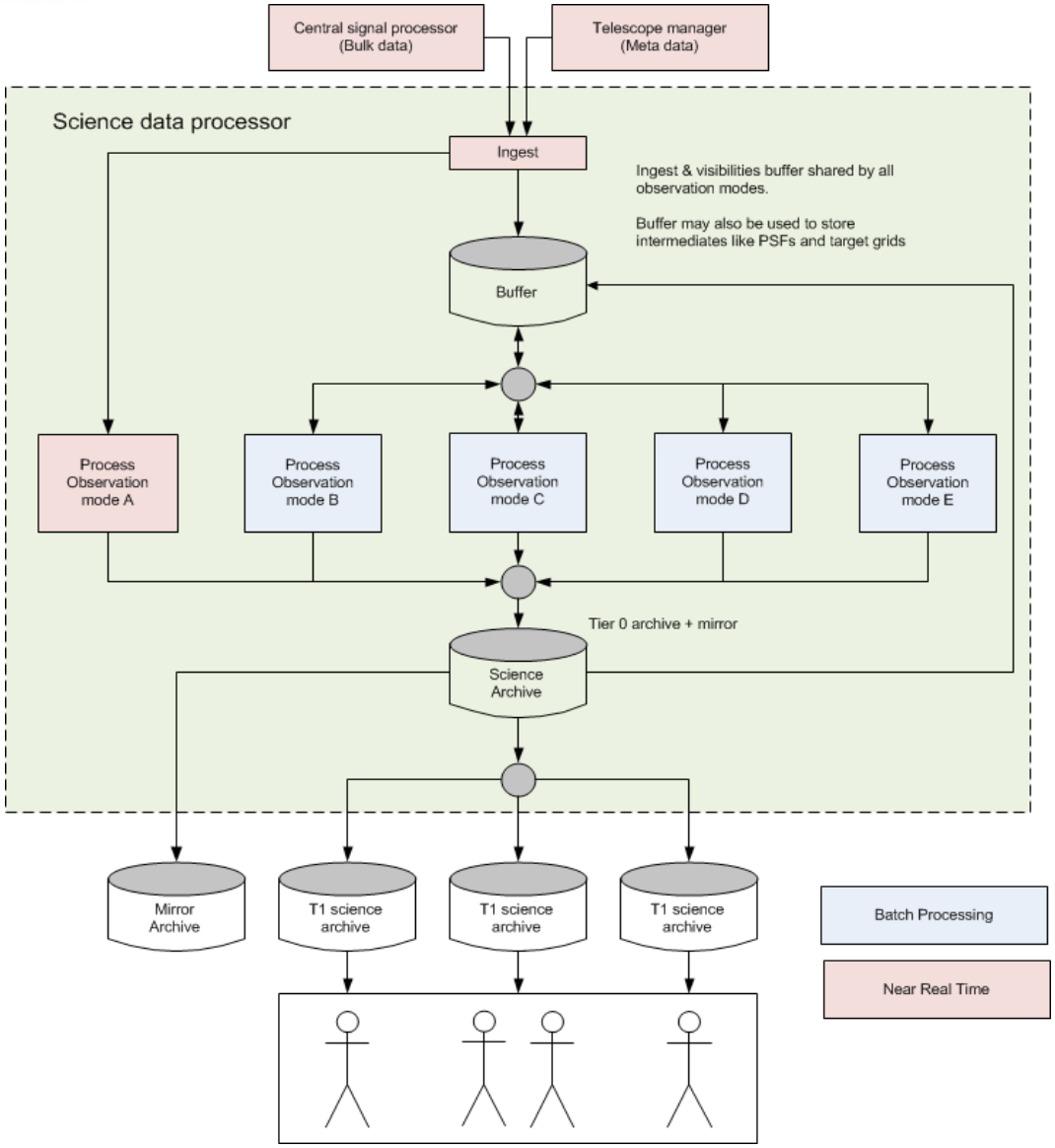
- Emphasis is on the framework to manage the throughput
  - Hardware platform will be replaced on a short duty cycle c.f. any HPC facility
  - Algorithms and workflow will evolve as we learn about telescopes

Approach: Co-design of software and physical layer architectures

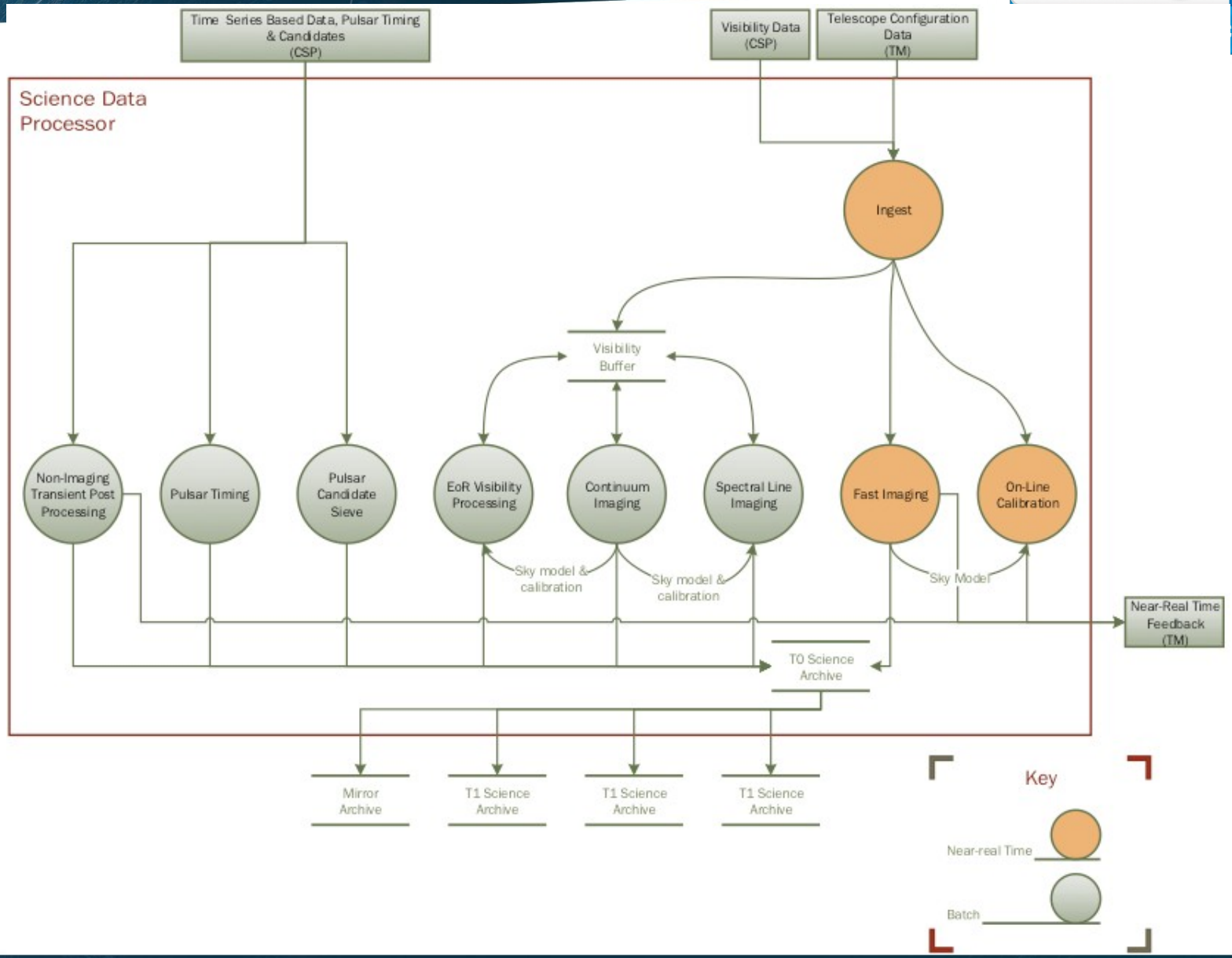
SDP context diagram with external interfaces



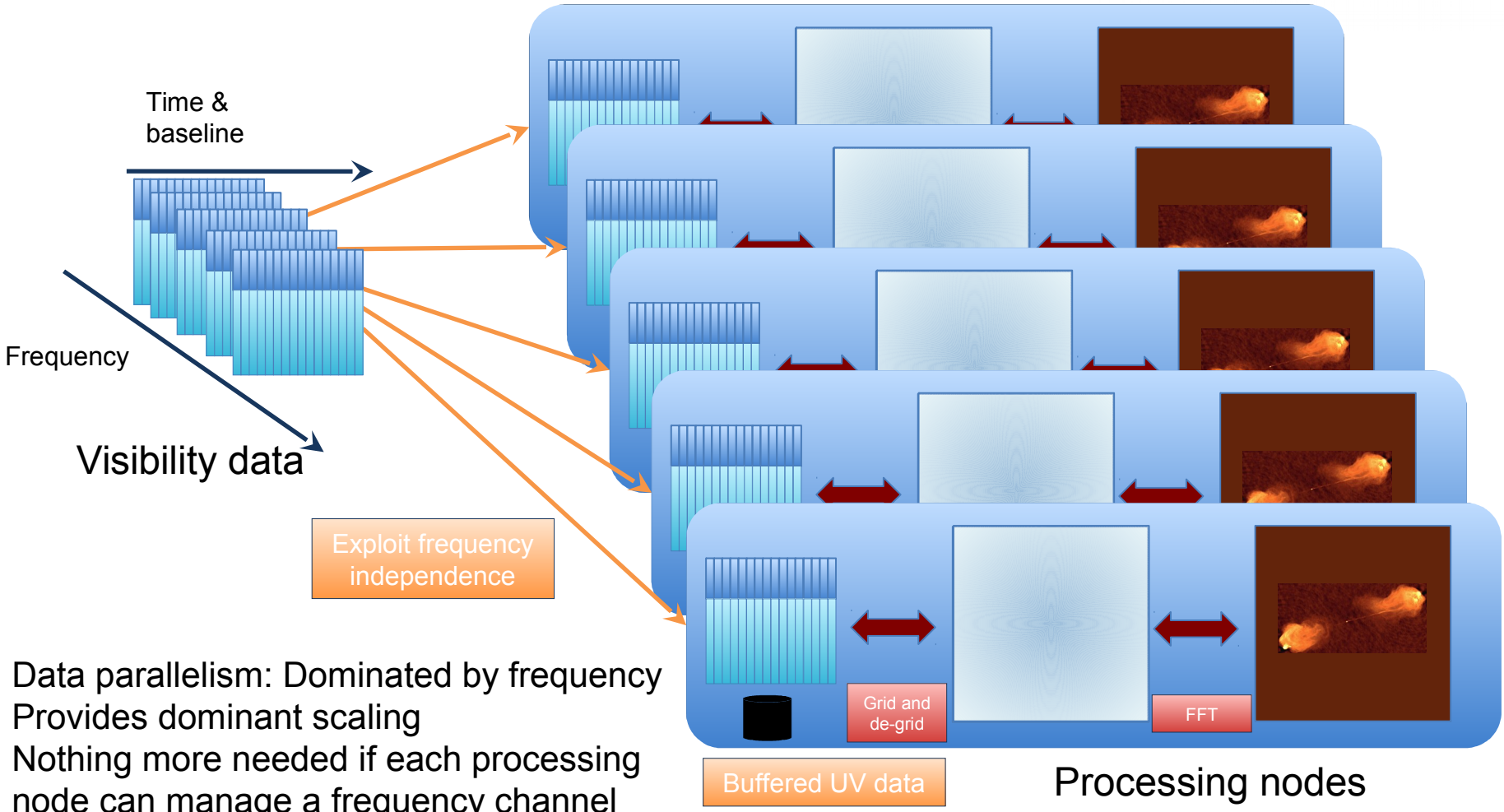
Overview of Science Data Processor



Overview of Science Data Processor

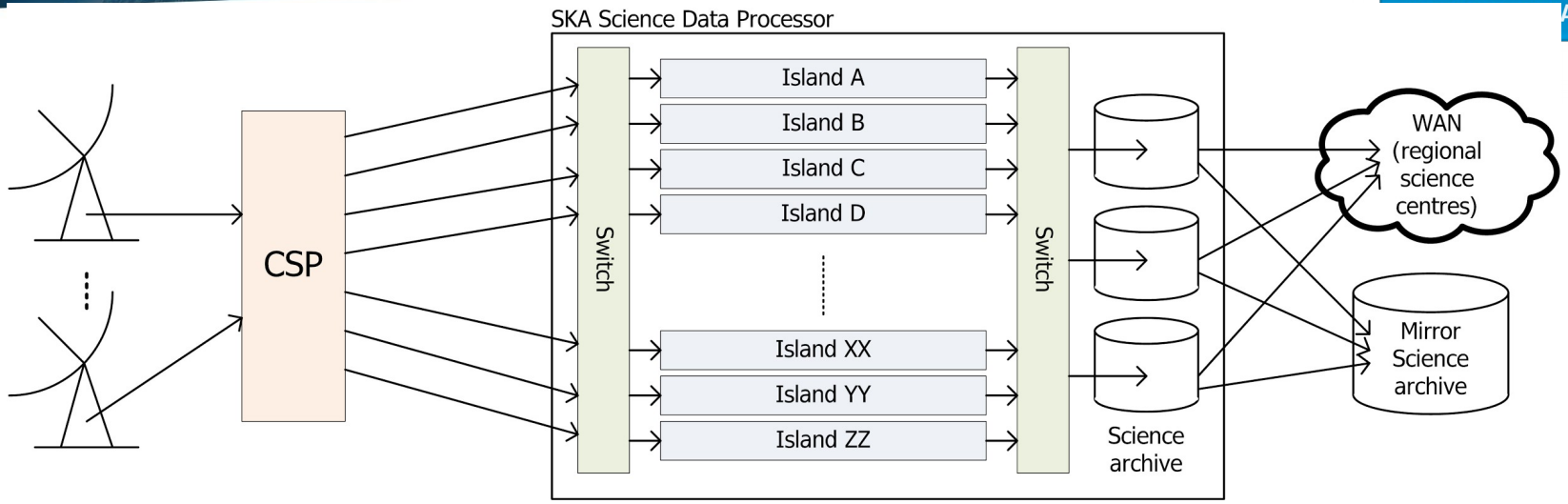


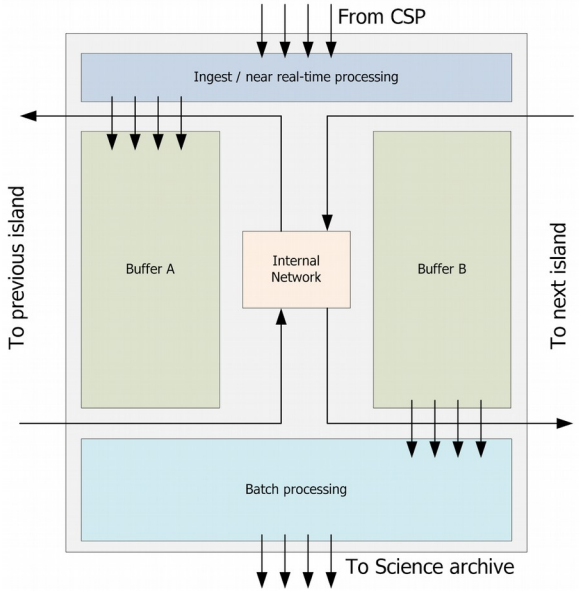
Data Driven Architecture

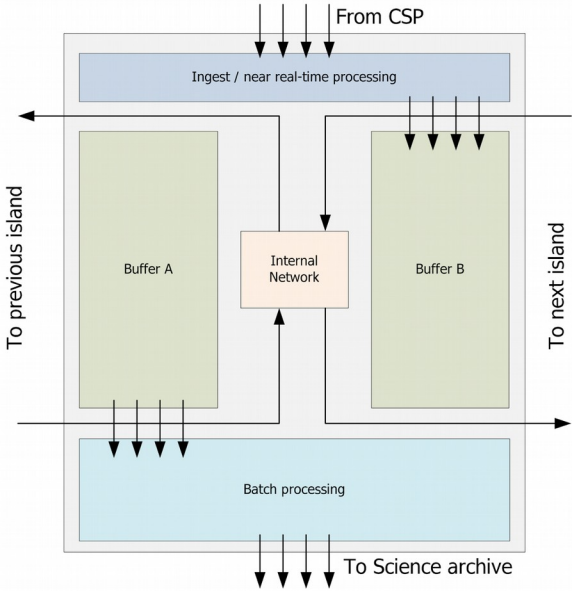


- Data parallelism: Dominated by frequency
- Provides dominant scaling
- Nothing more needed if each processing node can manage a frequency channel complete processing

Physical Layer





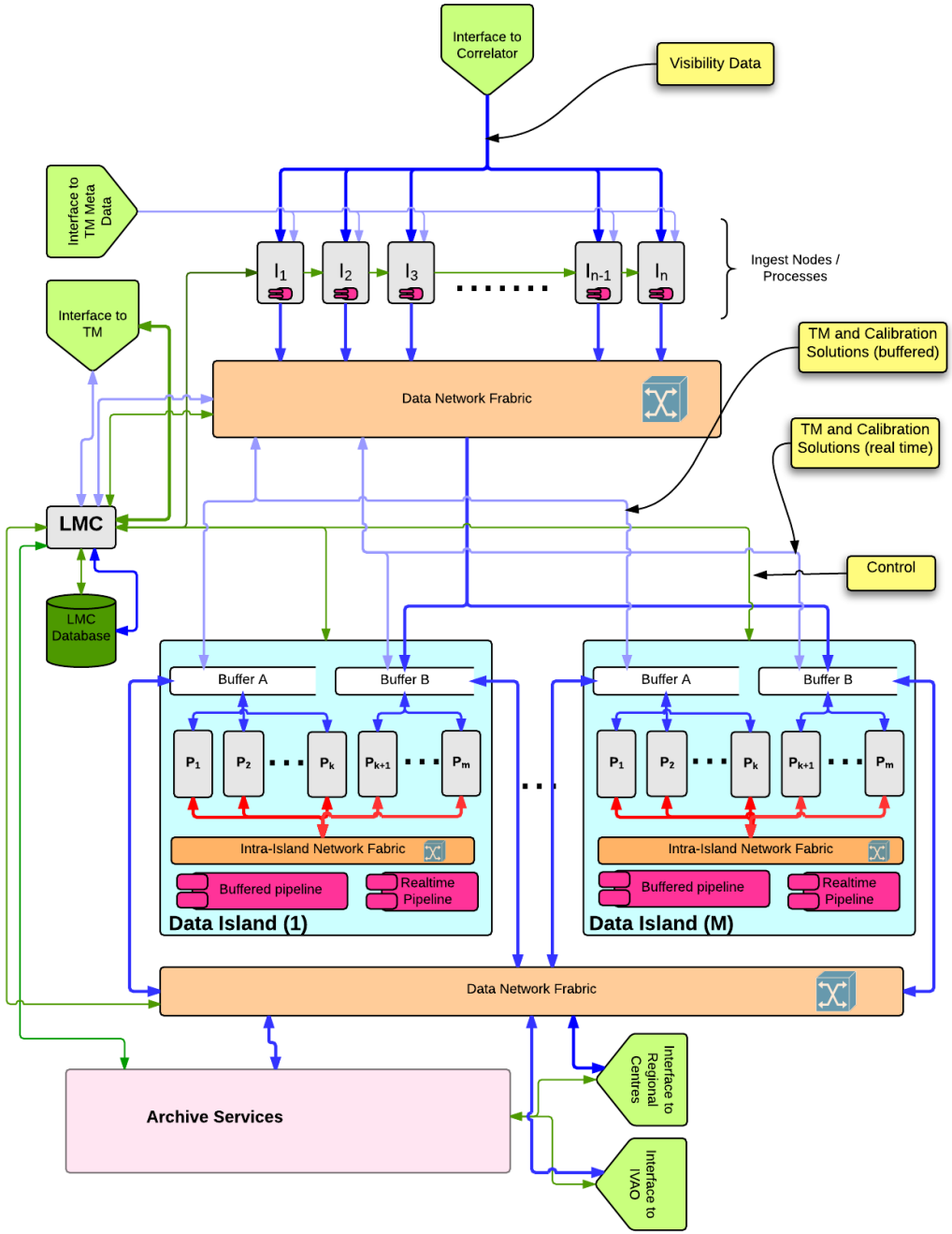


Physical Lay

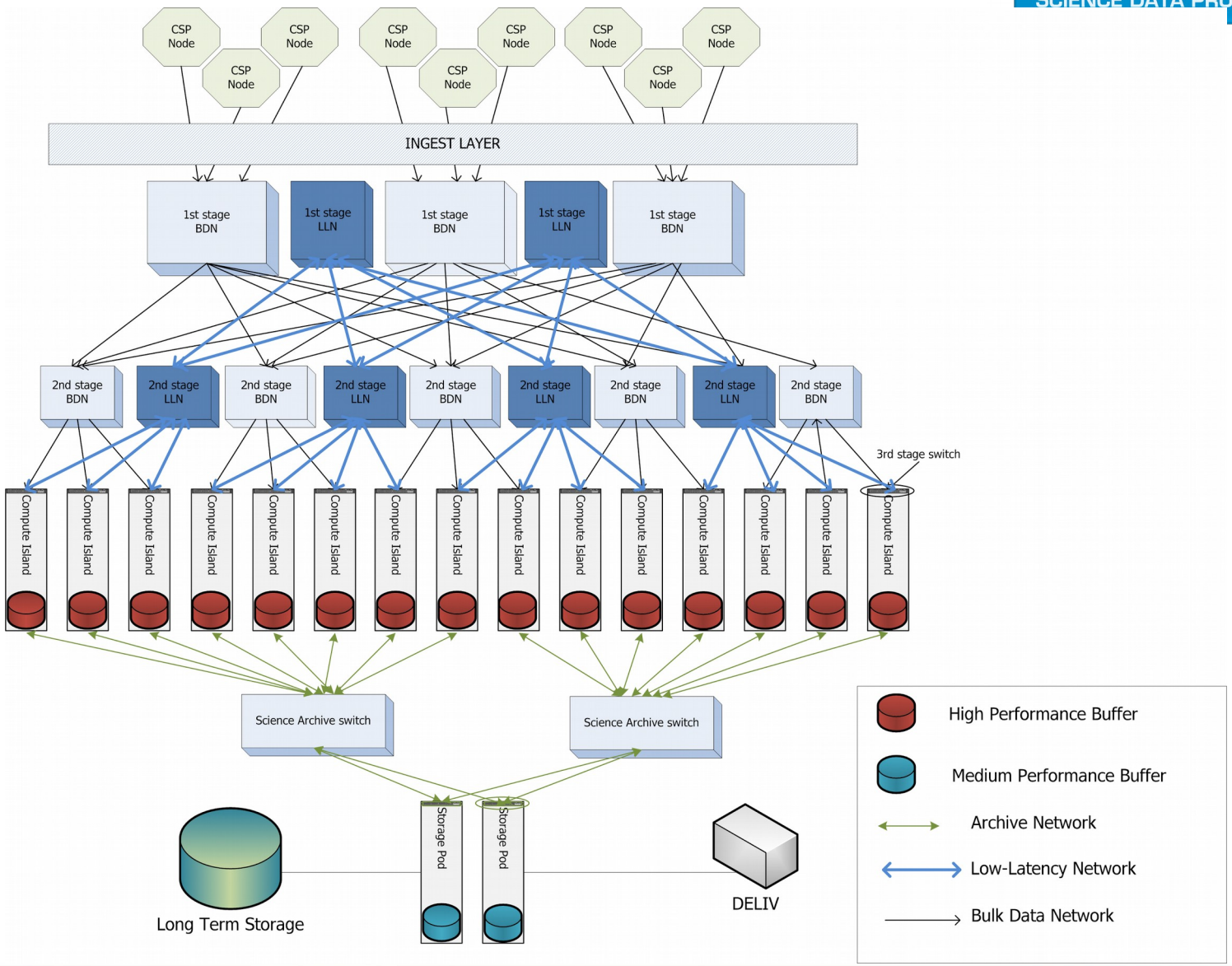
SKA SQUARE KILOMETRE ARRAY

SDP

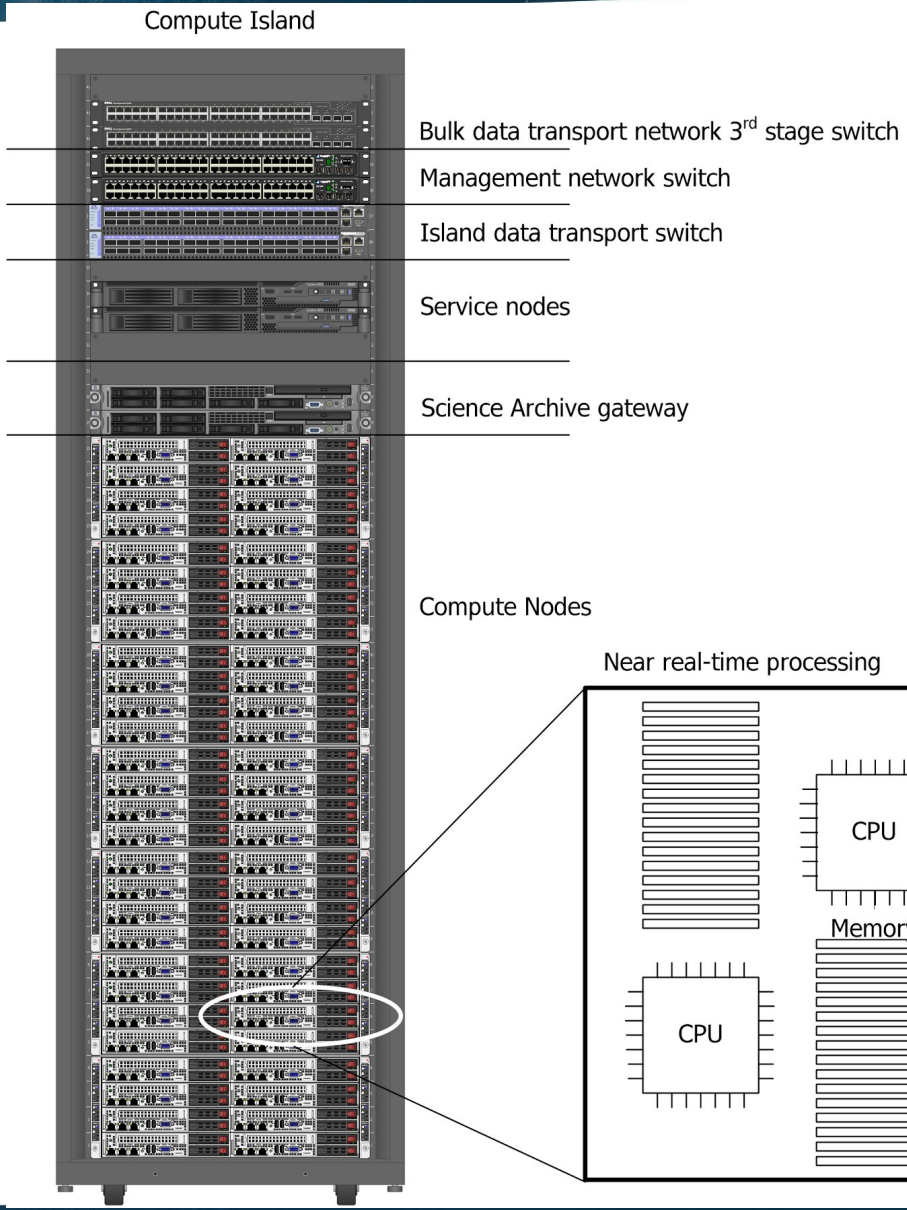
SCIENCE DATA PROCESSOR



Top level SDP network

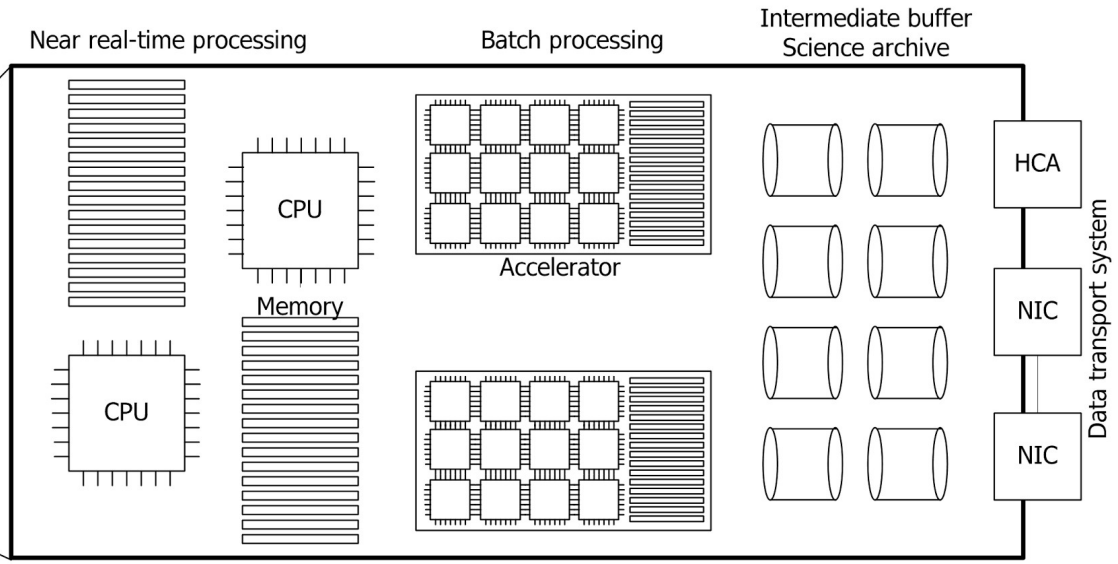


The Compute Island concept

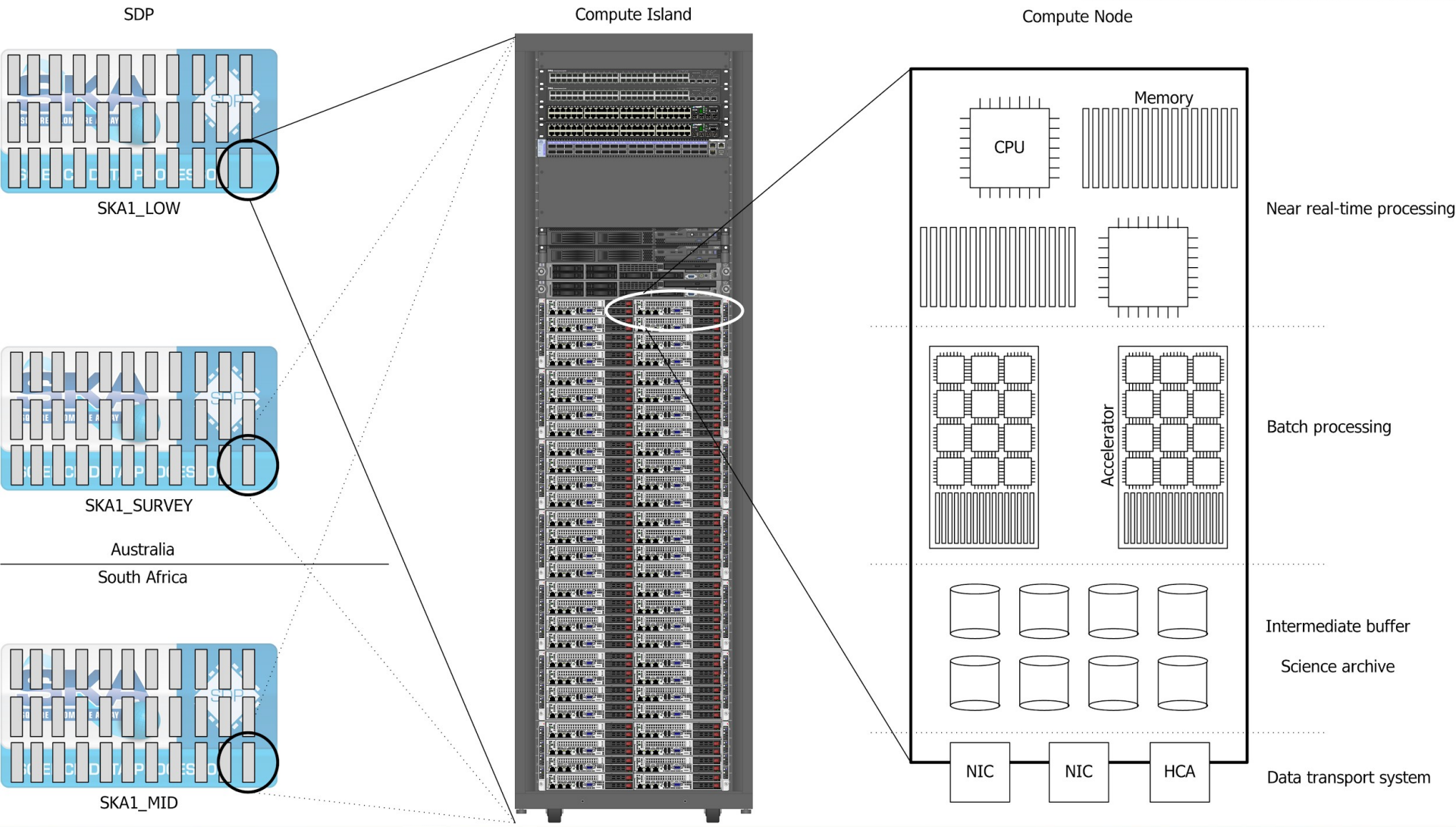


A Compute Island is a self-contained set of processing resources, capable of end-to-end processing of a chunk of UVW-space, including

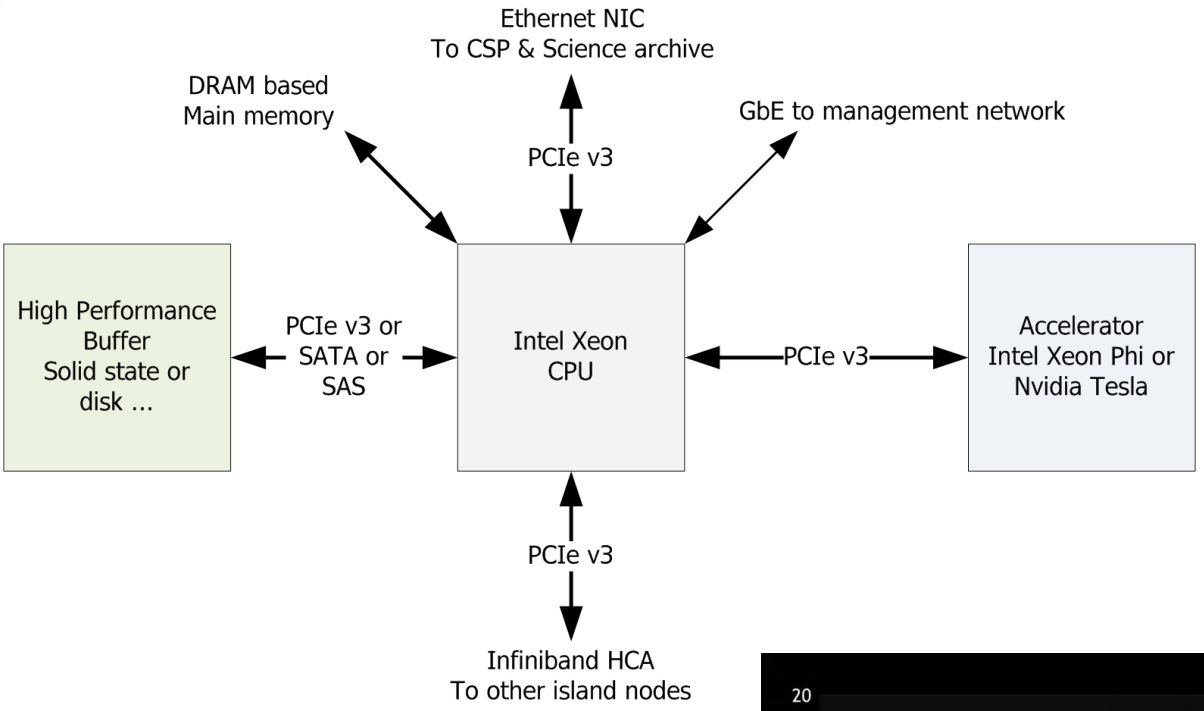
- Ingest
- Buffer
- Calibration & Imaging
- Archive



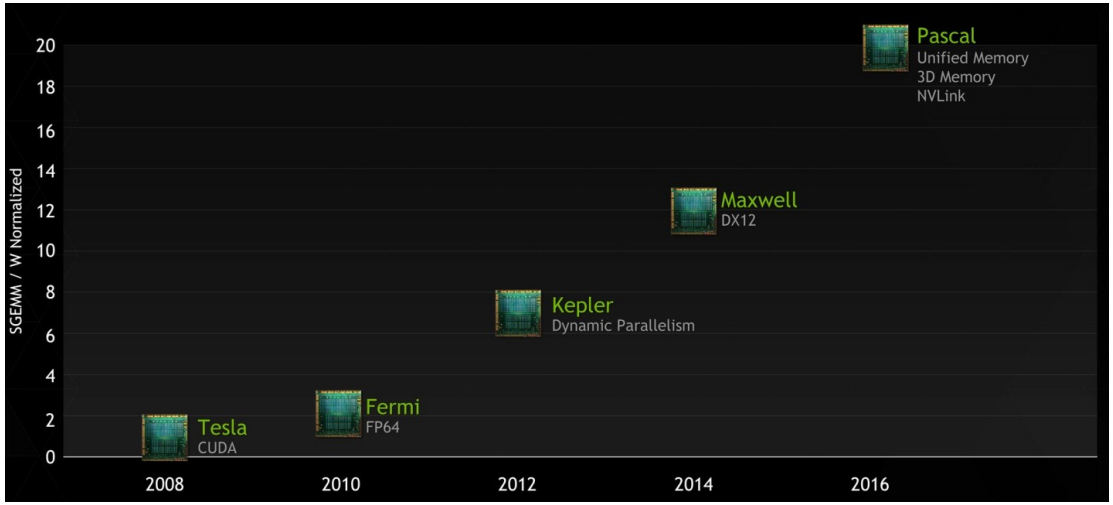
SDP scaling using Compute Islands



Compute Node Candidate Architecture



- Heavily inspired by COBALT node
- Used for prelim costing & scaling
- Many alternatives possible
  - ARM
  - OpenPower + CAPI
  - FPGA
  - DSP
  - Custom hardware
  - combinations of above
  - etc...

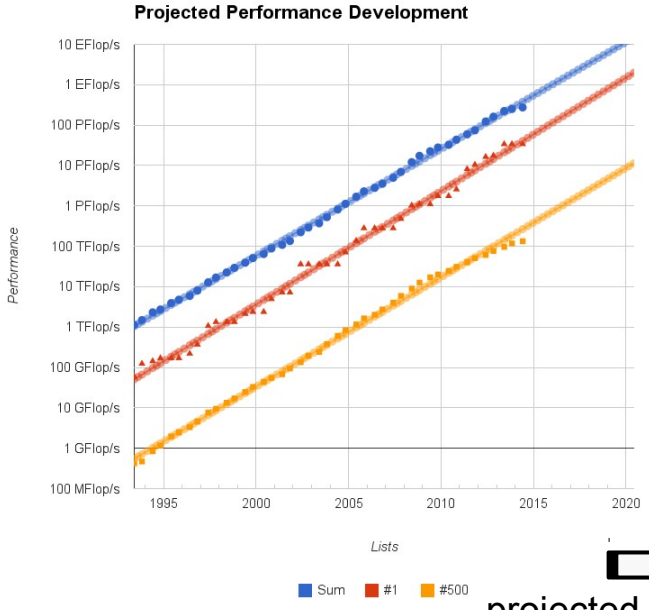


Required SDP capacity

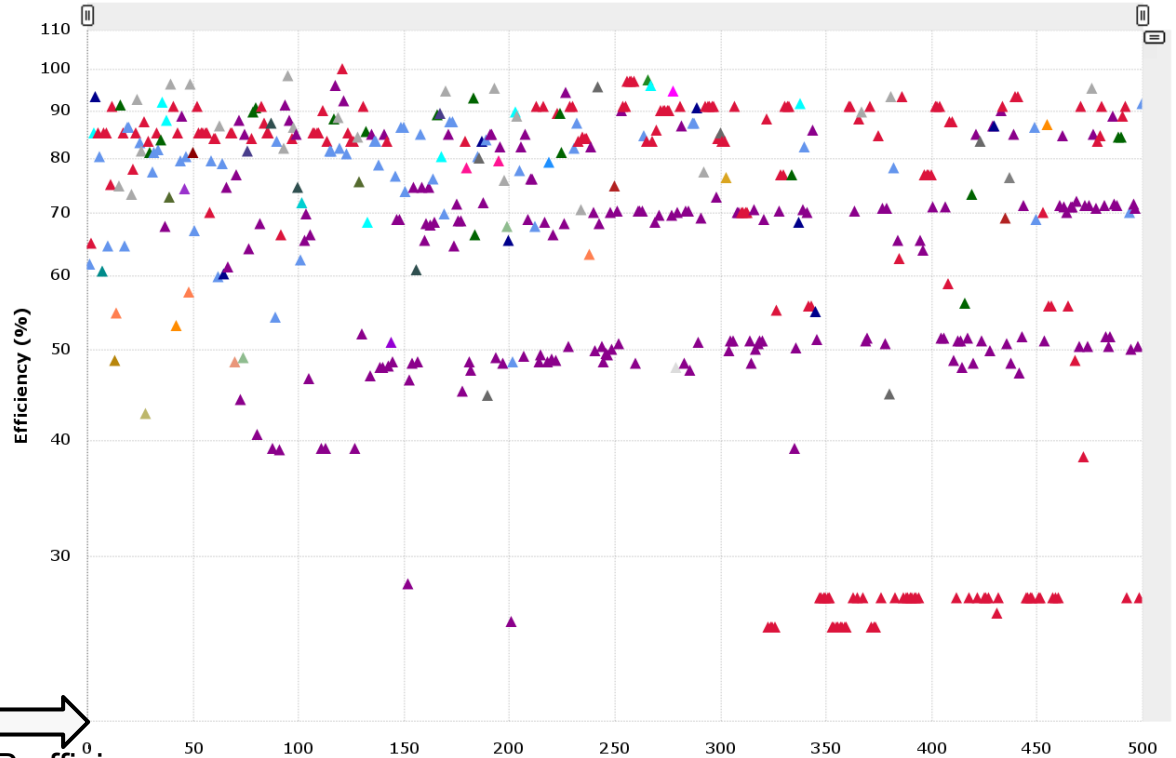
Required SDP computational capacity depends on

1. Input bandwidth
2. Computational intensity of the data reduction
3. Computational efficiency in % of R_{peak}

Top 500 performance development



Top 500 computational efficiency



projected SDP efficiency

Required SDP capacity

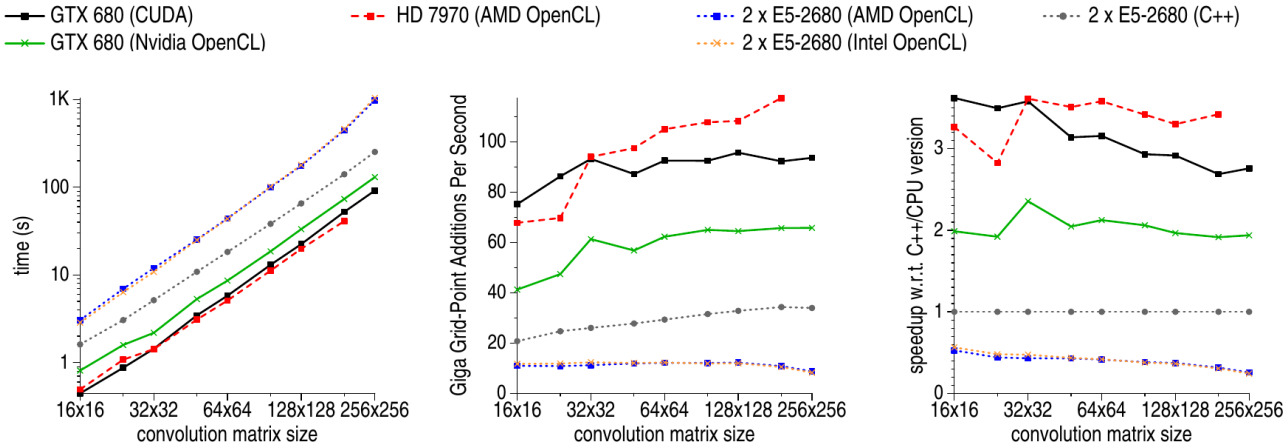
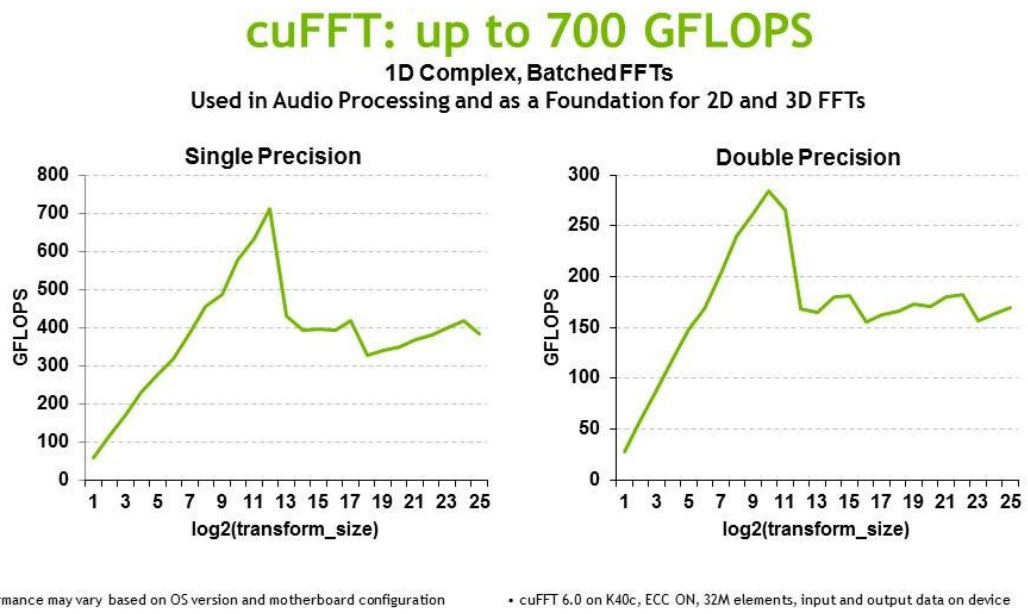


Figure 4: Performance of our new strategy for



SKA

SQUARE KILOMETRE ARRAY

SDP

SCIENCE DATA PROCESSOR