www.csiro.au

Simulations of ASKAP data processing

Tim Cornwell
Ben Humphreys
Maxim Voronkov
Matthew Whiting
Australian Square Kilometre Array Pathfinder

CSIRO

An illustration of the speed of the Pathfinder

- ATCA image of Centaurus
- Required 1200 hours observing on the Australia Telescope Compact Array in Narrabri

- The Pathfinder will take about 10 minutes

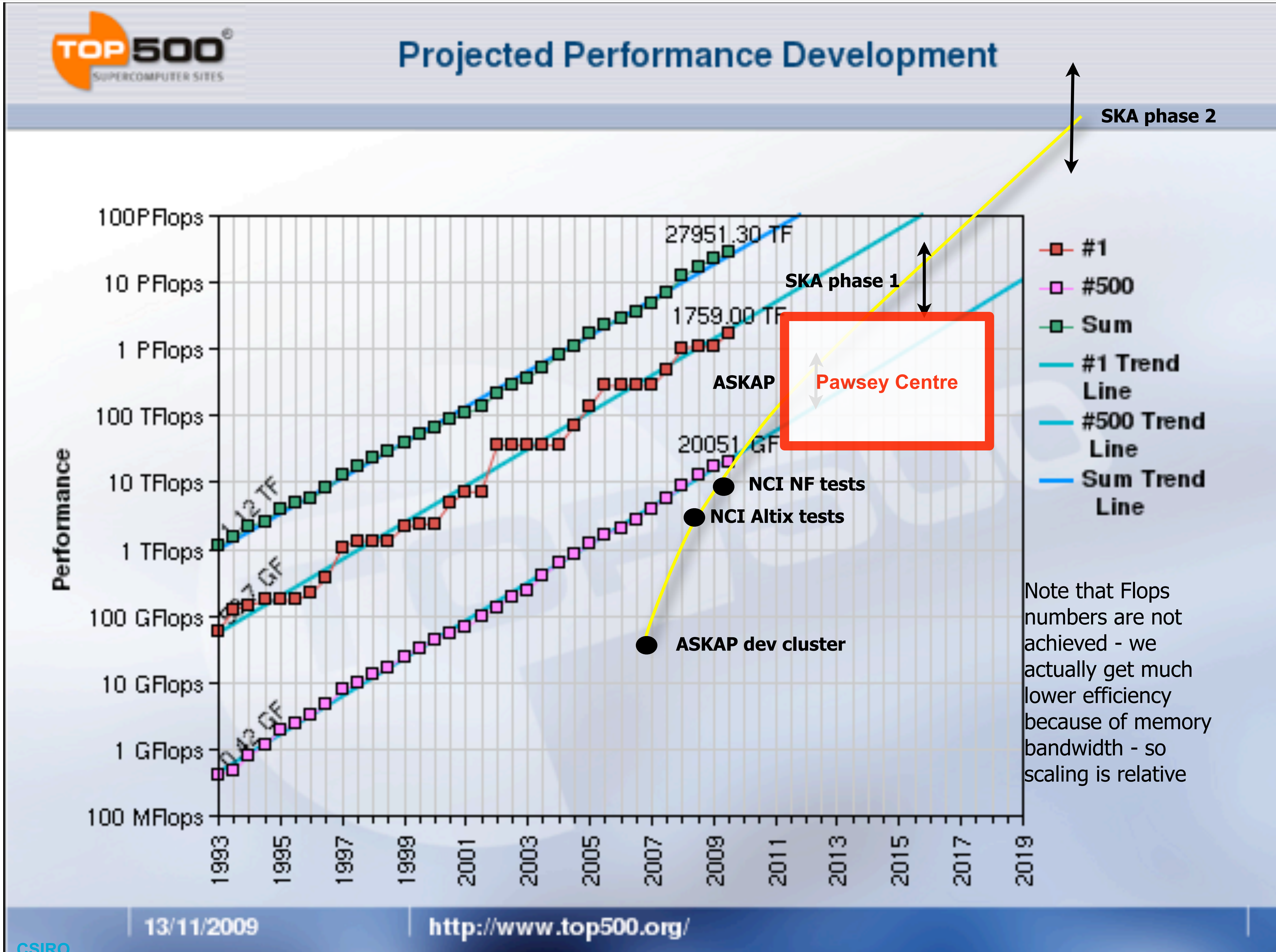
ASKAP data processing

- Large field of view implies large data rate
  - Data rate = 2.8GB/s
- Data processing load
  - ~ 100 TFlop/s for 30 arcsec resolution
  - ~ 1 PFlop/s for 10 arcsec resolution
- Process data from observing to archive with no human decision making
  - Calibrate automatically
  - Image automatically
  - Form science oriented catalogues automatically
- Scaling to 100 TFlop/s and then 1 PFlop/s still in progress
  - Using NCI Constellation to get to ~ 20 - 30 TFlop/s
  - Require Pawsey Petascale system (2013) to get to 1PFlop/s
- Goal is to get 80% efficiency at full scale
- Requirements for processing
  - x86-64 cluster or supercomputer
  - 1 PFlop/s, 40Gb/s input, large memory

Pawsey HPC Centre for SKA Science

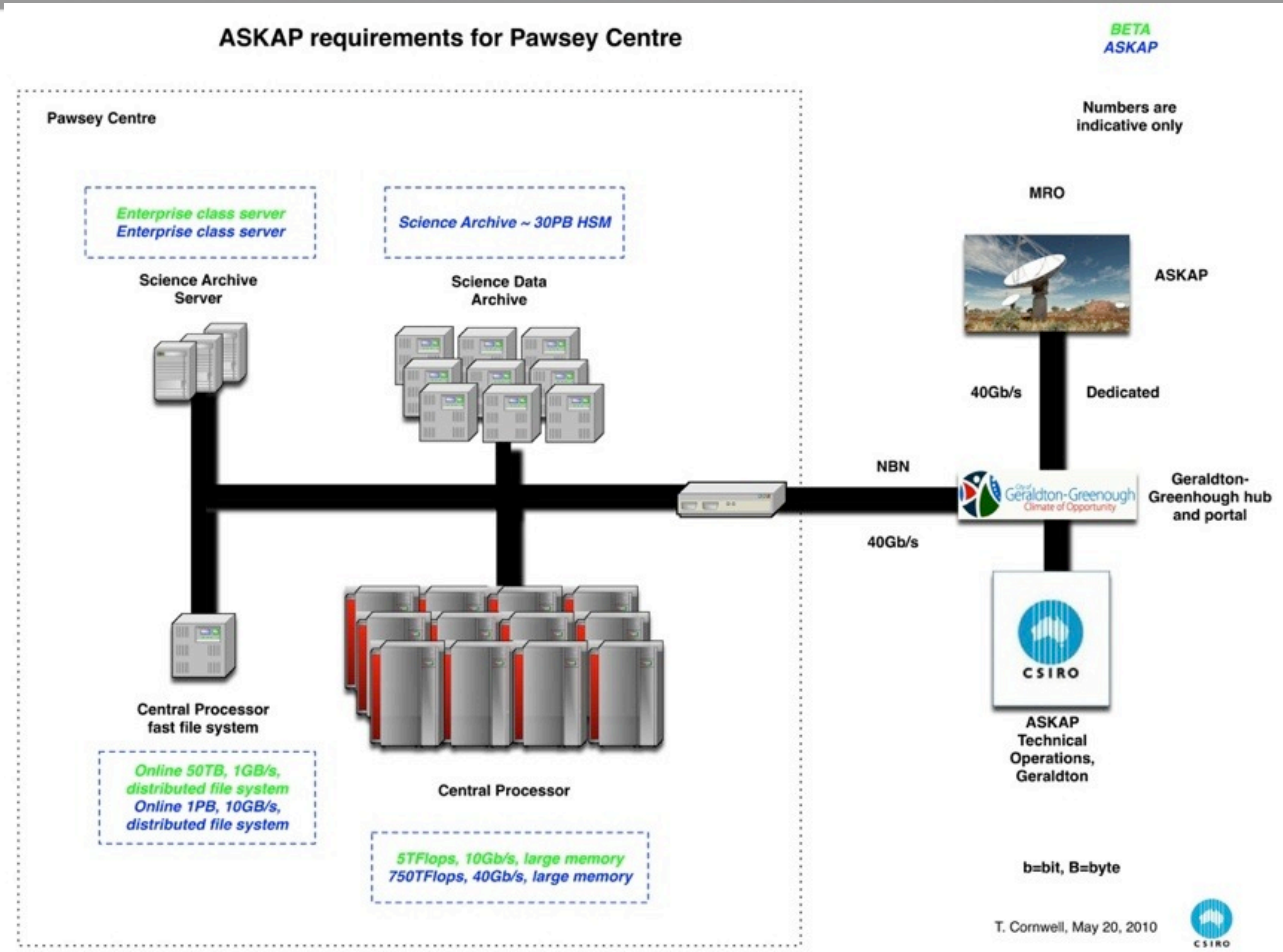
- AU$80M to build Petascale system by 2013
- Radio astronomy guaranteed 25%, ~ 25% more merit-based
  - ASKAP, MWA prime radio astronomy users
- CSIRO managing construction project
- iVEC will operate
- MRO to Pawsey Centre network connection ~ 40Gb/s
  - Many dark fibres that can be lit for SKA

Climbing Mount Exaflop



Note that Flops numbers are not achieved - we actually get much lower efficiency because of memory bandwidth - so scaling is relative

Possible ASKAP computing configuration



Simulation, Scaling, Verification, Validation

- Comprehensive ASKAP/SKA program
  - Simulation: Can we simulate and process observations?
  - Scaling: Can the processing scale up to many cores?
  - Verification: Does the processing do the right thing?
  - Validation: Does the processing do it correctly?
- ASKAP has developed highly distributed simulation and reduction software
  - Current effort ~ 10 FTE-years, expect ~ 10 more FTE-years or more
- Simulate ASKAP data reduction for Survey Science Teams
  - Well underway
  - Using Intel 32 core/128GB loaner and NCI National Facility Vayu
- Simulate and process SKA observations
  - Work upward in scale and complexity
  - Contribution to SKA PrepSKA
- Scaling processing to 80% efficiency on 1 PFlop/s resource
- Verification and validation
  - Simulations
  - Comparisons to other packages *e.g.* CASA
  - Processing ATCA, EVLA, LOFAR data

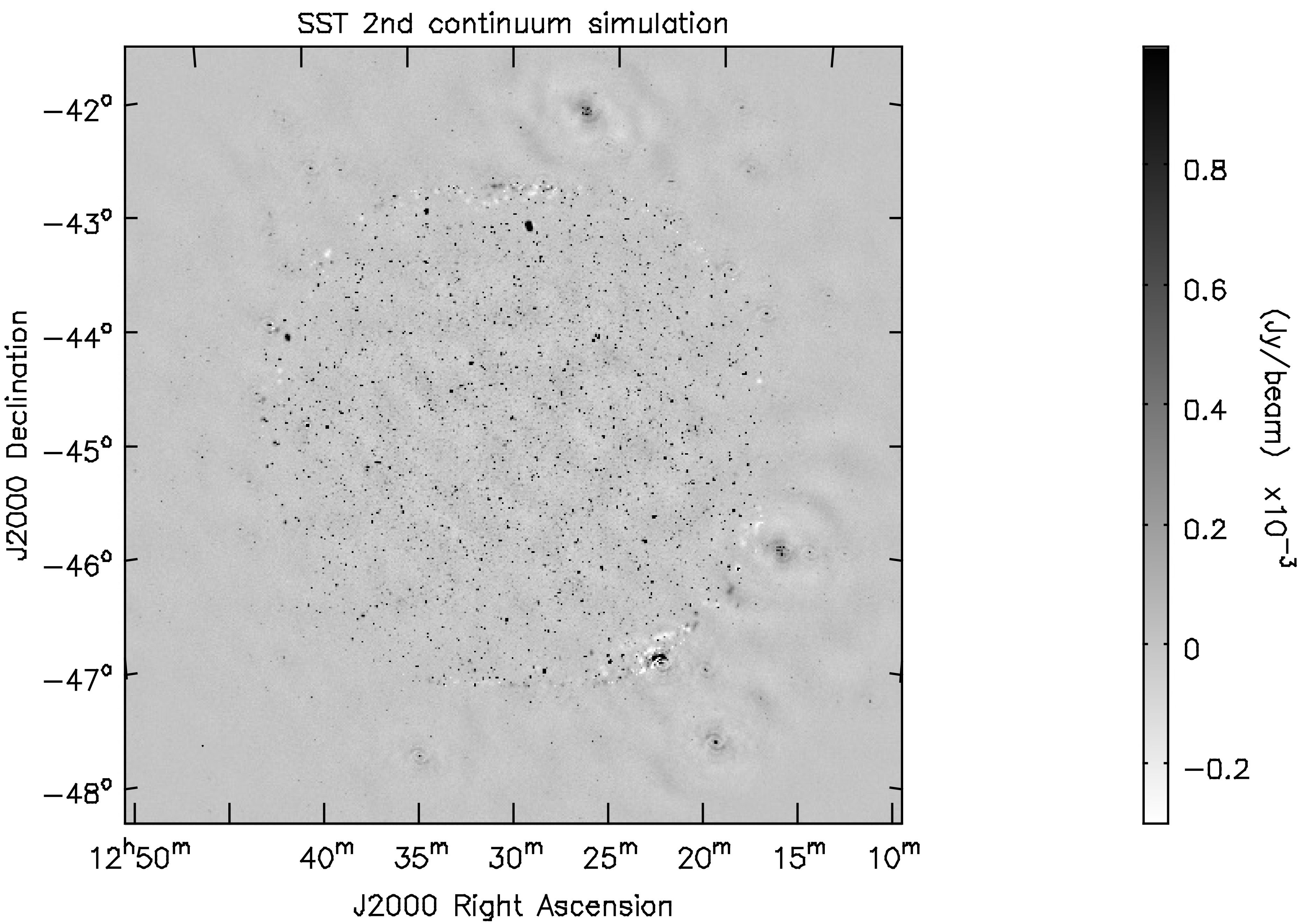
ASKAP imager

- All synthesis code written from scratch
  - Legacy code (e.g. CASA) not suitable for large scale distributed processing at high efficiency
- C++/MPI code
  - Distributes processing over multiple nodes
- Capabilities
  - Imaging using AWProjection or AProjectionWStack
    - Blocked Airy disk or ATCA primary beam model
  - MultiScale CLEAN
  - MultiScale CLEAN + Multi-Frequency Synthesis (Urvashi PhD)
  - SNR-based CLEANing
  - Wiener filtering post-imaging instead of usual weighting pre-imaging
- In testing now
  - ASKAP simulations
  - ATCA reprocessing
  - LOFAR commissioning

ASKAP simulations

- Consumers
  - ASKAP Science Survey Teams
  - ASKAP Computing Team
- Simulations ~ data challenges
- Excellent way to find problems
- Drives improvements in functionality, performance, and scaling
- Performed using ASKAP software running on one of:
  - 8 node 4 core Intel cluster, purchased 2007
  - 32 core Intel machine, 128 GB donated by Intel
  - Sun Constellation at National Computational Facility
- Results available from ASKAP Redmine site
  - http://pm.atnf.csiro.au/askap/wiki/sup/
- Or Matthew Whiting's page:
  - http://www.atnf.csiro.au/people/Matthew.Whiting/ASKAPsimulations.php

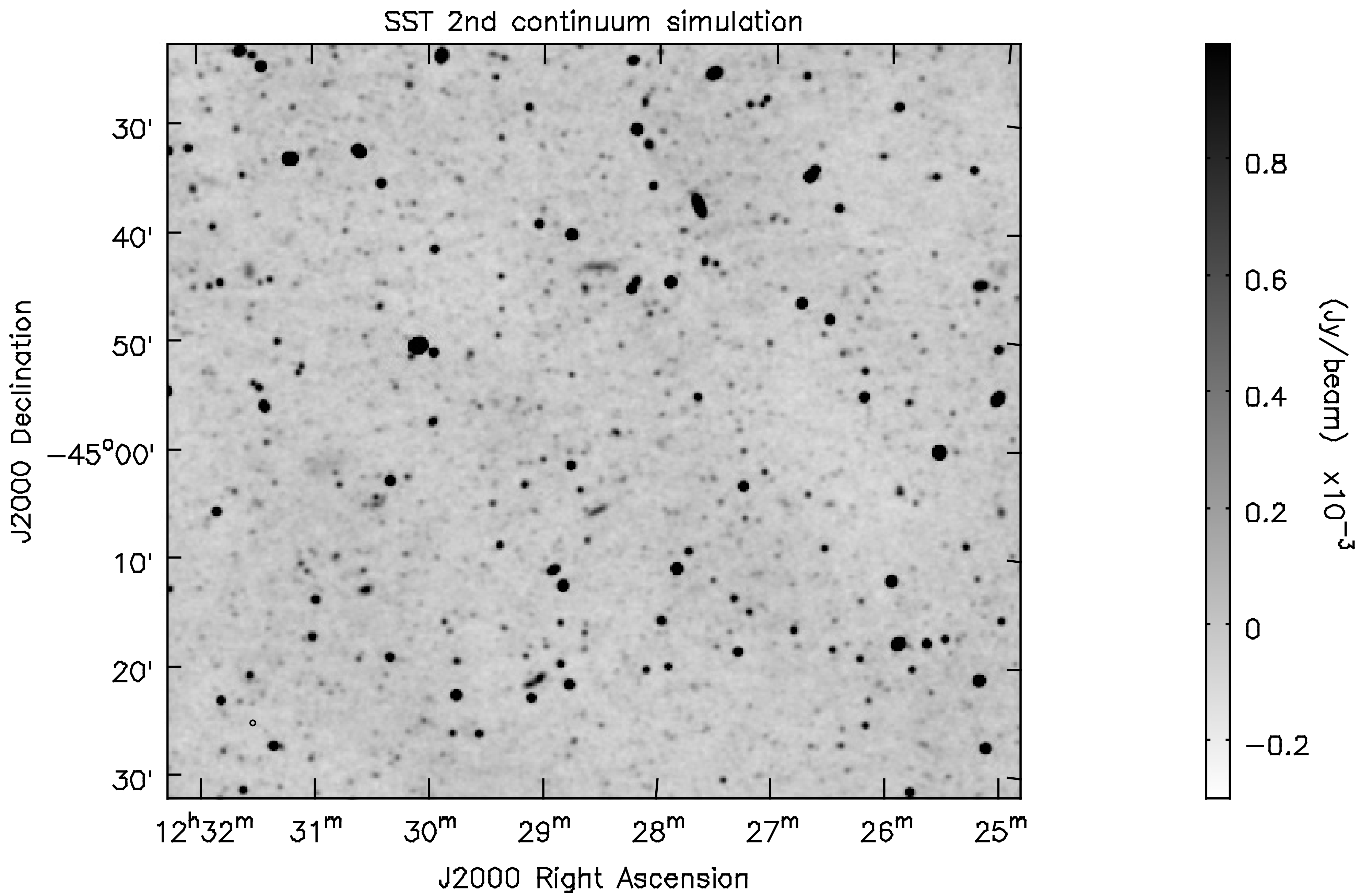
SST2 continuum image: 30 arcsec resolution



- SKADS sky model
- 2km configuration
- +/- 4 hours
- 120 2.53 MHz channels
- 20s integration
- 32 PAF beams separated by 0.5deg
- 11.5 uJy/beam noise
- AWProject with variable support
- 30" resolution
- Made possible by pre-production system loaned by Intel

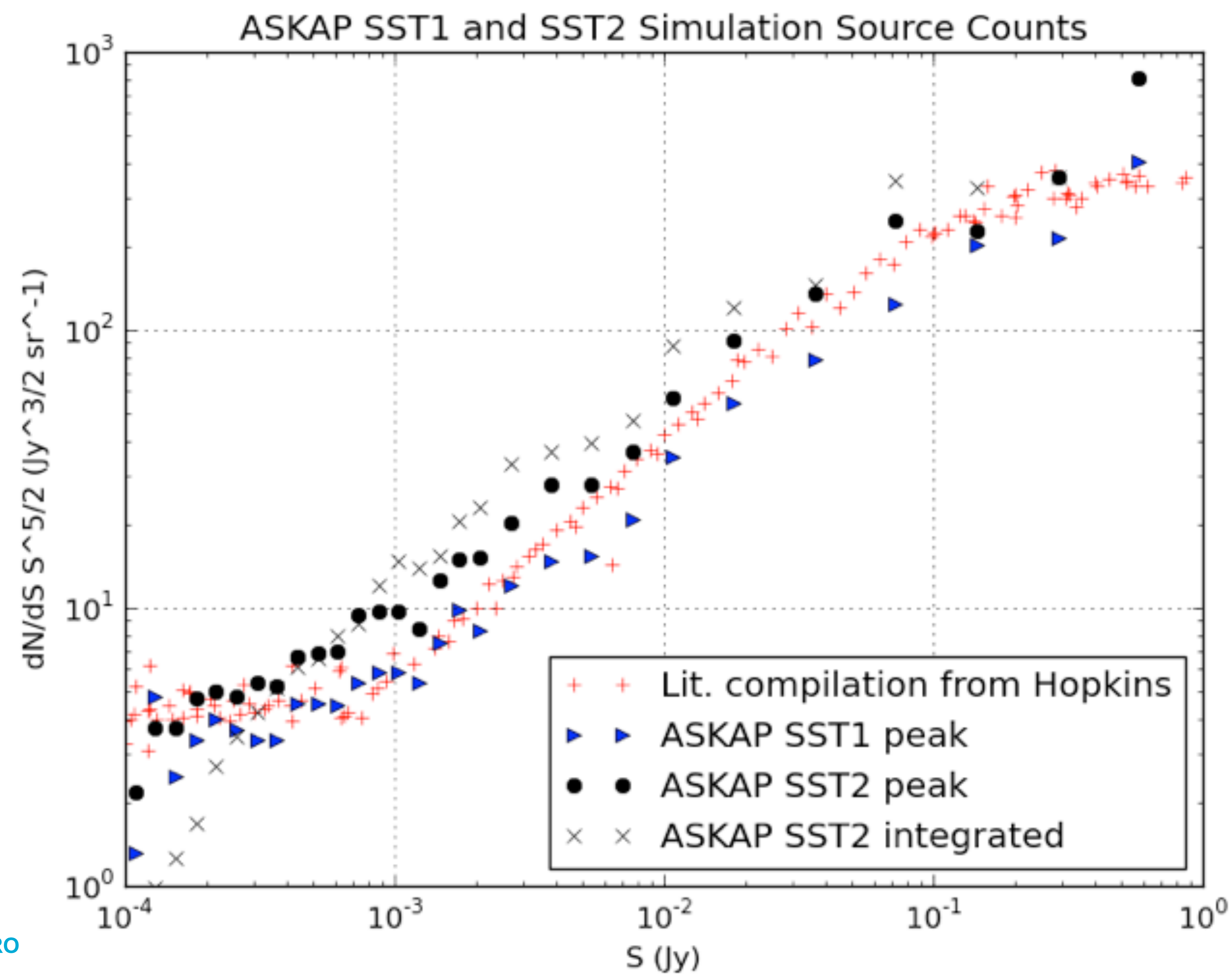
- No preconditioning
- Image equalised
- April 16, 2010

SST2 continuum image (zoomed)



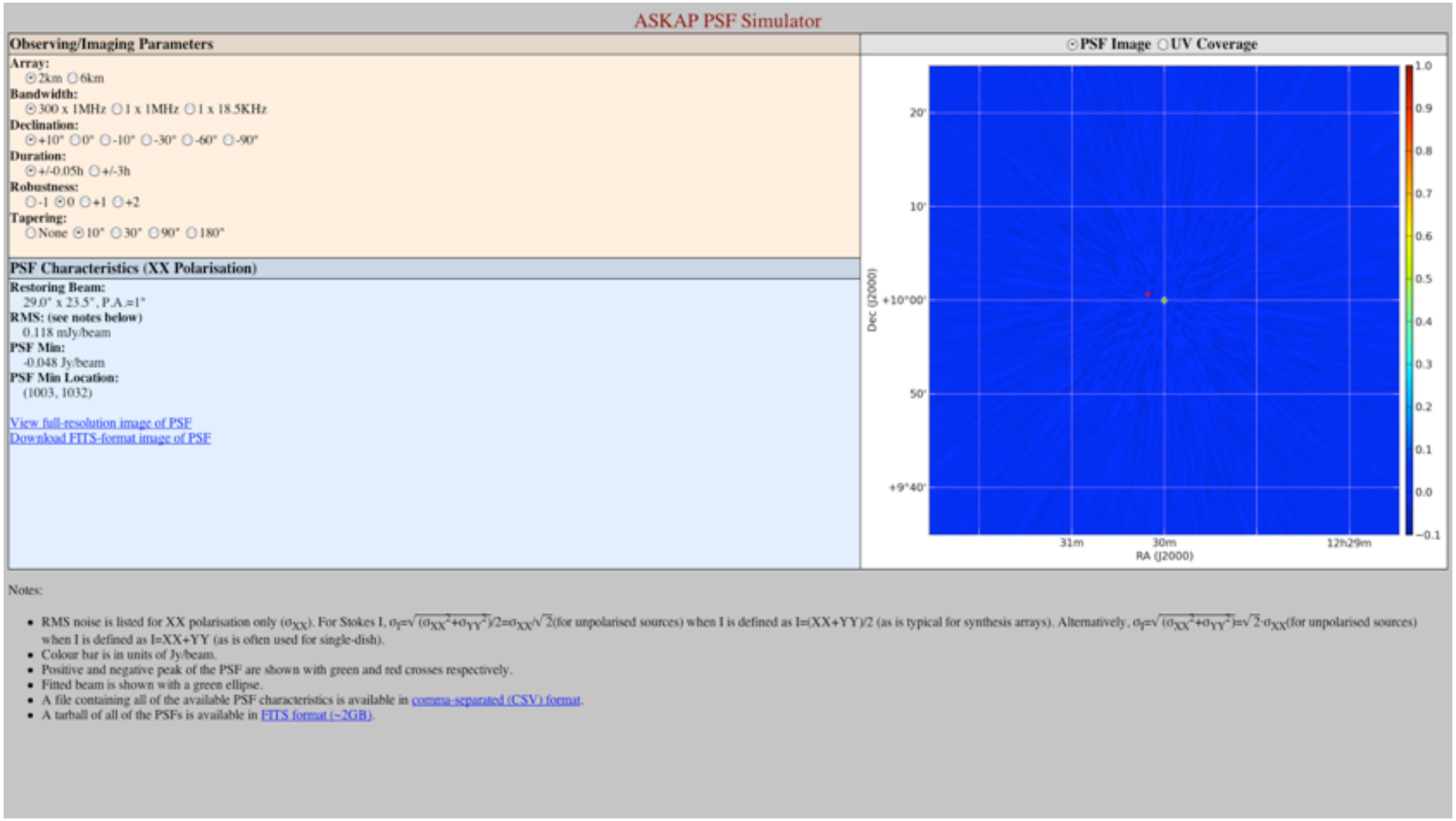
Comparison with actual source counts

- Test of SKADS model and ASKAP simulations
  - Performed by Tara Murphy of VAST Survey Science Team
  - Using Matthew Whiting's Duchamp program

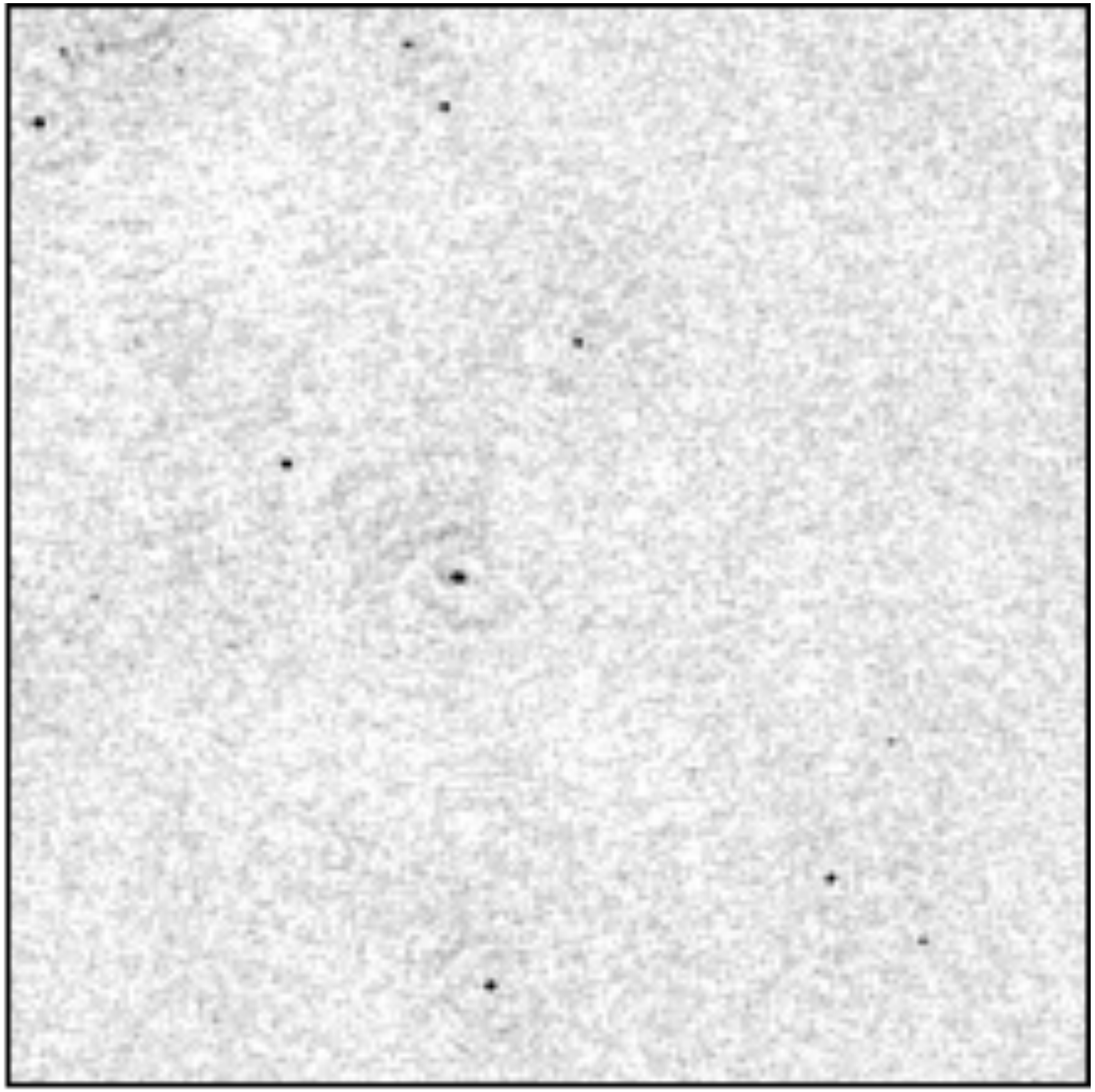


ASKAP PSF simulator

- Statistics, images for selected ASKAP PSFs (3600!)
- Can download as JPG or FITS files
- http://www.atnf.csiro.au/people/Emil.Lenc/ASKAP/psf/view.html



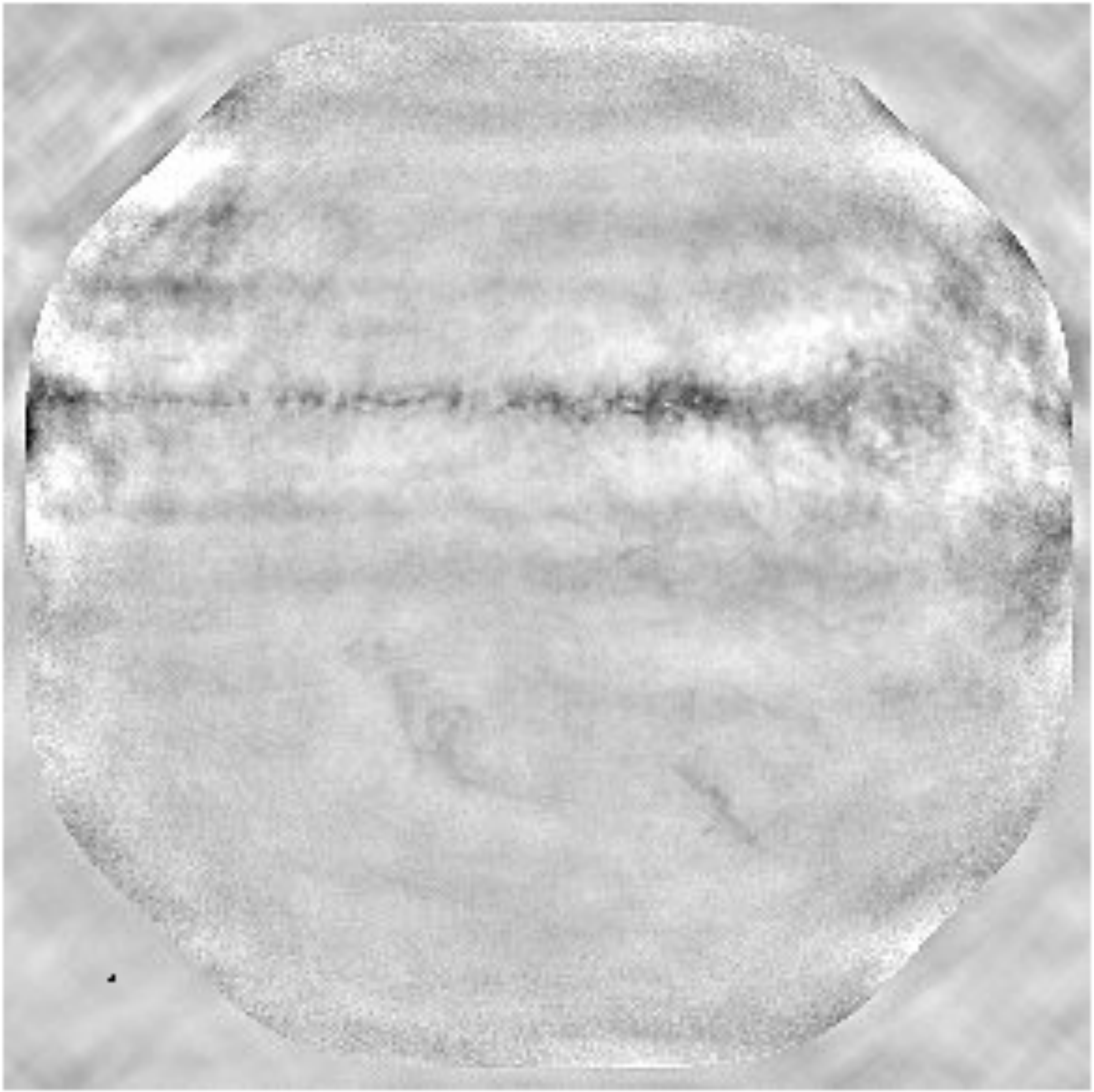
SST2 spectral line image



- SKADS S3-SAX sky model
- 2 km core of the ASKAP configuration (30 antennas only)
- 32 idealised beams, spaced in a rectangular grid 0.5° apart
- Spatially uniform noise
- 1024 channels of 92.5 kHz width
- 1327.39 to 1422.0175 MHz.
- 8 h, made up of 5 s integrations.
- Total data volume = 5.5 TB
- No preconditioning or deconvolution
- Processed on NCI Vayu

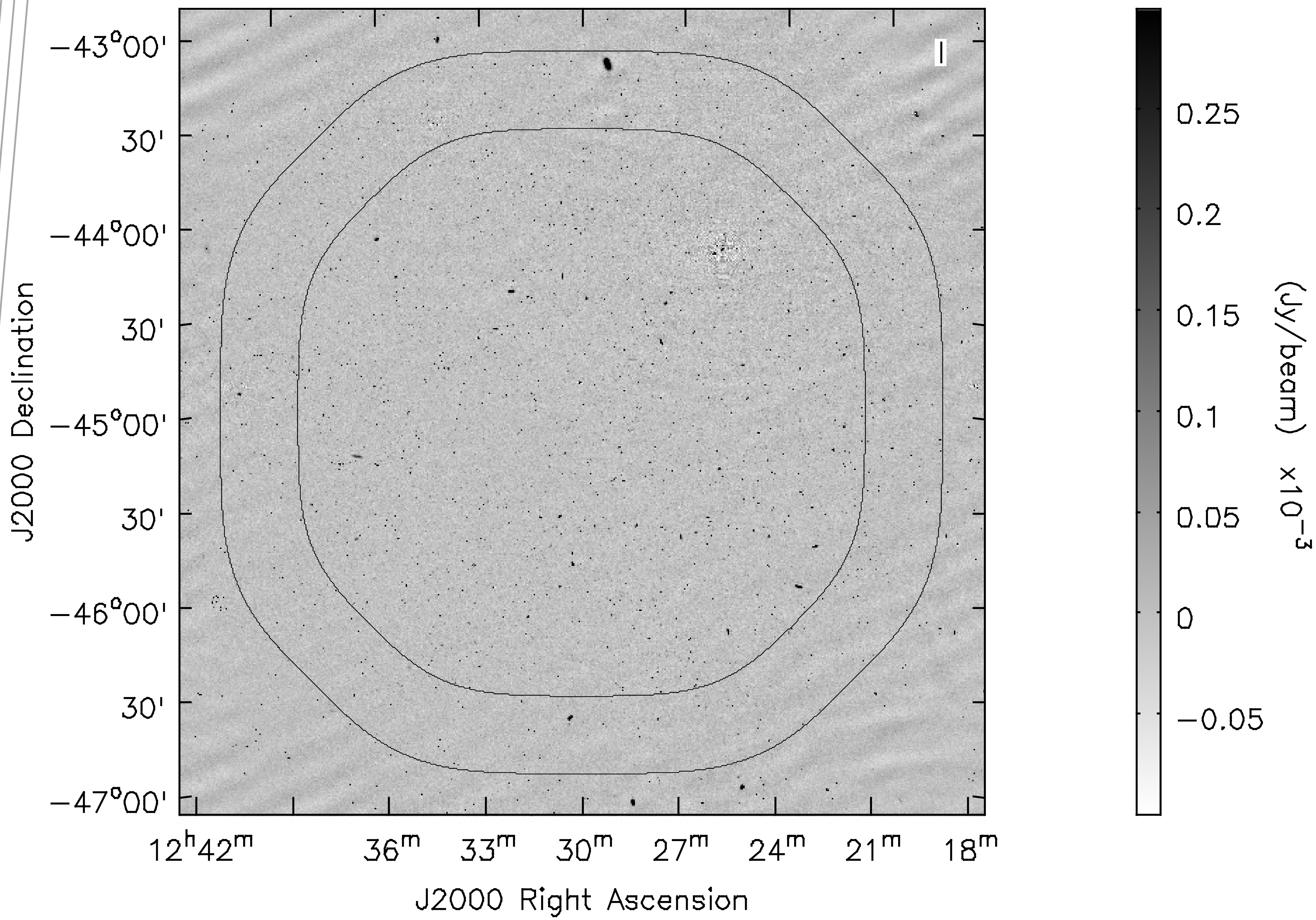
- April 16, 2010

GASKAP - Galactic plane spectral line



- Galactic plane shrunk to fit inside ASKAP FOV
- 2km configuration
- +/- 4 hours
- 646 3.91kHz channels
- 60s integration
- 32 PAF beams separated by 0.5deg
- AWProject with variable support
- 30” resolution
- Wiener filtering, robust +0.25
- Processed on NCI Vayu machine - Sun Constellation
- Image not deconvolved
- 30 July 2010

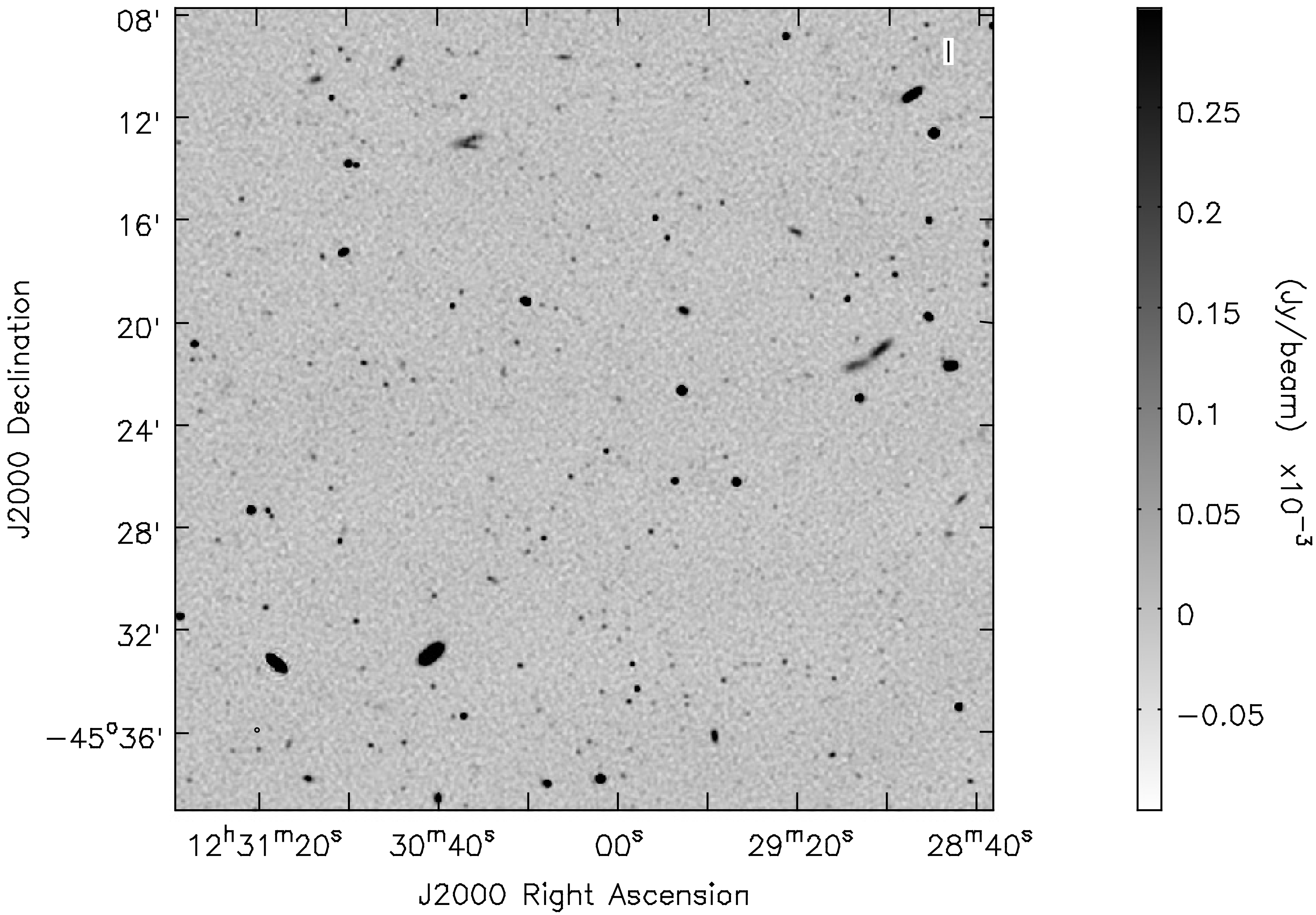
SST3 continuum image: 10 arcsec resolution



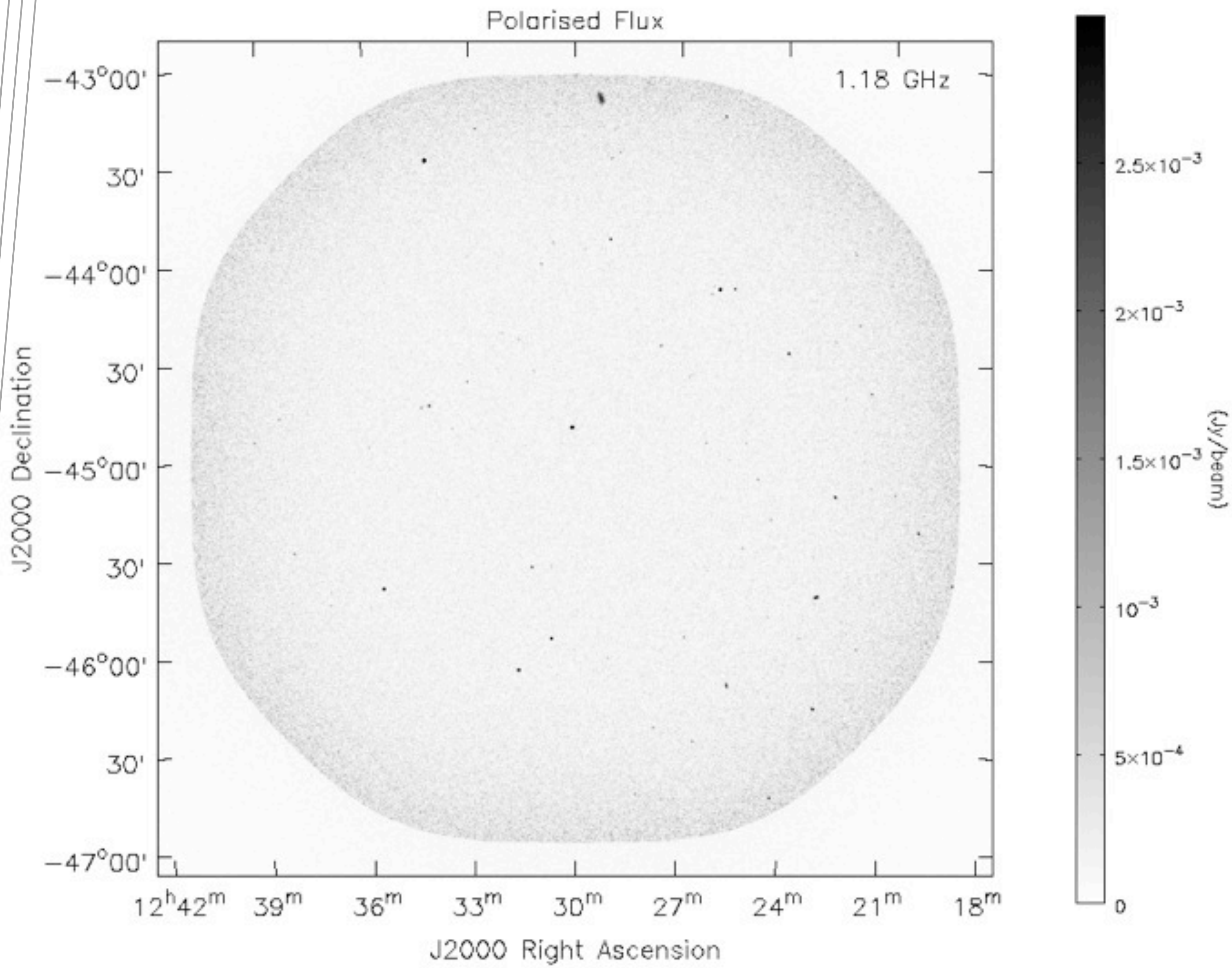
- SKADS sky model
- 6km configuration
- +/- 4 hours
- 128 2.53 MHz channels
- 20s integration
- 32 PAF beams separated by 0.5deg
- 14.1 uJy/beam noise
- AProjectWStack
- No wplanes in simulation or imaging
- 10" resolution
- Made possible by pre-production system loaned by Intel

- Image equalised
- May 25, 2010

SST3 continuum image (zoomed)

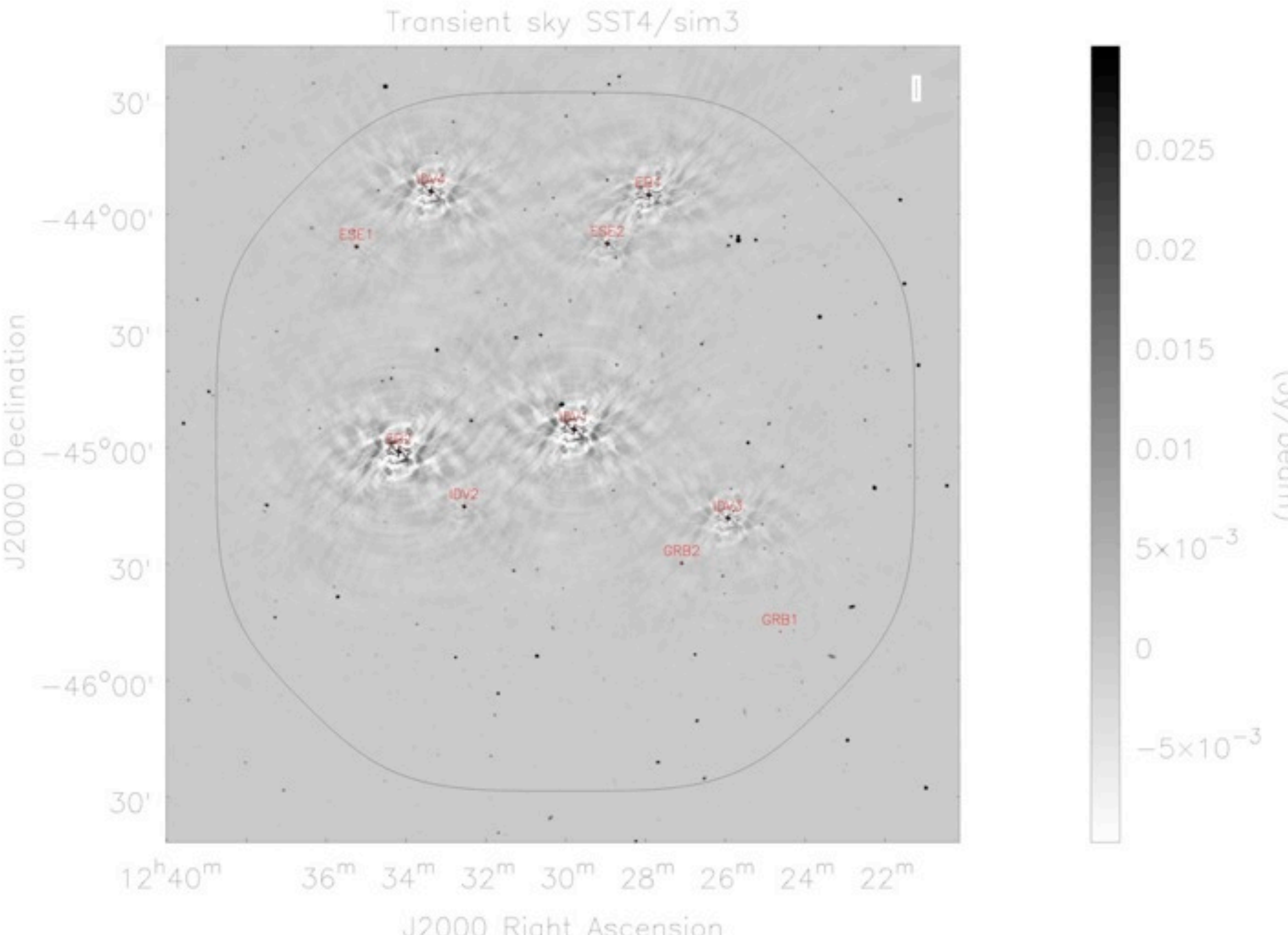


Polarised image



- The input model used was an updated version of the SKADS S3-SEX catalogue (Wilman et al 2008) provided by Jeroen Stil of the POSSUM team.
- SKADS catalogue augmented by Q & U fluxes for each component plus a rotation measure made up of a random "internal" component and foreground.
- 2km core (ie. 30 antennas only)
- PAF has 32 idealised beams, on rectangular grid 0.5deg spacing
- 8hr observing, 10s integrations
- 32 channels, each 8MHz in width, 1420MHz to 1172MHz
- Each channel was simulated and imaged separately, then combined into a cube.
- Normalised Wiener filtering with robust = +0.25, Gaussian tapering with 27.6 x 27.6 arcsec.
- Processed on NCI Vayu

SST4 continuum + transients image



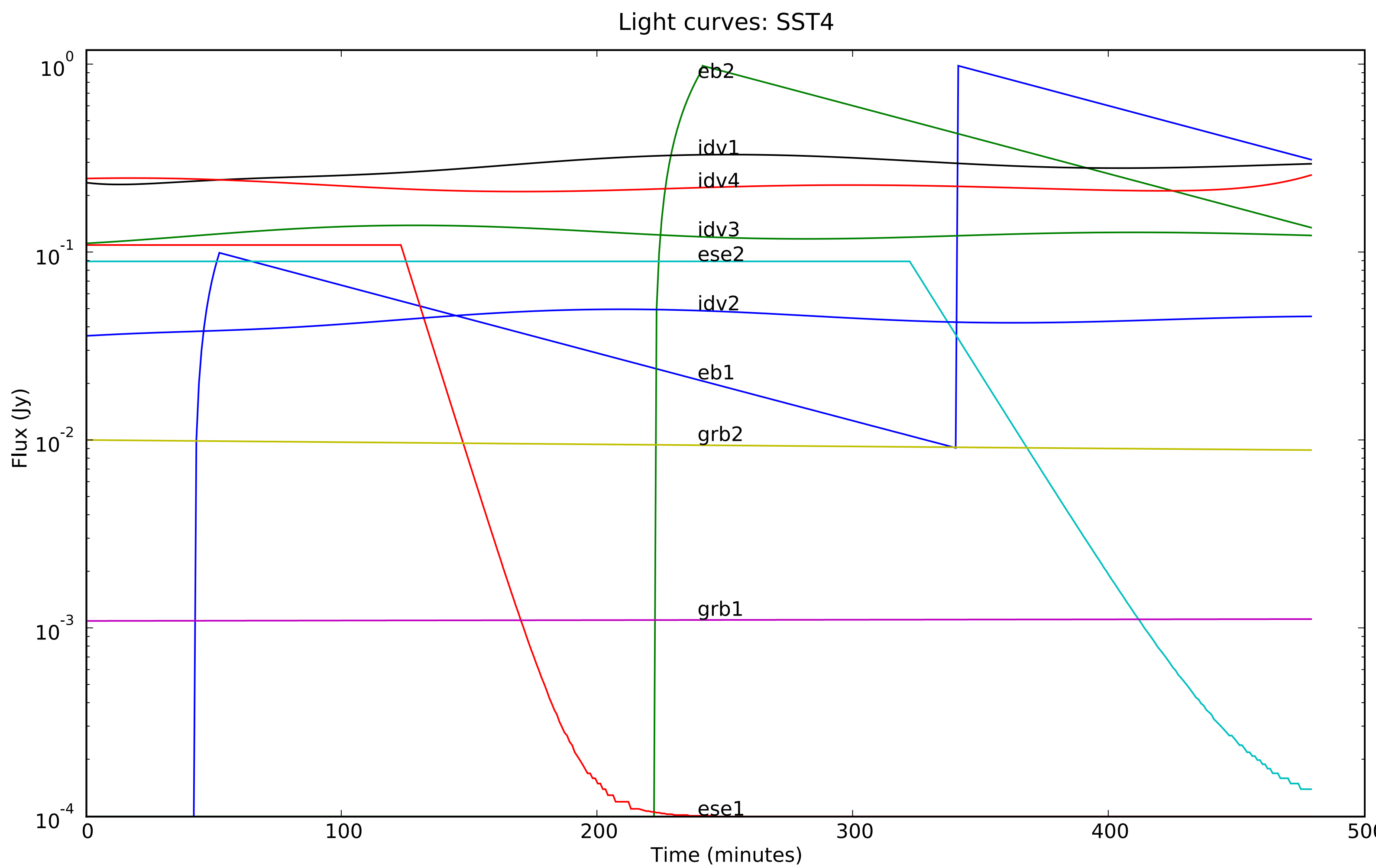
- SKADS sky model + transients from VAST
- 2km configuration
- +/- 4 hours
- 128 2.53 MHz channels
- 20s integration
- 32 PAF beams separated by 0.5deg
- 11 uJy/beam noise
- AProjectWStack
- No wplanes in simulation or imaging
- 30” resolution
- Made possible by pre-production system loaned by Intel

- Image equalised
- May 25, 2010

- Simulation not a realistic set of transient sources
- Considerable sidelobes within the primary beam
- Noise limited outside the affected primary beam

Transient light curves

- Type of sources
  - Extreme bursters (rare and very troublesome)
  - Extreme Scattering Events (rare and troublesome)
  - Gamma Ray Burster (common and not trouble)
  - Intra Day Variables (common and troublesome)
- Early conclusion - may have to track and remove worst IDVs
- VAST to provide more detailed information



Next simulation deliverables

- Science Integration 5 ends September 30
- VAST (Slow transients via commensal observations)
  - Scintillation of all compact sources in field
  - Detailed and realistic light curves for other transients
- WALLABY (HI survey)
  - Full resolution (6km)
  - Limited field of view (~ 2 by 2 degrees)
  - ~ 300 channels
  - Extended sources
- DINGO (Deep HI point-and-stare)
  - Fields near dec=0

End of November….

- FLASH (HI absorption)
  - Low frequency (down to 700MHz)
  - Absorption against background sourcs
- EMU
  - Spectral indexes without and with correction (MFS)
- DINGO
  - Detection of nodes in cosmic web

Lessons learned from simulations

- Simulations/data challenges have many benefits
  - Understanding basics of telescope performance
  - Driving software to higher performance
  - Testing computer architectures at scale
- Constant battle to constrain memory use
  - Buffering to disk not an option
- MSClean has good performance but very large memory use
  - Looking at improvements/alternatives
- Single threaded MSClean must be distributed
  - NCI National Facility expected > 75% utilisation
  - Currently working on this

- Plan to continue at increasing scale and complexity throughout development and commissioning of ASKAP

Processing for next generation telescopes

- **Physics + algorithms + operational model + software + hardware**
  - All can be changed to some degree as needed
- **Physics**
  - Requires Fourier transform + Fresnel transform from data space into image space
  - Limited number of tradeoffs available in telescope design
- **Algorithms**
  - Multiple algorithms can implement the same physics ☺
  - Mostly high data flow/low computational complexity ☹
- **Operational model**
  - All data must be processed in ~ real time, no buffering possible?
  - ASKAP uses barriers on multiple steps on all data
  - SKA can afford to drop some fraction of data
- **Software**
  - ASKAP is currently C++/MPI
  - All synthesis code re-written from scratch
  - International Exascale Software Project
- **Hardware**
  - Most processing is currently low computational complexity
  - Pay attention to data flow at all levels of hierarchy

ATNF/ASKAP
Tim Cornwell
ASKAP Computing Project Lead

Phone: +61 2 9372 4261
Email: tim.cornwell@csiro.au
Web: www.atnf.csiro.au

www.csiro.au

Thank you