Computational challenges for SKA and SKA-Low

Tim Cornwell, Square Kilometre Array
SKA data processing challenges

- Wide field imaging
  - ASKAP, MWA, LOFAR
- Imaging with aperture arrays
  - LOFAR, MWA
- Imaging with phased array feeds
  - ASKAP
- Imaging with wide bandwidth
  - ASKAP, VLA
- Calibration and correction of direction dependent effects
  - LOFAR, MWA, MeerKAT, ASKAP, VLA
- All organizations represented at CALIM meetings
  - Challenges being addressed
- High performance computing lags
Types of processing

• SKA-Mid
  – Standard imaging
  – Pulsar timing and searching

• SKA-EOR
  – Standard imaging
  – Experimental EOR detection

• SKA-Survey
  – Standard (ASKAP-style) imaging
Costing

- The SKA Board wishes to set a cost cap
- SKA Office has to advise the Board by July 2013
- Constructing cost estimates for Baseline Design
- Baseline Design is not overly prescriptive
- Best guesses needed
SDP software platforms

• Assume one software platform for all science cases
  – Specializations and optimizations as necessary e.g. AA, PAF
• No existing software package can be used
• Many existing algorithms will fail to scale
• Third-party apps (e.g. casacore, wcslib) must be rewritten for multi-threading
  – Roughly €10M in existing casacore
• Software platform must be supported on multiple hardware platforms
  – Mitigate risk by facilitating use of non-SKA hardware
SDP hardware platforms

- Architecture of computing platform unknown
  - Many nodes of many cores (GPU or MIC) most likely
- Could have different architectures across telescopes or applications
  - Plan explicitly for non-SKAO platforms elsewhere
- Procurement delivery is staged to maximise capabilities and match to growing needs
  - As used at Pawsey Centre
- Shift procurement into operations to procure only when techniques are mature
SKA data processing rates

Note that Flops numbers are not achieved - we actually get much lower efficiency because of memory bandwidth - so scaling is relative.
Calibration and imaging cost model

- Five pipelines
  - Ingest, Calibration, Continuum, Spectral line, Transients

- Steps in processing
  - Gridding and degridding visibility data
  - Multi-Scale Multi-Frequency CLEAN

\[ T_{\text{clean}} = \mu_{\text{clean}} N_{\text{scales}} N_{\text{Taylor}} N_{\text{iterations}} N_{\text{psf}} \]

- Resources
  - Processing
  - Memory
  - Fast storage

Exploring the Universe with the world’s largest radio telescope
Typical costs

- Based on GPU scaling
- Runs all pipelines on longest baselines

<table>
<thead>
<tr>
<th>Telescope</th>
<th>Diameter</th>
<th>Baseline</th>
<th>Processor</th>
<th>Memory</th>
<th>Storage</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKA1_AA_Low</td>
<td>35</td>
<td>100</td>
<td>14.3</td>
<td>0.9</td>
<td>1.4</td>
<td>16.6</td>
</tr>
<tr>
<td>SKA1_Mid</td>
<td>15</td>
<td>100</td>
<td>19.2</td>
<td>4.8</td>
<td>1.3</td>
<td>25.3</td>
</tr>
<tr>
<td>SKA1_Survey</td>
<td>15</td>
<td>100</td>
<td>22.6</td>
<td>6.5</td>
<td>2.7</td>
<td>31.8</td>
</tr>
</tbody>
</table>
Gridding/degridding model

\[ T_{ws} = \mu_{\text{grid}} N_{\text{vis}} \left( \rho^2 \left( \frac{w_{\text{rms}}}{w_{\text{max}}} \right)^2 R_F^2 + R_A^2 \right) + \mu_{\text{FFT}} \left( 2 \rho R_F \left( \frac{T_{\text{obs}}}{T_A} \right) + N_{\text{int}} \right) N_{\text{pixels}}^2 \log_2 \left( N_{\text{pixels}}^2 \right) \]

\[ + \mu_{\text{reproj}} \frac{N_{\text{pixels}}^2 h_{\text{obs}}}{\rho} \]

- AW snapshots algorithm
  - Identify best-fit plane in uvw space
  - Use AW Projection to move points onto this plane
  - Update plane at rate chosen to minimize total work
  - Some projection mandated by Earth curvature
  - Correct for coordinate distortion of each image plane

- Superior to snapshot imaging and AW Projection
  - Less CPU, memory

- Diagonal or general Mueller matrices

- Update or change model in the future as appropriate

Exploring the Universe with the world’s largest radio telescope
Data handling

- Assume that single pass processing is insufficient
- Hence data must be buffered for ~ days to allow multiple calibration passes
  - Might be continuum only but could be all data
- Require large multi-day visibility data buffer for all telescopes
- Assume average throughput must = 100%
Cost of field of view

Fresnel number (blue). Number of pixels per axis (green)
CPU-based scaling numbers

- Performance measured by four numbers
- Can be benchmarked by small programs
  - tConvolve $\mu_{wp}, \mu_{FFT}, \mu_{reproj}, \mu_{clean}$
- Numbers shown are from Pawsey Centre
- ASKAP Real Time Computer
  - 200TF system costing €3.2M
- Expect SDP to update cost model regularly during pre-construction and construction
  - Many assumptions that can fail at scale

<table>
<thead>
<tr>
<th>$\mu_{grid}$</th>
<th>8.0e-13 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{FFT}$</td>
<td>1.1e-11 s</td>
</tr>
<tr>
<td>$\mu_{reproj}$</td>
<td>9e-10 s</td>
</tr>
<tr>
<td>$\mu_{clean}$</td>
<td>1.0e-12s</td>
</tr>
<tr>
<td>Cost (2012)</td>
<td>€1.6M</td>
</tr>
</tbody>
</table>
GPU-based scaling numbers

- Somewhat speculative
- Grid, clean measured
- FFT, reproject scaled
- Substantially better than CPU

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\text{grid}}$</td>
<td>8.0e-14</td>
</tr>
<tr>
<td>$\mu_{\text{FFT}}$</td>
<td>1.1e-12</td>
</tr>
<tr>
<td>$\mu_{\text{reproj}}$</td>
<td>9e-11</td>
</tr>
<tr>
<td>$\mu_{\text{clean}}$</td>
<td>2.0e-13</td>
</tr>
<tr>
<td>Cost (2012)</td>
<td>€1.6M</td>
</tr>
</tbody>
</table>
Cost of resolution

Processing for AA_Low, 35 m diameter, vs baseline, GPU

- Ingest
- Calibration
- Continuum
- Spectral Line
- Transient
- All

Exploring the Universe with the world’s largest radio telescope
Cost of diameter, fixed field of view

Processing of AA_Low, 100km baseline, vs station diameter, fixed FOV

Exploring the Universe with the world’s largest radio telescope
Things we need to know for costing

• Observing scenarios
• Calibration of processing numbers
• Firmer numbers for
  – Maximum baseline
  – Station diameter
  – Number of mosaiced beams
  – Number of (science) spectral channels
• Thoughts on beams, polarisation, etc.
• Other science cases