Processor for the Square Kilometre Array Telescope

Bojan Nikolic
Astrophysics Group, Cavendish Lab
University of Cambridge
Email: b.nikolic@mrao.cam.ac.uk

Project Engineer & Architecture Task Leader
SKA Science Data Processor Consortium
A major, integral, part of the next generation low-frequency radio telescope

A major computing project

Has requirements that are very distinct from current large-scale computing applications

In design phase, construction to start in 2017/18 (substantial portion of budget already announced)

Has a data-parallel, dataflow based architecture with an unusual high-availability model
SKA Data Processor is “Sensor” Big Data

• **Big Data** -- *relatively* low **Information**
  – Processing: extracts information, reduces volume of data by a large factor
  – Processing algorithms *not* diverse: the diverse analyses are done on the reduced data
  – Pervasive data parallelism in early stages of processing

• **Imperfect & irregular** measurements
  – Removing the measurement artefacts is the major challenge (but allows for much cheaper sensors)
  – Requires iterative algorithms (~10 iterations) & joins of parallel processing outputs
  – These *imperfections & irregularities* -> general computing *not* traditional streaming signal processing
SKA Data Processor is:

• Floating point based (FFTs, stencils, etc)
• Program “state” much smaller than input data
• One application with several modes – not a huge diversity of algorithms and applications
• No intrinsic fine-grained bulk synchronous time-stepping
• Networking communications dominated by input data distribution and reductions of intermediate results
• Can handle loss of data and intermediate results (similar to Monte-Carlo)
What is the Square Kilometre Array (SKA)?
Scientific Context – a partner to ALMA, EELT, JWST

Credit: A. Marinkovic/XCam/ALMA (ESO/NAOJ/NRAO)

Credit: ESO/L. Calçada (artists impression)

Credit: Northrop Grumman (artists impression)

Credit: SKA Organisation (artists impression)
Scientific Context – a partner to ALMA, EELT, JWST

ALMA:
- 66 high precision sub-mm antennas
- Completed in 2013
- Budget ~1.5 bn USD

Credit: A. Marinkovic/XCam/ALMA(ESO/NAOJ/NRAO)

JWST:
- 6.5m space near-infrared telescope
- Launch 2018
- Budget ~8 bn USD

Credit: Northrop Grumman (artists impression)

European ELT
- ~40m optical telescope
- Completion ~2025
- Budget ~1.1 bn EUR

Credit: ESO/L. Calçada (artists impression)

Square Kilometre Array
- Two next generation low-frequency arrays
- Completion ~2022 for Phase 1
- Budget 0.65 bn EUR for Phase 1 Construction

Credit: SKA Organisation (artists impression)
## What will be the Square Kilometre Array (SKA)?

<table>
<thead>
<tr>
<th><strong>Radio Telescope</strong></th>
<th>Makes Images of the Sky at radio (5m-3cm) wavelengths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~100 more sensitive than current telescopes</td>
</tr>
<tr>
<td></td>
<td>Complements ALMA, JWST (successor to Hubble), and E-ELT</td>
</tr>
<tr>
<td><strong>Currently in Design</strong></td>
<td>Construction begins 2018</td>
</tr>
<tr>
<td></td>
<td>Full operations expected at end 2022</td>
</tr>
<tr>
<td></td>
<td>Significant funds already committed by participating countries</td>
</tr>
<tr>
<td><strong>Major Engineering Project</strong></td>
<td>Two remote desert sites</td>
</tr>
<tr>
<td></td>
<td>&gt;100k receiving elements</td>
</tr>
<tr>
<td><strong>Major ICT Project</strong></td>
<td>Subject of this talk!</td>
</tr>
</tbody>
</table>
The “known-unknowns”

HIGH-PRIORITY SCIENCE OBJECTIVES
Epoch of Re-Ionisation
Epoch of Re-ionization

Furlanetto et al. (2003)

$z = 18.3$

10 Mpc comoving
$\Delta v = 0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=16.1$

10 Mpc comoving
$\Delta v=0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=14.5$

10 Mpc comoving

$\Delta v=0.1$ MHz
Epoch of Re-ionization

*Furlanetto et al. (2003)*

$z=13.2$

$10 \text{ Mpc comoving}$

$\Delta v=0.1 \text{ MHz}$
Epoch of Re-ionization

Furlanetto et al. (2003)

z=12.1

10 Mpc comoving
Δν=0.1 MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=11.2$

10 Mpc comoving

$\Delta v=0.1$ MHz
Epoch of Re-ionization

*Furlanetto et al. (2003)*

$z = 10.4$

10 Mpc comoving

$\Delta v = 0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z = 9.8$

10 Mpc comoving
$\Delta v = 0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=9.2$

$10 \text{ Mpc comoving}$

$\Delta \nu=0.1 \text{ MHz}$
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=8.7$

10 Mpc comoving
$\Delta v=0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=8.3$

10 Mpc comoving
$\Delta v=0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=7.9$

10 Mpc comoving
$\Delta \nu=0.1$ MHz
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=7.5$

$10 \text{ Mpc comoving}$

$\Delta v=0.1 \text{ MHz}$
Epoch of Re-ionization

Furlanetto et al. (2003)

$z=7.2$

10 Mpc comoving
$\Delta v=0.1$ MHz
The SKA will detect around 30,000 pulsars in our own galaxy, 2000 msec pulsars → accurate clocks

Relativistic binaries give unprecedented strong-field test of gravity expect ~100

Timing net of ms pulsars to detect gravitational waves via timing residuals

Expect timing accuracy to improve by ~100
Nano-Hertz range of frequencies

- MBH-MBH binaries: resolved objects and stochastic background
- Cosmic strings and other exotic phenomenon
- Timing residual ~10s ns $\to$ need ms pulsars
Finding the unexpected

Hubble Deep Field (HDF)

~ 3000 galaxies

Very Large Array observation of HDF

~ 15 radio sources
Finding the unexpected (2)

Hubble Deep Field (HDF) ~ 3000 galaxies

Simulation of SKA Observation
What are we trying to find/measure?

- Gravitational waves (\( \sim \) nanoHz signal)
- Structure in the pristine primordial gas due to the first galaxies forming (> Mpc signal)
- Evolution of structure of interstellar gas in galaxies over the second half of the Universe’s age
- One-off flashes in radio and use them for Cosmology
- Imprint of primordial fluctuations on the distribution of galaxies in the nearby Universe
- The history of formation of stars in the Universe
- Dust -> pebbles -> boulders -> planets
THE SKA OBSERVATORY
The SKA Observatory – Phase 1: “SKA\textsubscript{1}”

People, buildings, roads, ground works, communications, electrical power, maintenance….

SKA\textsubscript{1}-Low: Aperture Array, 512 stations each with 256 antennas

SKA\textsubscript{1}-Mid: 180 15m-diameter dishes with single pixel receivers

Computers! Digital signal processing & general purpose computing
Role of computing/processing in SKA:

COMPUTING IS THE MAJOR PART OF SKA TELESCOPES BY DESIGN
Large “D” – vs – Large “N”

GBT 100-m diameter telescope

SKA LFAA prototype array

No 1 aim: collect as many photons as possible
No 2 aim: maximum separation of collectors -> achieve high angular resolution
Visibility:

\[ V(B) = E_1 E_2^* = I(s) \exp(i \omega B \cdot s/c) \]

Resolution determined by maximum baseline:

\[ \theta_{\text{max}} \sim \frac{\lambda}{B_{\text{max}}} \]

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

\[ \theta_{\text{dish}} \sim \frac{\lambda}{D} \]
KEY CHARACTERISTICS OF RADIO INTERFEROMETRY IMAGE PROCESSING
Radio Telescopes make **noisy** measurements

Illustration only – not real data

Sum of **all**
Can clearly see the fringes!
Sampling

- Each pair of telescopes results in a measured “visibility”
  - Loosely equivalent to a sample in the Fourier domain
- Inevitably areas without any samples:
  - At high radius -> limit of attainable resolution
  - At small radius -> limit of largest detectable structure
- Regular distribution of telescopes leads to a regular distribution of samples in uv plane

Illustrative example for the 27 antenna JVLA!
Sampling

- Earth rotation effective moves the sampling point of each pair of telescopes
- This limits the possible integration time duration
- Fills in the gaps between the previous samples (but not the central gap and outside the limit)

Illustrative example for the 27 antenna JVLA!
Radio Interferometers sample sparsely & irregularly

Sampling

- More rotation…

Illustrative example for the 27 antenna JVLA!
Redundant + Sparse Sampling

With SKA1 there will be very many individual visibilities in each observation:

- Accumulation of visibilities improves the signal/noise
- Fills in most of the uv plane

In general however to retain best signal/noise significant unevenness in sampling of the uv plane must be accepted.

-> Deconvolution

Illustrative example for the 27 antenna JVLA!
Measurements are imperfect – corrupted by slowly changing mechanical, electrical & atmospheric effects.

Iterative & joint solving for the image of the Sky & Calibration

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

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

- Further data parallelism in locality in UVW-space
- Use to balance memory bandwidth per node
- Some overlap regions on target grids needed
- UV data buffered either on a locally shared object store of locally on each node
Data-parallelism schemes

- Exploit frequency independence
- Distribute visibility data, duplicate target
- Gather and accumulate target grids

To manage total I/O from buffer/bus distribute Visibility data across nodes for same target grid which is duplicated
- Duplication of target provides fall-over protection
Radio Interferometry Distinguishing Features

- Data are intrinsically noisy (the galaxy, atmosphere, ground, interference, receivers)
- Measurements sparsely & irregularly sampled
- Accurate direct inversion of measurements into images not possible -> iterative method required
- Slowly varying instrumental and atmospheric effects corrupt the observed data -> jointly solve for the images of sky and the corrupting effects
- High degree of data parallelism & intermediate data dependencies can be expressed as tree reductions
• Noise -> lots of data to reduce noise
  – Data can’t be pre-averaged (because calibration & sampling issues)
• Sparse & Irregular sampling -> gridding to put data on a regular grid
• Iterative solving -> streaming solution not possible, must keep a substantial amount of data until reduction converged
SKA SCIENCE DATA PROCESSOR REQUIREMENTS
SDP Top-level Components & Key Challenges

Science Data Processor

Telescope Manager

SDP Local Monitor & Control

Data Processor
- High Performance
  - ~100 PetaFLOPS
- Data Intensive
  - ~100 PetaBytes/observation (job)
- Partially real-time
  - ~10s response time
- Partially iterative
  - ~10 iterations/job (~3 hour)

Long Term Archive
- High Volume & High Growth Rate
  - ~100 PetaByte/year
- Infrequent Access
  - ~few times/year max

Delivery System
- Data Distribution
  - ~100 PetaByte/year from Cape Town & Perth to rest of World
- Data Discovery
  - Visualisation of 100k by 100k by 100k voxel cubes

Regional Centres & Astronomers

CSP

Science Data Processor
Visibility Buffer / Data Islands

Data Island size TBD, depends on load-balancing, resilience and message passing efficiency.
• $N_{cu} R_{cu,FLOPS} > 130 \text{ PetaFLOPS}$
• $M_{cu,work} > 160 \text{ GigaByte}$
• $N_{cu} R_{cu,bw} > 260 \frac{\text{PetaByte}}{s}$
• $N_{cu} R_{cu,io} > 70 \frac{\text{TeraByte}}{s}$
• $N_{cu} M_{cu,buf} > 100 \text{ PetaByte}$
• $R_{cu,FLOPS} \geq 1 \text{ TeraFLOPS}$
Global Networking

- Tree reductions

\[ N_{di} R_{di,net} > 200 \text{TeraByte/s} \]

Total injection into the tree reduction -- not a bisectional bandwidth req (unless tree reduction not supported)
Hardware challenges

- Power efficiency
- Getting the right mix of storage, processing and networking capabilities
- Reasonably standard programming model, programming environment
  - Commissioning of large radio telescopes takes years -> continuity essential
Software Challenges

- Achieving high efficiency: op-ex
- Dealing with the failures
- Adaptability to future system and node architectures
- Minimise development & delivery risks
- Maintenance and Enhancement over planned 50yrs observatory lifetime
SKA Science Data Processor

ARCHITECTURE HIGHLIGHTS
• Hybrid programming model:
  – Dataflow at **coarse-grained** level:
    • About 1 million tasks/s max over the whole processor (-> ~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
  – 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
Why?

• Shared memory model essential at fine-grain to control **working memory** requirements

• Dataflow:
  – Load-balancing
  – Minimisation of data movement
  – Handling failure
  – Adaptability to different system architectures
- Classify arcs in the dataflow graph as precious or non-precious
- Precious data are treated in usual way – failover, RAID, etc.
- Non-precious data can be dropped:
  - If they are input to a map-type operation then no output
  - If they are input to a reduction then result is computed without them
- Stragglers outputting non-precious data can be terminated after a relatively short time-out
The non-precious data concept - illustration
Current Status

- **Established:**
  - Overall requirements, system interfaces
  - System decomposition

- **Provisional**
  - Sub-system requirements, internal interfaces
  - Key architectural choices

- **Next steps**
  - Identifying baseline technologies for all sub-systems
  - Identifying stable APIs and technologies for verification before commencement of construction