ASKAP Pipeline processing and simulations

Dr Matthew Whiting
ASKAP Computing, CSIRO
May 5th, 2010
ASKAP Computing Team Members

• Team members
  • Marsfield: Tim Cornwell, Ben Humphreys, Juan Carlos Guzman, Malte Marquarding, Tony Maher, Euan Troup, Max Voronkov, Matthew Whiting, Xinyu Wu (from May)
  • Narrabri: Dave Brodrick
  • Parkes: Simon Hoyle
  • Dwingeloo: Ger van Diepen
  • Socorro: Urvashi Rau

• Mostly software engineers
  • Astronomers - Tim, Max, Matthew
Using ASKAP

- Synthesis image of Centaurus A
  - Required 1200 hours observing on the Australia Telescope Compact Array

- ASKAP will take ~ 10 minutes

Image: Ilana Feain, Tim Cornwell
Response to Imaging Challenges

- Large size of data flow presents challenges for processing
- Control complexity of the data and manage computing load
- 3-axis mount partially in response to this
  - Removes need for beam rotation in software
- Maintain an accurate Global Sky Model
  - Used to subtract bright sources from continuum sky before processing
  - Derived during early operations for entire sky
  - Science imaging uses GSM as starting point - small differences only
- Post-gridding weighting schemes used
  - Gridding/degridding dominates computing load (90%)
- PSF good enough to not need full deconvolution of spectral-line and transient pipeline data
- Distributed processing necessitates custom-built pipeline software
  - Existing tools not parallelisable (CASA, miriad, …)
Spectral-line Pipeline overview

- Keep full 16384 channels
- Two polarisations only
- Global Sky Model subtracted prior to imaging
  - Use previously-made model and predict forward
  - GSM needs to be of sufficient S/N to not degrade sensitivity
- No deconvolution in imaging
  - Maybe for a few nearby galaxies, or Galactic objects, but single major cycle only
  - Array configuration means PSF is good enough
- Cataloguing done immediately
  - Measure as much as we can while we can
- Create image cutouts, moment maps and spectral plots
- Will have data quality evaluation built in to pipeline
Other Pipelines

• **Continuum mode**
  - Average visibility stream to 256 channels. Keep full Stokes
  - Able to do full imaging with deconvolution over full 30 sq.deg.
  - Always have a good continuum model of the sky
  - Automatic cataloguing
    - Keep continuum model up to date for calibration
    - Create science catalogues

• **Transient mode**
  - Correlator outputs every 5 seconds
  - Average visibilities to 32 channels
  - Remove sky model and look for new / varied sources
  - Raise alerts and maintain light-curves
  - Search on longer timescales as well
Source finding

• ASKAP pipelines will have automatic source extraction and cataloguing built in

• Capable of:
  • Forming science catalogues for continuum, emission line, absorption line and transient data products
  • Keep Global Sky Model up to date
  • Produce associated data products for each source: moment maps, integrated spectra plots, cutouts

• Integrate with imaging pipeline, to make use of data while in memory
  • Large data products, so want to minimise disk I/O
  • Utilise distributed processing to minimise memory usage per cpu
**Duchamp Source Finder**

- *Duchamp* is an open-source astronomical source finder optimised for spectral-line cubes
- Stand-alone program developed outside ASKAPsoft, but library used in ASKAP pipelines as 3rd-party package
- Basic algorithm is:
  - Define threshold (use cube statistics)
  - Search each 2D image or 1D spectrum, recording detected objects
  - Combine objects to form 3D structures
  - Measure parameters from combined sources
- Smoothing or wavelet reconstruction possible to enhance objects above the noise
- Range of graphical output to help understand detections

ASKAPsoft source detection

- Source detection currently built upon *Duchamp*
- Uses code as a library, and adds functionality on top
- Parallel-processing done in coarse-grained fashion:
  - Split up image in user-defined manner
  - Each chunk searched separately with Duchamp algorithms
  - Any sources at/near edge are handled by a master process
  - Reduces memory impact on any single cpu
- Other features not (yet) in stand-alone package
  - For 2D case, have Gaussian fitting capabilities
    - Decompose detected sources into Gaussian components
    - Functionality comes from casacore library
  - Variable detection threshold across image
  - Integration with multi-frequency synthesis output

- Form of ASKAP source detector(s) to be determined in consultation with science teams (ASKAP Working group 2)
  - Developing protocols to convert algorithms into pipeline code
ASKAP End-to-End Simulations

• Construct simulations of the full pipeline processing
  • Sky Model → Telescope Simulator → Imager → Analysis

• Necessary for testing pipeline software
  • How does imaging work with appropriate range of source properties?
  • How does imaging scale to appropriate sizes?
  • How does source extraction perform?

• Necessary for testing science questions
  • How does imaging pipeline treat various source types?
  • How does source extraction perform?
  • What survey strategies will be appropriate?

• Coordination of these activities across SSPs managed through an ASKAP Working Group

• ASKAP Computing responsible for running simulations with prototype pipelines and disseminating resulting data
Sky Models

- Telescope simulator needs a “true” sky to observe
- Provide with image of sky model
- Current simulations start with SKADS Simulated Skies
  - S3-SEX (Willman et al 2008) and S3-SAX (Obreschkow et al 2009)
- Convert catalogues into images by pixelising each component
  - Both point sources and extended (not diffuse) sources
  - S3-SAX has double-horned analytic spectral profiles
- Precess to desired location on sky
Telescope simulator & imager

- “Observe” the sky model to produce visibilities
- Required inputs:
  - Sky model
  - Array configuration and synthetic beam distribution
  - Correlator configuration (frequencies, polarisations)
  - Integration time etc
  - Gridding parameters

- Imager then reads visibilities to create image
- Parallel processing via master-worker arrangement under MPI
- Highly scalable if no cleaning required
  - Have not been cleaning spectral-line cubes, so able to run at high efficiency on 256 cores at NCI National Facility
  - Do not have distributed MS clean working yet, so cleaning (e.g. Continuum data) is not as efficient in cpu use
Continuum simulation
Continuum simulation (zoom)
Spectral-line simulation

- Use S3-SAX as input model
- 2km core of the ASKAP configuration
  - 30 antennas only, giving natural resolution of 30"
- Coarser channels than planned for ASKAP
  - 92.5 kHz: binned 5x
  - Increase redshift coverage in simulation while keeping number of channels manageable
- Simulate an 8-hour observation, with 5-sec integrations
  - Large amount of data: 5.5 TB of measurement sets
- Simulating done per channel
- PAF simulated with 32 synthetic beams
  - Initial simulation: spaced at 1° separation
  - Not critically sampled, so noise variations seen
  - 2nd simulation had 0.5° separation - noise more uniform
Spectral-line simulations
Example sources

#018: J123020–451214  12:30:20.89, −45:12:14.41, 6585.292 km/s
  F_{1.2}=0.100 Jy km/s, F_{peak}=0.0034 Jy/beam, S/N_{max}=3.78
  W_{RA}=0.50, W_{DEC}=0.50, W_{3o}=129.4 km/s, W_{vel}=78.579 km/s, W_{vel}=122.334 km/s
  Centre: (233.9, 182.5, 675.5), Size: 23 voxels, Range: [233.235, 182.184, 673.679]

#019: J123041–451705  12:30:41.11, −45:17:05.02, 6706.989 km/s
  F_{1.2}=0.281 Jy km/s, F_{peak}=0.0058 Jy/beam, S/N_{max}=6.50
  W_{RA}=0.67, W_{DEC}=0.67, W_{3o}=78.650 km/s, W_{vel}=95.393 km/s, W_{vel}=61.220 km/s
  Centre: (212.6, 153.5, 669.6), Size: 39 voxels, Range: [211.214, 152.155, 668.671]

#020: J123052–451647  12:30:52.42, −45:16:47.16, 6859.538 km/s
  F_{1.2}=0.49 Jy km/s, F_{peak}=0.0124 Jy/beam, S/N_{max}=13.87
  W_{RA}=1.00, W_{DEC}=1.00, W_{3o}=59.532 km/s, W_{vel}=72.658 km/s, W_{vel}=81.708 km/s
  Centre: (200.7, 155.2, 662.1), Size: 100 voxels, Range: [198.203, 153.158, 660.664]
Source-finding Test Cubes

- Provide small cubes with noise reduced by ~30x
- Many more sources for testing source finders
- Show PSF structure for the brightest sources (cube is not cleaned)
- Use two declinations: $\delta = -45^\circ$ and $\delta = -3^\circ$
- Central part of field only, so sensitivity very uniform
Future simulations

• Want to simulate different ASKAP capabilities
  • Polarisation: S3-SEX + RM mask + %pol mask [end May]
  • Transients: coarsely sampled with known variable sources inserted [end May]
  • Galactic HI: simple spectral-line observation initially [July]
  • Diffuse continuum
• Additional extragalactic HI?
  • Alternative simulation inputs?
  • Larger, resolved galaxies?
  • Need to make appropriate use of limited resources
• Welcome input via ASKAP working group 1
  • Next meeting will be early June
Thank you