ASKAP Science Processing

ASKAP-SW-0020

Version: 3.0
Date: 26/02/2016
Project: ASKAP

Prepared by: Tim Cornwell, Ben Humphreys, Emil Lenc, Maxim Voronkov, Matthew Whiting,
Daniel Mitchell, Stephen Ord, Daniel Collins
Reviewed by: Matthew Whiting, Daniel Mitchell, Stephen Ord, Daniel Collins, Juan-Carlos Guzman
Review reference : https://jira.csiro.au/browse/ASKAPSDP-1786
Approved by: Juan-Carlos Guzman Date: 04/03/2016

Keywords: computing, science, processing

Change Record

Issue	Date	Author	Sections/Pages Affected	Remarks
0.1	10/02/2011	T.J. Cornwell	All	Initial version
1.0	28/02/2011	All authors	All	Gold version
2.0	20/12/2011	All authors	All	Second release, responding to feedback from the Survey Science Teams
3.0	26/02/2016	Matthew Whiting, Daniel Mitchell, Stephen Ord, Daniel Collins	All	For ASKAP SDP Production Readiness Review

Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 Copyright

1.2 Important Disclaimer

1.3 Summary and Scope

This document describes all science-processing tasks represented in the ASKAP specifications for which ASKAP has prime responsibility. The emphasis is on a complete and accurate description of the processing steps. It is acknowledged that some areas still require further research or design, and consequently this document will be revised regularly as the processing evolves to the final state.

This document describes the operation of a fully-developed system, operating at the point where all requirements are met. The time-scales for the implementation of different features are not discussed.

This document is written primarily for the ASKAP Survey Science Teams (SSTs). It is expected that readers have a broad understanding of radio interferometry at the post-graduate level or above. Some sections are highly detailed and may be skipped at first read if desired.

The processing of high-time resolution and VLBI observations is outside of the ASKAP specifications and is not discussed.

API and user documentation for the ASKAP Central Processing tools and pipelines is out of scope for this document.

1.4 Status

The third release of this document has been produced for the March 2016 ASKAP Central Processing Production Readiness Review.

This document reflects our understanding as of early 2016. Some areas of uncertainty remain and these are indicated in the relevant sections.

This document will be revised regularly.

1.5 Glossary

Acronym	Definition
AAF	Anti-aliasing filter
ASKAP	Australian SKA Pathfinder
ASKAPsoft	Software for the Australian SKA Pathfinder
ATCA	Australia Telescope Compact Array
CASS	CSIRO Astronomy and Space Science
CSIRO	Commonwealth Industrial and Scientific Research Organisation
CP	Central Processor
FFT/IFFT	Forward/Inverse Fast Fourier Transform
FIRST	Faint Images of the Radio Sky at Twenty-cm (sky survey)
GCF	Gridding convolution function (including both image plane effects and anti-aliasing filter)
GSM	Global Sky Model
HSM	Hierarchical Storage Management
Linear Mosaic	Least squares combination of residual images
LSM	Local Sky Model
MADFM	Median Absolute Deviation from the Median
MFS	Multi-Frequency Synthesis
MRO	Murchison Radio Observatory
MSMFS	Multi-Scale Multi-Frequency Synthesis
NRAO	National Radio Astronomy Observatory
NVSS	NRAO VLA Sky Survey
PAF	Phased Array Feed
PSF	Point Spread Function
RFI	Radio Frequency Interference
SIAP	Simple Image Access Protocol
SKA	Square Kilometre Array
SMMT	Spatially multiplexed mosaicking telescope
SNR	Signal-to-Noise Ratio
SST	Survey Science Team
SUMSS	Sydney University Molonglo Sky Survey
TAP	Table Access Protocol
TBD	To be decided
TMMT	Temporally multiplexed mosaicking telescope
TOS	Telescope Operating System
VLA	Very Large Array
VLBI	Very-long-baseline Interferometry
VO	Virtual Observatory
WSRT	Westerbork Synthesis Radio Telescope

2 ASKAP CONCEPTUAL TELESCOPE MODEL

2.1 Overview

The Australian Square Kilometre Array Pathfinder is a wide field of view radio synthesis telescope designed to optimise survey speed. The majority of observing with ASKAP will be on a number of known survey projects.

The specification [12] of ASKAP is given in Table 1.

Table 1: ASKAP specification

Number of dishes	36
Dish diameter	12 m
Max baseline	6 km (30 dishes inside 2 km)
Resolution	10”
Sensitivity ($A_e/T_{\rm sys}$)	65 m^2 K^{-1}
Survey Speed	1.3×10^5 m^4 K^{-2} deg^2
Observing frequency	700 - 1800 MHz
Field of View	30 deg^2
Processed Bandwidth	300 MHz
Channels	16200
Correlator integration time	5 seconds
Focal Plane Phased Array	188 elements
Digitisation levels	12 bits
Dynamic range	50 dB

Fast survey speed requires two things - a wide field of view, and the ability to process large volumes of measured data.

- The large field of view is achieved in ASKAP by a novel technology - phased arrays at the focus of the antennas. The signals measured by the phased array are amplified, digitised, and combined to produce multiple beams on the sky. The combined signals are then correlated with signals from the corresponding beam on all other antennas. The net result can be conceived of as a large number of conventional radio synthesis arrays running simultaneously.
- To keep up with the large data rate, all processing steps from the output of the correlator to science qualified images, spectra, and catalogues are performed in automated pipelines running on a highly distributed parallel processing computer. These steps include flagging bad data, calibration, imaging, source-finding, and archiving.

The primary purpose of this document is to describe each of the pipeline processing steps in detail. Much of the processing will be familiar to those with experience manually processing data from radio synthesis arrays. However, there are some novel approaches required to deal with the large data volume, the novel aspects of the telescope such as the phased array feeds (PAFs), and the very demanding science requirements such as high dynamic range across the full field of view.

In this section, a high level, conceptual description of our model for ASKAP and the required science processing is provided.

2.2 Overview of ASKAP

The key physical locations in ASKAP are as follows:

- Observing takes place at the telescope located at Murchison Radio Observatory.
- Data are transmitted over a dedicated link to the Pawsey High Performance Computing Centre for SKA Science located in Perth.
- Science processing occurs on the Galaxy supercomputer located at the Pawsey Centre.
- The science products are archived to the CSIRO ASKAP Science Data Archive (CASDA), which is also located at the Pawsey Centre.
- Observations are monitored from the Science Operations Centre in Marsfield, Sydney.
- Technical operations are headquartered in Geraldton.

The system architecture of ASKAP is shown in figure 1. For clarity, only one antenna is shown - all antennas have the same connectivity to the MRO building. The Telescope Operating System, which is described below, is connected to all elements of the MRO architecture, providing control and monitoring.

2.3 Data flow

ASKAP is fundamentally a real-time telescope - the data volumes require that the processing occur promptly. The data flow is represented in Figure 2: Each antenna sends the signals from 188 PAF elements to the MRO processing building for digitisation and filtering to 1MHz wide channels across the observing band. These are sent to the beamformer for summation into primary beams. All primary beams from each antenna are then sent to the correlator, where all correlations between equivalent beams are calculated. The visibilities are then sent to the off-site Central Processor, located at the Pawsey Centre in Perth, WA, for processing into science results. The data rate from the correlator to the CP is 2.5 GB s^{-1}, about 70 TB in 8 hours, 70 PB in 1 year.

The data sizes preclude traditional off-line processing of the visibility data. For a fast file system of 10 GB s^{-1}, reading and writing the visibility data of 80 TB takes about 8000 s. Writing two spectral cubes (images and sensitivity = 58 TB) to the archive over the same fast file system would require 440 s for 30 arcsec resolution, and 5700 s for 10 arcsec resolution. If both occur at the same time, then the total I/O time for an 8 hour observation at 10 arcsec resolution would be 3.8 hours – about half the observing time. Thus source finding is performed within the data processing pipeline, before the images are written to storage.

The storage sizes are very substantial as well. For the 30 arcsec resolution, we require 1.1 PB for a full continuum sky survey, and 6 PB for the full spectral line sky survey with only two polarisation products kept. The 10 arcsec resolution images are roughly 10 times larger. Thus the worst case, storing the full 10 arcsec spectral line survey with full polarisation, would require 120 PB. The data from the telescope is processed as it is collected. The data products from the telescope are classified as in Figure 3.

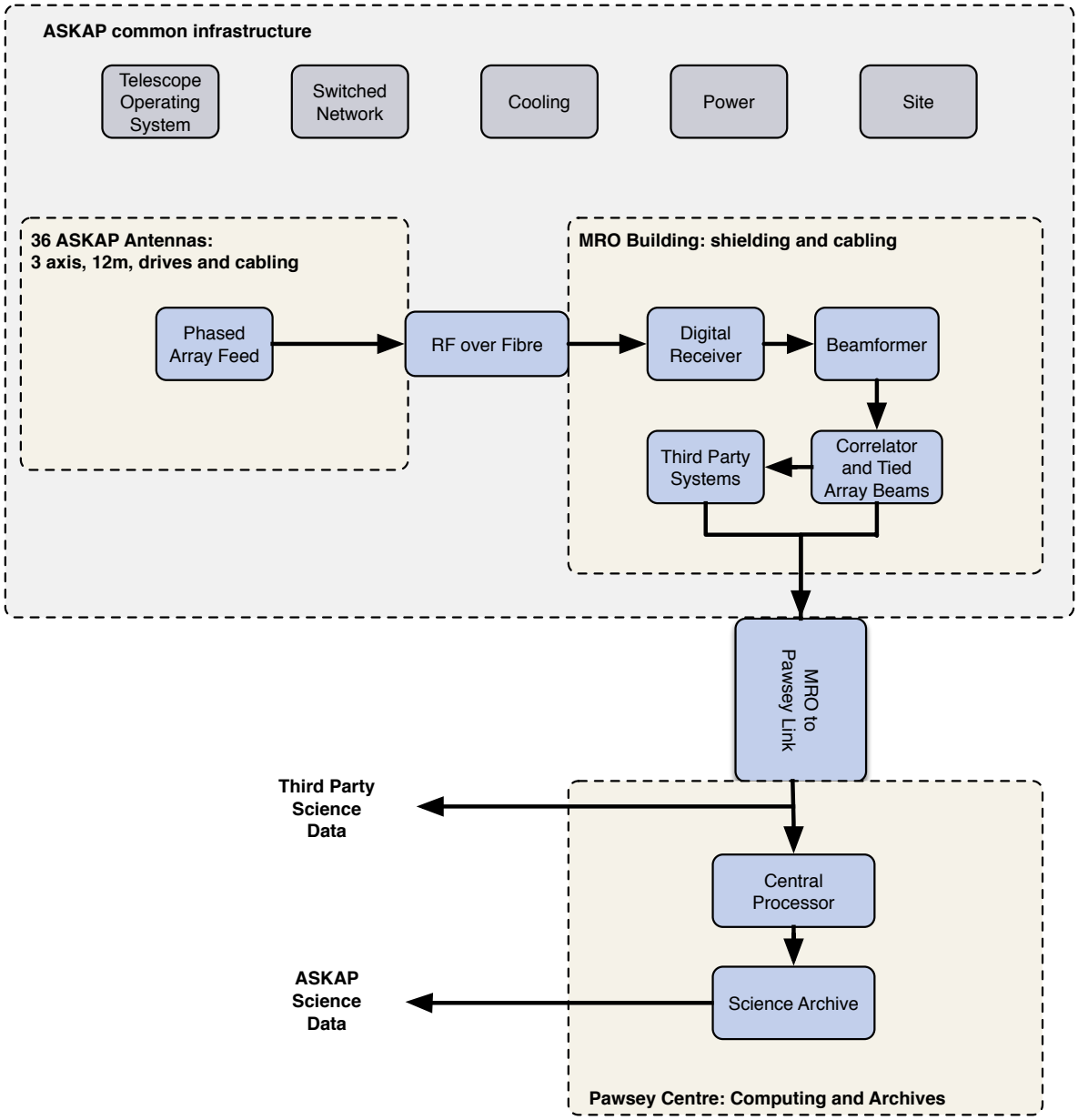


Figure 1: ASKAP System Architecture (adapted from [12])

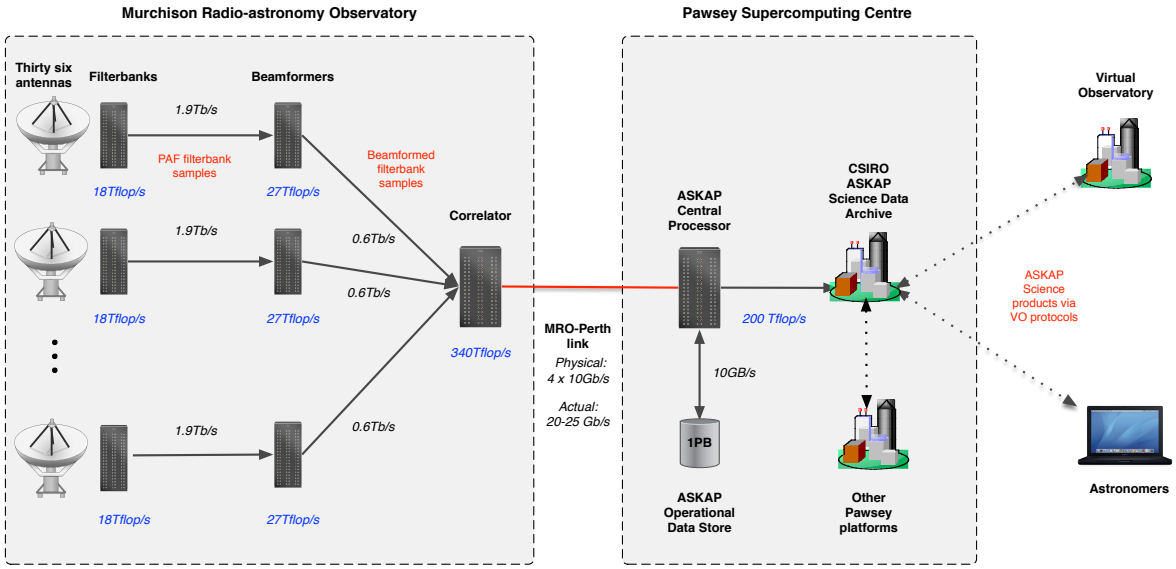


Figure 2: Data flow for ASKAP

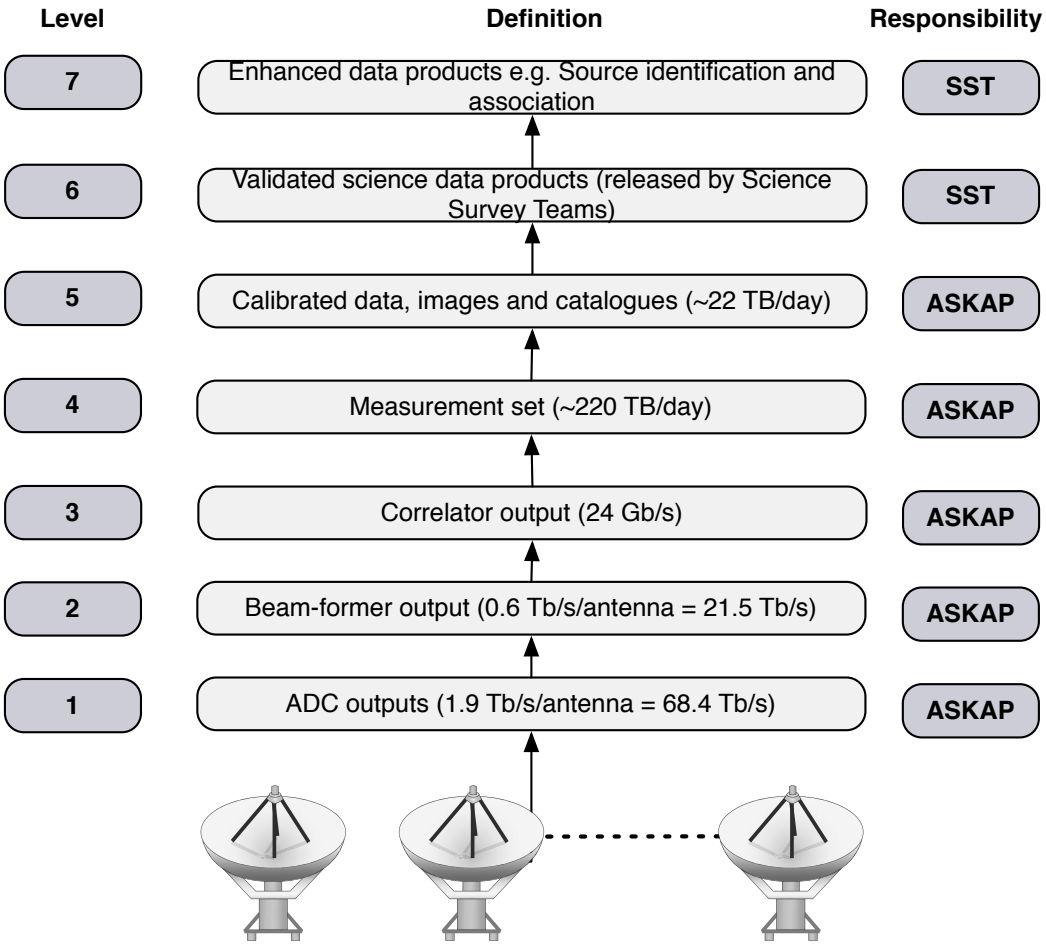


Figure 3: Classification of data products.

2.4 Overview of observing

ASKAP has been designed from the beginning to conduct surveys. By the ASKAP User Policy, at least 75% of the observing time is dedicated to surveys. For surveys, the implication is that we know and can record the intent and context of any given observation, and that similar observations have been processed before.

ASKAP will operate between 0.7 and 1.8 GHz on baselines no longer than 6 km. Calibration in this regime is quite straightforward – both the troposphere and ionosphere are relatively benign compared to higher and lower frequencies. Based on experience with the Australia Telescope Compact Array, we expect phase calibration of the atmosphere to be required no more than a few times an hour.

The radio sky is bright at these frequencies. Due simply to cosmologically distant radio sources, a typical ASKAP field will have about 50 Jy of flux in compact or slightly resolved sources. The entire continuum sky can be observed to about 1 mJy beam^{-1} within one day. Hence we can assume that we have an accurate model of the sky, a global sky model, for all subsequent observations. This simplifies the processing substantially.

Although single dominant sources are rare, and will be visible in just one primary beam, we nearly always have enough information and SNR in the entire field to calibrate the antenna/beam gains on a 60 second time scale. Calibration of the G term (broadband complex gains) is calculated on this time scale and fed back to correct the array in real time. The calibration tables can be inserted at either of two points: at the beamformer via modifications to the beamformer weights, or at the data ingest point. The former is mandatory if the tied array beams are used.

The correlator only calculates correlations between phased array beams (rather than the PAF elements). At the calibration and imaging steps, we cannot straightforwardly determine the element gains. Instead the element gains are estimated at the beamformer by correlating signals from radiators on the surface with the signals received by the elements. Our current assumption is that this form of calibration is sufficient to make the element gains consistent, and the overall complex gain can be derived from the regular astronomical calibration.

ASKAP is a Spatially Multiplexed Mosaicing Telescope (SMMT). Most radio telescopes that perform mosaicing are time multiplexed – meaning that the observations of different regions (pointings) are taken at different times. For ASKAP, all the field of view is observed at one time – by forming multiple beams spanning the object using a phased array feed.

Mosaicing theory has been worked out by a number of authors. The key results for a Time Multiplexed Mosaicing Telescope (TMMT) are:

1. A large field of view may be imaged using a number of interferometric observations spanning the object.
2. To prevent aliasing of spatial structure from one spatial frequency to another, the field of view must be sampled at roughly $\lambda/(2D)$ radians, where D is the antenna diameter.
3. The sky brightness may be estimated using linear and non-linear algorithms.
4. For imaging objects larger than the antenna beam, a single dish image will be necessary.
5. Mosaicing with different antenna primary beams is conceptually straightforward but can be computationally expensive.
6. The dynamic range of mosaic images can be limited by a number of effects, overall array calibration, feed illumination, knowledge of the antenna primary beam, surface and pointing accuracy.

All of these results hold for a SMMT such as ASKAP, though the details may vary.

Since the telescope is calibrated quite accurately at all times, and there is a model of the sky, we can always remove the predicted visibility before any deconvolution. All subsequent images are then differential only. The differential image may still require cleaning but the amount of computation required is lessened considerably. Following processing, the sky model can optionally be added back to the differential image for science analysis.

Data re-weighting (e.g. uniform weighting and visibility tapering) cannot be performed as usual since the extra pass through the observations for spectral line imaging is not affordable. Instead, we use pre-conditioning of the dirty image and point spread function to apply the weighting after the visibilities are gridded. Since the long integration Fourier plane coverage for the 2 km configuration is so good, all the pre-conditioning has to do is to even out the bumps in sensitivity in the Fourier plane. We accomplish this by robust weighting to remove the bumps, and tapering to restore the overall Gaussian shape. This loses a small amount of SNR and reduces the close-in side lobes by up to an order of magnitude.

Each deconvolution cycle requires significant processing. As a consequence, the deconvolution is designed to require little if any special processing. The multi-scale algorithm estimates the sky on a number of scales at once. Consequently smoothing the deconvolved image will accurately recover the full flux on scales significantly larger than the synthesised beam. The multi-scale algorithm also provides a list of candidate objects for use in the source finding.

One unusual aspect of the deconvolution is that the signal-to-noise varies across the field, being roughly constant within the PAF envelope, and tending to zero away from the envelope. Sources outside the PAF envelope can be deconvolved given sufficient SNR. Hence the deconvolution must explicitly take into account the SNR.

Following the imaging, the images and spectral cubes are passed to the source finder. The output from the source finder is archived. An official catalogue release is prepared by the Science Survey Teams.

2.5 ASKAP Array Configuration

The array configuration was chosen with three main goals [10, 11]:

- A highly compact configuration (baselines less than 300 m): this is necessary for low surface brightness science (Galactic science and the cosmic web) and for pulsar surveys.
- A medium configuration (baselines less than 2 km): this is optimised for surveys for extragalactic HI in emission. It also serves well for polarization and transient source surveys.
- A long-baseline configuration (baselines to 6 km): this is optimised for high resolution continuum surveys. It overcomes the confusion limit at 10 μJy rms, whilst retaining some surface brightness sensitivity to extended objects.

The chosen configuration (see Figure 4) meets these three goals. There is a 2 km-scale configuration with 30 antennas, and an extra 6 on baselines up to 6 km. One antenna close to the centre of the array has been moved slightly from the originally-published configuration to improve the sampling on baselines less than 300 m.

Imaging at full spectral and angular resolution is out of scope for ASKAP given the high computing requirements. For observing with a high end frequency at 1.4GHz, full spectral resolution imaging is only

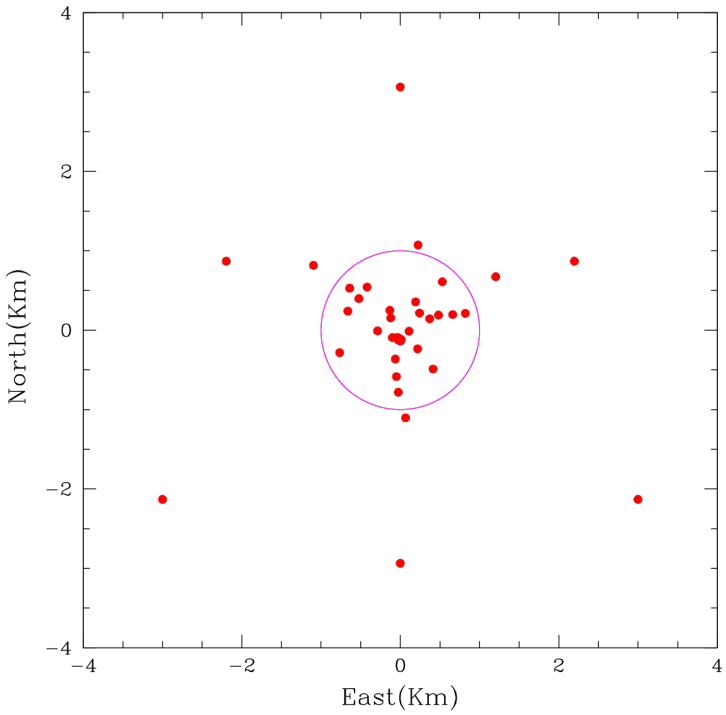


Figure 4: ASKAP Array Configuration. The circle radius is 1 km. The inner 30 antennas are placed to provide a high quality naturally weighted synthesised beam at 30 arcsec resolution. The outer antennas are placed to provide 10 arcsec resolution.

possible for baselines with length less than 2km. Note that this does include some baselines to the outer six antennas.

2.6 ASKAP Antenna Design

The dynamic range specification for ASKAP, 50 dB, is demanding, especially given the wide field of view. Similar values have been reached with the VLA and WSRT but only for a single dominant source centred at the delay-tracking centre (which is usually chosen to be the same as the antenna pointing direction). With such a specially chosen source configuration, many effects that limit dynamic range become second order. For a wide field of view telescope such as ASKAP (and SKA), bright sources will necessarily lie at random locations, meaning that such effects become first order. Stable, well-understood and accurately calibrated systems are then necessary to reach the dynamic range goals. For a system with a phased array feed, there are an increased number of possible causes of instability – there are many more active elements (such as amplifiers), and the entire system is liable to rotate with respect to the sky as the antenna parallactic angle changes. With this in mind, ASKAP has adopted "sky-mount" antennas, in which the antenna is fixed with respect to the sky. An equatorial mount provides one example of a sky-mount antenna but there are other options. ASKAP uses an altitude-azimuth mount with the antenna surface and feed leg assembly rotating around the optical axis. This three-axis system fixes the antenna sidelobes and PAF with respect to the sky.

2.7 ASKAP Phased Array Feeds

The wide field of view of ASKAP is obtained by using multiple feeds in the focal plane. Conventional horn feeds can be used for this purpose but the allowed spacing is limited by the requirement that each horn illuminate the primary or secondary reflector with good efficiency. Usually this limits the spacing to 2-3 wavelengths or more. For closer spacing, virtual horns must be used. Summing the outputs of multiple smaller feeds, none of which by itself can illuminate the antenna adequately, forms a virtual horn. After summation, the virtual horn can have very high efficiency in illuminating the reflector.

The weighted sum may be formed to obtain high efficiency (as needed for the most sensitive search for spectral line emission) or to minimize the exterior sidelobes (as needed for imaging the bright continuum sky). The ASKAP Phased Array Feed (PAF) comprises 188 elements distributed in a checkboard pattern over a 1.2 metre diameter circle, each including; a low noise amplifier (LNA), band defining filters and an electrical to optical convertor. The PAF covers the frequency range of 700MHz to 1800MHz. Analogue data from each of the PAF elements is transmitted via single mode fibre to the Digital Receiver. The PAF utilises RF over Fibre to send the analog signals to the MRO control building for digitisation and beamforming. The outline of the signal path is shown in Figure 7.

2.8 Beamformers and Correlator

In broad terms, in the next stage in the signal path the analogue signals from the PAF on each antenna are digitised by the digital receiver and then formed into a set of well-behaved beams on the sky, and then the beamformed signals are correlated across all antennas for each corresponding beam. This produces a set of correlation coefficients for each beam.

In more detail, the steps in the signal processing are:

1. Following the signal path in Figure 8 For each antenna:

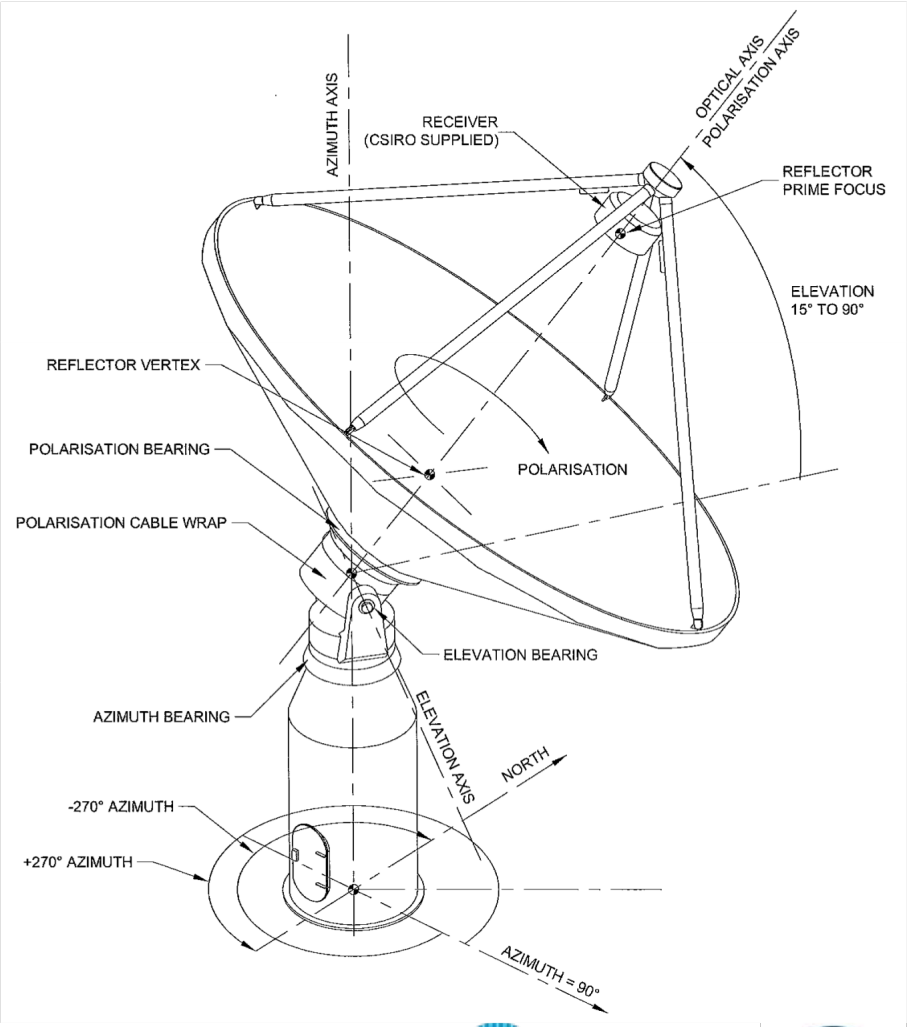


Figure 5: ASKAP Antenna Schematic. The antenna can track on three axes – the usual azimuth and elevation, and a novel, third axis, called the polarisation axis. Rotation about the third axis serves two purposes: first, the PAF is kept fixed on the sky, and second, the feed legs (and associated scattering) are also fixed with respect to the sky.

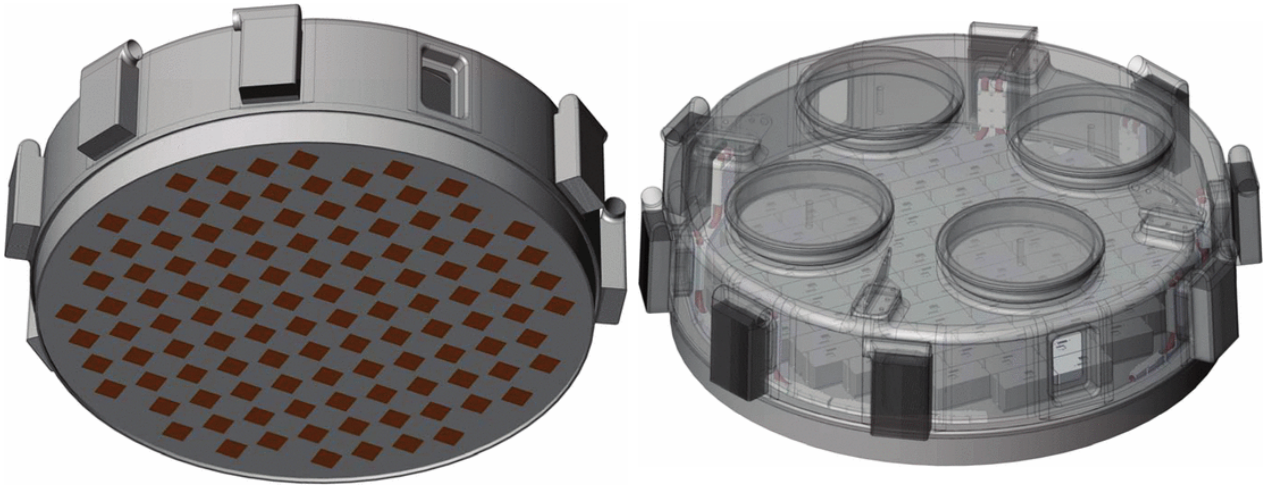


Figure 6: The MkII ASKAP PAF from below (left) showing the checkboard feed layout. The cutaway on the right shows the RF modules, called *dominoes* containing amplification and electrical to optical converters [14].

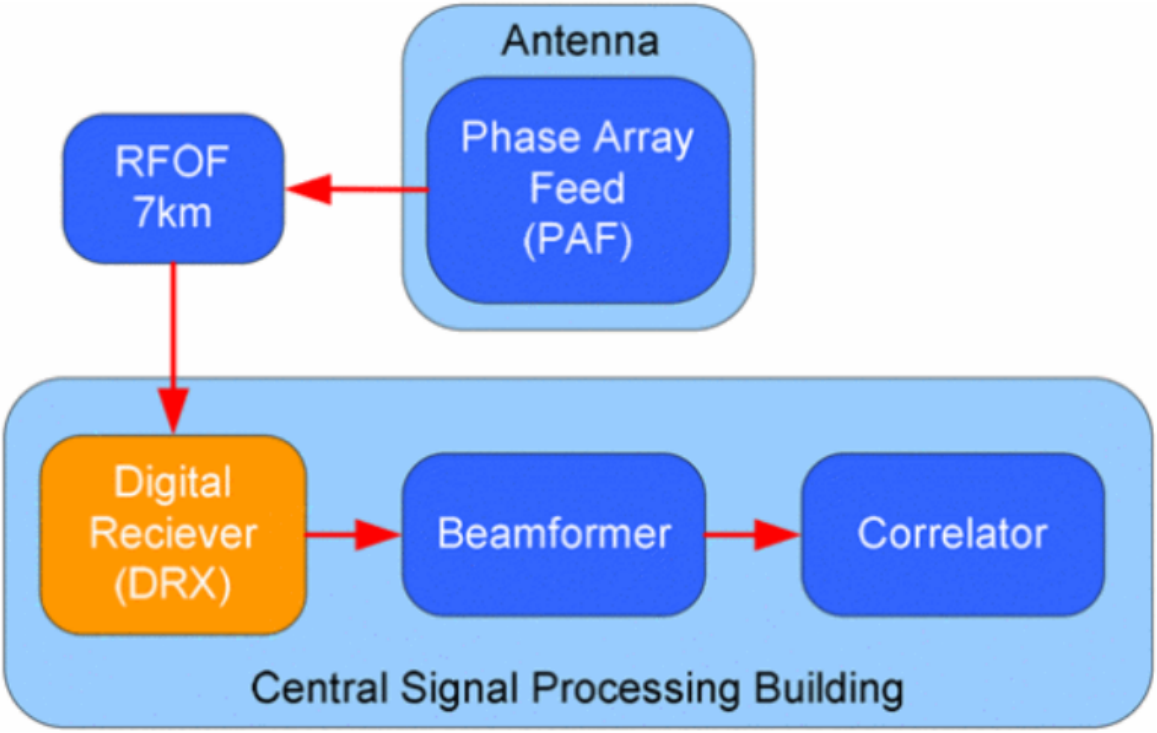


Figure 7: Schematic ASKAP signal path [13].

- The voltages at the 188 PAF elements are amplified and digitised to 12 bits
- Each PAF output is filtered by a 12 tap polyphase FIR into 384 1MHz channels each oversampled by a ratio of 32/27. The digitised samples are then sent to the beamformer.

2. The beamformer (see Figure 9) constructs PAF beams by:
   - A cross connect gathers all 192 inputs, 188 from the PAF and 4 more other sources, which may be utilised for calibration signals or RFI receptors.
   - An Array Covariance Matrix (ACM) is formed as input to the system that forms the beamformer weights.
   - 36 beams are each formed by the weighted sum of 64 PAF outputs. These can be weighted in very flexible ways so as to optimise the resulting beams according to a number of possible metrics - highest sensitivity, lowest sidelobes, nulling in particular directions.
   - Note- The weights can be set differently per antenna and per coarse channel.
   - The samples from each formed beam are then filtered to high resolution using polyphase filters. Each coarse channel is divided into 54, 18kHz fine channels
3. The beamformed, finely-channelised data streams are then fringe stopped for the centre of each beam and sent to the correlator.
4. The correlator performs cross-correlations between co and cross polarised beams on different antennas, resulting in $4 \times 36 \times 36 \times (36+1)/2 = 95{,}904$ distinct correlations per fine channel.
5. The correlations are integrated for 5s and then sent to the Central Processor for conversion into science products.

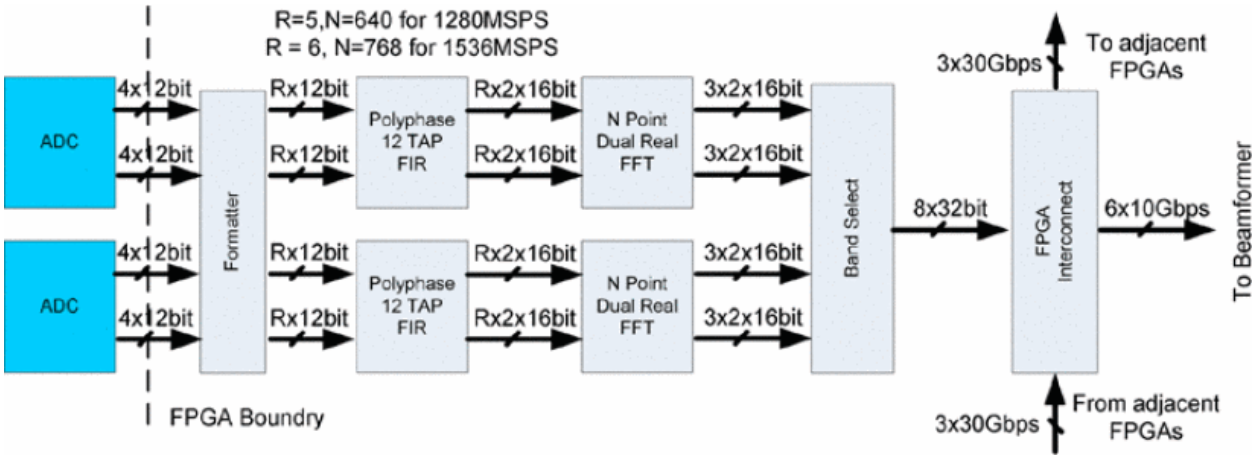


Figure 8: The ASKAP digital receiver, digitises and filters the output from the PAF and provides input to the beamformer [13].

Note that the specification for downstream processing after correlation is for 300MHz total bandwidth for a total of $54 \times 300 = 16200$ fine channels, each of width 18.518519 kHz.

The ASKAP specifications also call for a "zoom mode", with higher frequency resolution of $\sim$ 1 kHz channel width. Note that the precise width is yet to be determined, but will be within 1-10% of this figure. It is likely that $\sim$ 300 of the channels will be used for coarse channels to allow the continuum to be sampled.

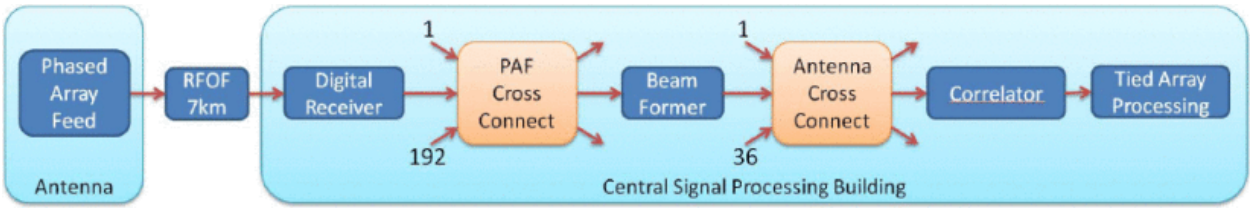


Figure 9: The signal path through the beamformer [14].

The 300MHz bandwidth will be able to be broken into discontinuous sub-bands, centred at different frequencies to allow coverage of different spectral lines (for instance, Galactic HI and OH). Note that accurate Doppler correction (Section 3.9.1) will be crucial for this mode of operation.

The capabilities of the beamformer and the correlator are limited by cost. The beamformer is limited in the number of beams produced and the number of PAF elements summed into each beam (roughly 60). The correlator is limited to 36 antennas and 36 input beams.

2.8.1 Sky Sampling

The 36 phased array beams are separated by λ/D, rather than the sampling $\lambda/(2D)$ required to image without any aliasing. This occurs because the voltages are correlated on the λ/D scale. As a result there are ripples in the sensitivity pattern (see figure 10).

In theory, the voltages can be reconstructed on the $\lambda/(2D)$ grid by sinc-interpolation. However the correlator has insufficient capacity to perform the necessary calculations. There are two distinct deleterious effects arising:

- The sensitivity possesses ripples at the 20-30% level.
- The reconstruction of structure in complex or extended fields could be in error because of the aliasing.

In addition, the sensitivity inevitably rolls off at the edge of the field due to loss of gain off-axis. Flattening the sensitivity is possible using 16 way interlacing - 4 fine scales to remove the variations at $\lambda/2D$ and 4 coarse scales to remove the roll-off in sensitivity at the edge of the PAF. Figure 11 shows the results of 4 way and 16 way interlacing. Such an approach would also prevent any aliasing of large scale structure.

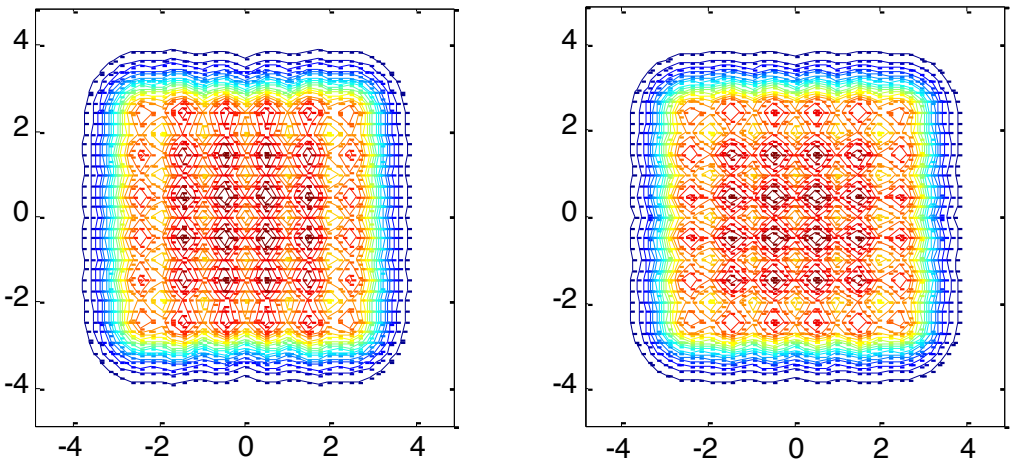


Figure 10: Sensitivity pattern for (left) one polarisation, (right) Stokes I. 5% contours. (From John Bunton and Stuart Hay).

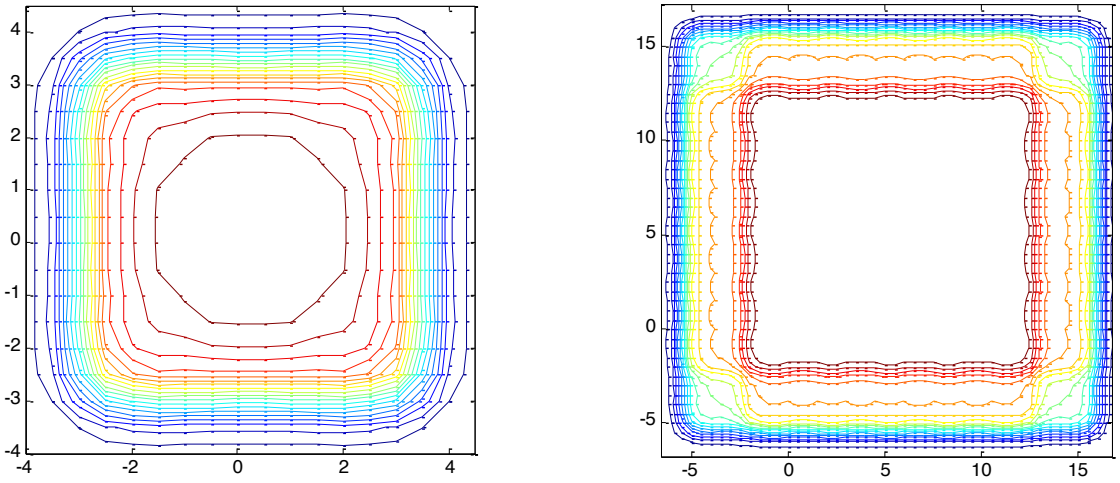


Figure 11: Sensitivity after (left) 4-way interlace, (right) 16-way interlace. 5% contours. (From John Bunton and Stuart Hay).

2.9 Computing Overview

The ASKAP architecture describes three computing components, the *telescope operating system (TOS)*, the *central processor (CP)*, and the *CSIRO ASKAP science data archive (CASDA)*. The telescope operating system is assigned the responsibilities relating to monitoring and control of the physical instrument. This includes antennas, beamformers, correlator and various other hardware sub-systems. The central processor, described further in section 2.11, is assigned the responsibilities of data processing and calibration of the instrument. The central processor, or rather the processing steps carried out within it, is the primary focus of this document. Finally, the science archive, CASDA, is assigned the responsibilities of permanent data archive and providing user access to this data.

The software components and processing pipelines that comprise the CP software are also known collectively as *ASKAPsoft*.

The telescope operating system and central processor are the computing subsystems responsible for the acquisition of data and initial online processing. These subsystems are further decomposed into the components illustrated in figure 12.

ASKAPsoft makes extensive use of third party libraries such as EPICS and casacore. For the most part, the science processing capabilities described in this memo have been written specifically for ASKAP.

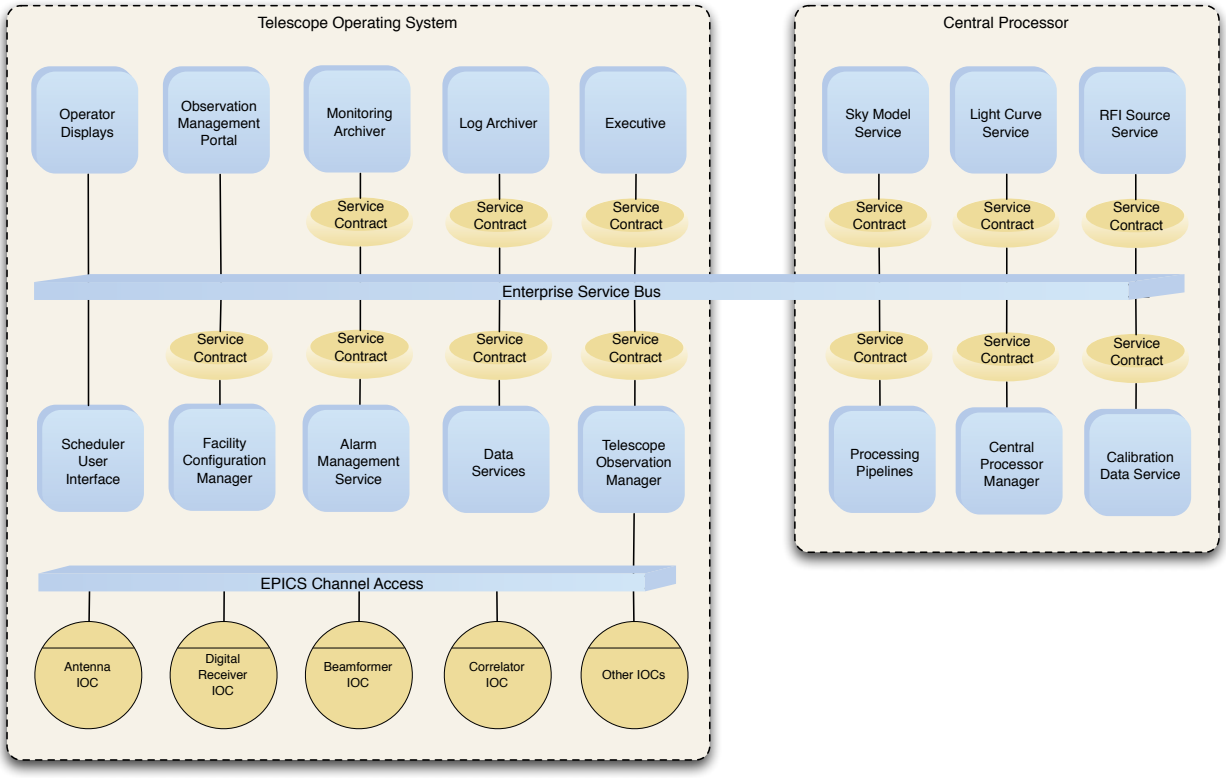


Figure 12: ASKAP Computing Architecture - Shows the components of the telescope operating system and central processor sub-systems.

The top-level components of the telescope operating system and the central processor and their responsibilities are:

- Executive - responsible for orchestrating an observation,
- Telescope Observation Manager (TOM) - responsible for monitor and control of the physical telescope hardware subsystems such as the antennas, beamformers, correlators, etc.,
- Processing Pipelines - are the automated data reduction and analysis pipelines. See section 2.12,
- Data Services - are responsible for permanent and temporary storage of data shared by the components of the online system,
- Sky Model Service - responsible for managing and providing access to the Global Sky Model (GSM), an all sky database including flux measurements, polarisation information and spectral indices,
- Light Curve Service - provides storage, management and retrieval of this light curve data,
- RFI Source Service - responsible for managing and providing access to a database of known RFI sources that may impact ASKAP observations,
- Calibration Data Service - responsible for storage and management of calibration parameters,
- Monitoring Archiver - responsible for collecting and archiving monitoring data from subsystems,
- Log Archiver - responsible for collecting and archiving of log messages generated by the subsystems,
- Alarm Management Service - responsible for managing/escalating alarm conditions in the system,
- Facility Configuration Manager - responsible for hosting and maintaining configuration data for hardware and software sub-systems,
- Scheduler User Interface - responsible for scheduling the scheduling blocks for execution by the executive,
- Operator Displays - responsible for presenting a user interface for control and monitoring of the instrument by an operator,
- Observation Management Portal - responsible for presenting a user interface for the input, modification and monitoring of observing programs by the astronomer or Science Survey Team.

ASKAPsoft will not be released publicly, for several reasons: it is in development, and will remain so for some time; much of the software will be highly specialised for the central processor or other ASKAP systems; and we do not have the resources to provide user support.

2.10 Observing with ASKAP

Usage of ASKAP is through serviced observing and the *Observation Management Portal (OMP)*. The OMP is a web-based user interface which allows the creation, submission, monitoring, and control of observations through the creation and management of *scheduling blocks*. Given ASKAP is primarily designed as a survey instrument, the OMP will be optimised for the construction and management of surveys involving observations of a large fraction of the sky.

Key functions of the OMP are:

- Create observations (scheduling blocks)

- Submit observations
- Monitor the state of observations
- Modify or delete observations (subject to restrictions imposed depending on what state the scheduling block is in)

The scheduling block is the primary object of execution for the online ASKAP system. It is a rather complex entity containing all the information necessary to execute sequentially and without interruption (except for exceptions/errors) an observation. This includes both the acquisition of visibilities and the data processing. After submission, these scheduling blocks are executed in a semi-autonomous manner under the supervision of an operator.

Essentially the scheduling block consists of an observation procedure and a set of observation parameters. For the standard observing modes, scheduling block templates are available. These scheduling block templates are essentially tested observing modes which can be tailored (via the parameters) to suit the needs of specific observations.

Note that SSTs will not directly create or manage the Scheduling Blocks. Rather, the SSTs will submit observation details to ASKAP Operations who will then use the OMP to create and schedule the observation. This will assist with the management of commensal observing programs among other operational benefits. SSTs will be able to use the OMP to monitor the state of the scheduling blocks related to them, however will be unable to modify them directly. This is a change from previous versions of this document, where it was indicated that SSTs would have direct access to all OMP functions.

2.11 Central Processor

The Central Processor is the ASKAP subsystem responsible for transforming the output from the correlator into science data products such as images, cubes, catalogues. The central processor is both a hardware and software subsystem, consisting of a high-performance computer and a suite of synthesis imaging and analysis software. The software suite is highly optimised and parallel, designed specifically to support processing high-rate data streams in quasi realtime.

The Central Processor hardware is provided and operated by the Pawsey Centre in Perth, Western Australia. The Central Processor hardware is provided by the *Galaxy* cluster, and by 16 nodes on the *Zeus* cluster.

Galaxy, a Cray XC30 system, comprises 536 compute nodes, of which 472 are CPU nodes and 64 have been provisioned for GPU processing. Presently the ASKAP software does not utilise GPU devices, and so only the CPU nodes are described further. Each of the 472 CPU nodes contain two 10-core Intel Xeon E5-2690V2 'Ivy Bridge' processors which share a total of 64 GB of main memory. This gives a total of 9440 cores, providing a peak performance (LINPACK) of 192.1 TeraFlops, and total memory of $\sim$31 Terabytes.

The Zeus nodes are primarily used for running the Ingest pipeline, while most other ASKAP data processing tasks will execute on Galaxy. Some of the CP services with low computational overheads may run on a Zeus node reserved for the services. If so, this will be allocated from the 16 Zeus nodes.

This compute capability is supported by a 1 PB high-speed distributed file-system capable of better than simultaneous 5 GB s^{-1} read and 5 GB s^{-1} write (10 GB s^{-1} total). This filesystem is shared by the Zeus and Galaxy systems, and is primarily used for buffering visibility data during processing, and buffering images, cubes and other data products prior to archiving in CASDA.

Free space on this file system is critical to telescope operations. Insufficient free space will cause observations to fail. Two related strategies are planned to mitigate this risk. First, an automated file purging policy will be used to minimise the accumulation of old files. Second, automated file system monitoring systems will send an alert to ASKAP operators when free space falls below a critical threshold, allowing operators to free space manually (or to identify why the automated purging was inadequate).

The processing pipelines, described further in section 2.12, that run on the Central Processor are written largely in C++, building on top of the casacore [6] library, and multiple other third party libraries. The synthesis processing code has been written from scratch and does not rely on third party libraries for the synthesis-specific algorithms. However, some of the algorithms are improved versions of those used in CASA.

Whilst the central processor software supports the execution of a large number of parallel data processing pipelines (to support commensal observing strategies) available compute and storage resources impose some practical limitations. For the purpose of capacity planning a working model exists which defines three broad categories of imaging/processing pipelines: 1) A pipeline processing all 16200 channels, termed here as *Large-N*; 2) A pipeline processing an averaged set of channels less than or equal to 300, termed *Small-N*; and 3) a transient detection pipeline which further averages channels to approximately 30, requiring substantially less aggregate processing and I/O capability than either of the other two pipeline categories. In addition to these *data reduction* pipelines there exist two system pipelines, the ingest pipeline and the calibration pipeline, both described in more detail later in this document.

The standard continuum pipeline will use multi-frequency synthesis to represent the sky brightness by Taylor terms in frequency (see Section 3.8.7 for details). An alternative approach would be to deconvolve channels separately and creating a "continuum cube", most likely with all Stokes parameters. It is expected this will be possible for between $\sim$ 30 and 300 channels, depending on the processing capability (this is still to be determined at time of writing). The output of this would be suitable for Rotation Measure Synthesis - see Section 3.9.3

At the time of writing, the expected capacity of the central processor will allow execution of the following processing pipelines in parallel:

- Small-N pipeline – 10 arcsec resolution at 1.4 GHz (6 km baselines), full stokes
- Large-N pipeline – 30 arcsec resolution at 1.4 GHz (2 km baselines), single stokes
- Transient detection pipeline

As our understanding of ASKAP processing improves, the number of concurrent pipelines and their capabilities may be revised. High angular resolution Large-N is omitted from that list since the compute and memory requirements, being many times larger than imaging with the 2 km baselines, are prohibitively high.

2.12 Processing pipelines

All observed data from the correlator are processed automatically all the way through to the science archive. The processing pipelines must perform the following steps:

- Ingest visibility data from correlator, and meta data from the Telescope Operating System (TOS),
- Flag for known radio frequency interference, working from a data base of known RFI sources,

- Identify and flag (unknown) radio frequency interference, saving candidate identification, and identify and flag further bad data
- Correct for ionospheric Faraday rotation.
- Solve for calibration parameters using a least squares fit of the predicted visibility (from previously obtained model of the sky) to the observed visibility,
- Apply calibration parameters, predicting forward from the last previous solution,
- Average visibility data to required temporal and spectral resolution,
- Subtract the current sky model from the visibilities,
- Construct an image
    1. Grid the data using convolutional resampling,
    2. Fourier transform to the image plane
    3. Deconvolve point spread function (if warranted)
- Find sources in the resulting image or cube.
- Perform automated quality evaluation on the images and source catalogues.
- Save science data products to the CSIRO Science Data Archive (CASDA).

The ingest pipeline is shown in Figure 13, and the subsequent pipelines in Figure 14. The pipelines are constructed from a small number of monolithic applications: a calibrator, an imager, a source-finder and a fast ("transient") imager:

- The calibrator calculates the calibration of various terms, working from the Local Sky Model (LSM), producing calibration tables.
- The imager performs imaging, self-calibration, and peeling, starting from the local sky model, producing an update to the LSM, and calibration tables.
- The source-finder finds and fits sources in final images, producing source catalogues.
- The fast imager is a lean and fast imaging application designed specifically for transient imaging. It has fast execution and low latency.

An array calibration pipeline runs continuously, producing calibration tables that are sent to the ingest system to be applied. A transient pipeline also runs continuously, making images every integration period (5 s). The other pipelines run at the end of observing. Continuum processing runs before any spectral line imaging so that the latter can use self-calibrated calibration tables and the final Local Sky Model. The transient pipeline finds and fits time-variable sources with low latency (< 5 s after the end of an integration). These can then be used in the continuum and spectral line imaging.

Each transient source has a light curve determined by the transient pipeline. This is conceptually just a table of source brightness on a regular cadence of *5s*. During the subsequent calibration and spectral line imaging, any especially bright sources will be modelled in the model visibility.

It is recognised that, in addition to the 5 s cadence used in the transient pipeline, it will be important to search for transient and variable sources at larger cadences, enabling detection of fainter sources than is

possible on 5 s integrations. This could be accomplished in a number of ways, principally either a longer-cadence version of the transient pipeline described above, or through offline UV processing. At time of writing the implications for the processing have not been fully worked through, and so we can not give details about likely implementations.

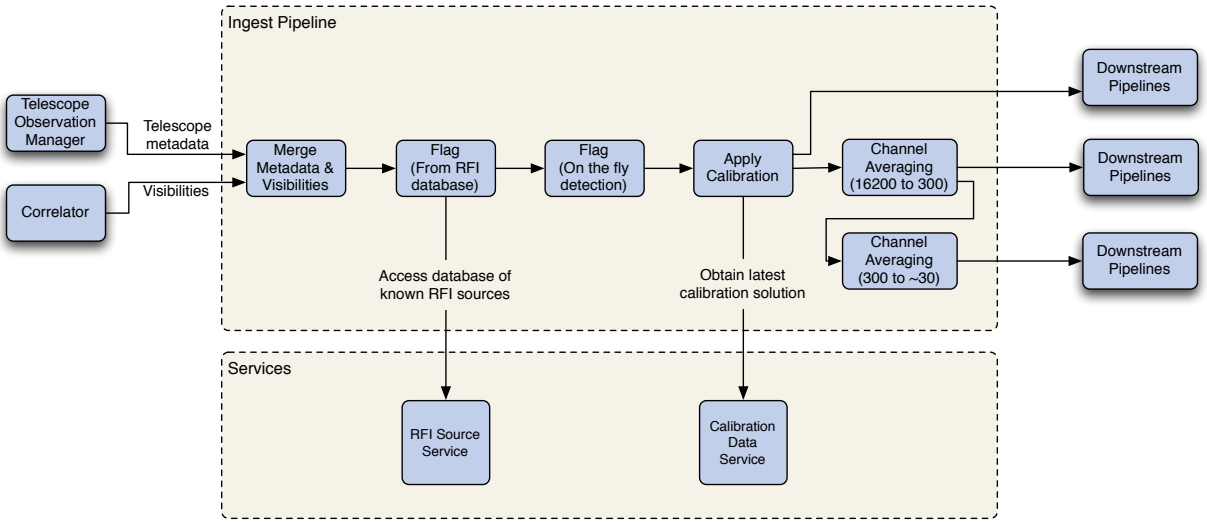


Figure 13: Schematic of the ASKAP Central Processor ingest pipeline.

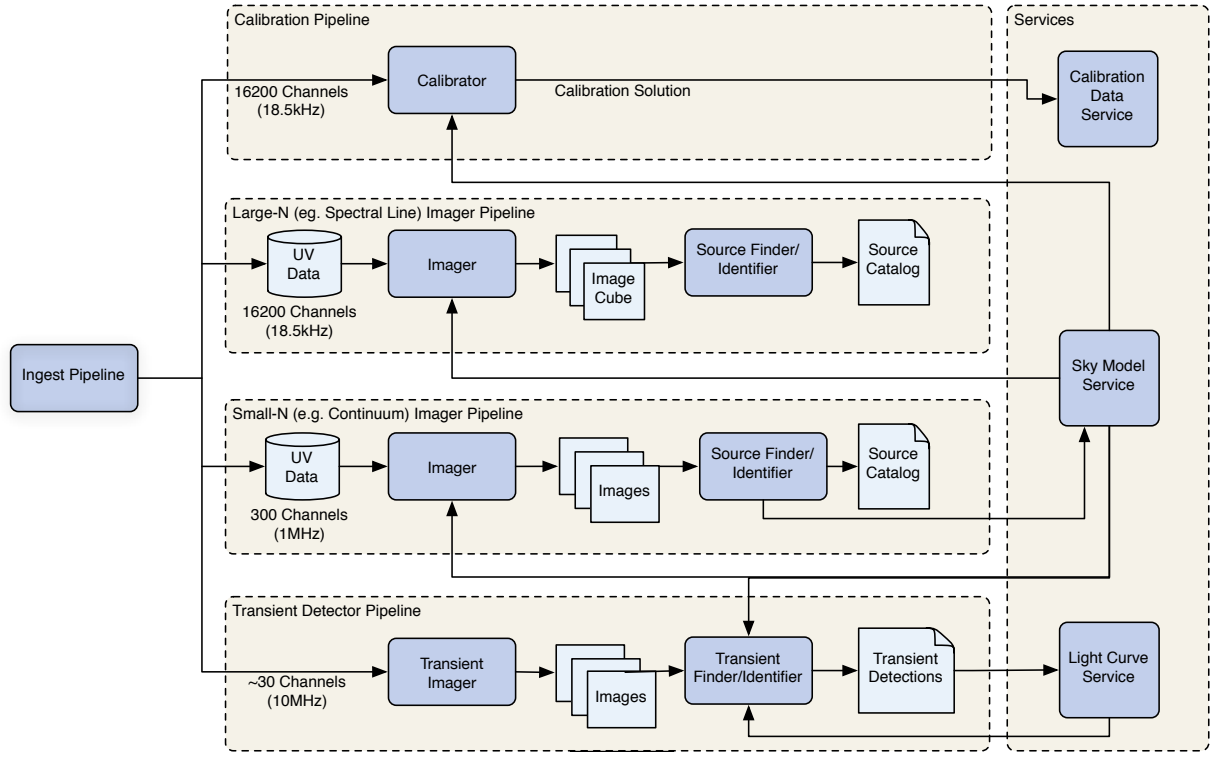


Figure 14: Schematic of ASKAP Central Processor calibration, imaging and transient detection pipelines.

3 DESCRIPTION OF PROCESSING STEPS

With the conceptual model of the telescope laid out, the processing steps can be described in detail.

3.1 Data Ingest

The visibility data from the correlator are combined with the meta-data from the TOS to provide the necessary information for subsequent processing. The data from the correlator are subject to RFI excision, flagging, calibration, and averaging. Each pipeline receives data averaged specifically for that purpose. Since the correlator stops the fringes for the nominal direction of each PAF beam, the losses due to averaging are minimised. The default averaging is 1 MHz in frequency, resulting in no more than 2% smearing in the 6 km configuration for a source two degrees from the centre of a given PAF beam.

3.2 RFI Excision

RFI excision is based on a catalogue of known sources of RFI. For example, the RFI database contains information on:

- Frequencies affected
- Expected strength, including minimum, maximum, and median
- Location of RFI
- Site when seen
- Fiducial position on Earth
- Orbital elements
- Emission times and durations (*e.g.* 6am to 12pm, WST)

Using this information, the observed frequencies, times, and direction in which a given source will be seen is calculated. The affected data are flagged appropriately (see section 3.3). The RFI database is updated regularly by operations, based on information from a variety of sources including automated flagging; a human examines statistics of the data flagging to find new RFI sources and enter them into the database.

3.3 Identification and removal of bad data

Bad visibility samples are found and flagged using a number of tuneable heuristics. The following are available, however only a subset of these will be possible on-the-fly (dynamic statistical thresholds in particular will not be feasible):

- Standard interferometric selections (baselines, frequencies, beams, etc.)
- Field elevation outside of range
- XX, YY, XY, YX outside of range

- Maximum absolute value of Stokes V exceeds some threshold

The thresholds used in the last two techniques can be defined by the user, or they can be set to some multiple of the local standard deviation (e.g. 3 or 5 sigma). The mean and standard deviation are estimated using robust statistics (i.e. quartiles rather than moments), to avoid biases from a small number of strong interference peaks.

Other heuristics could be added to the flagger if needed, for instance

- Maximum absolute change in XX, YY, XY, YX exceeds some threshold
- Deviation from fit in frequency exceeds some threshold

Since we possess an accurate model of the sky, we could use the statistical operators on the residual visibilities, if that is deemed necessary.

Autocorrelations and the TOS monitor information are also used to flag the data. The monitoring archiver is queried for relevant events and the data are flagged appropriately. In the event that new information becomes available, such as cross correlations with a reference horn pointing to a specific radiator, these can be incorporated into the system. Whether they are incorporated into the database, the TOS monitor information, or the automated flagging depends on the specific setup.

3.4 The Sky Model

The Global Sky Model (GSM) encapsulates our knowledge of the sky over the ASKAP frequency range. It will be used to derive accurate model visibilities for calibration and continuum removal. The GSM will be initialised from the ATCA calibrator list and other relevant sources such as SUMSS and NVSS.

The GSM will be updated following observations of a given field. Eventually the GSM will contain all sources relevant to ASKAP observations. For a threshold of 1mJy, there will be approximately 3 million sources over 3 π steradians, 90 per square degree, 5400 per ASKAP field of view.

The GSM will be composed from multiple objects:

- Discrete component listings, including strength, position, shape information, spectral indices, polarisation, Faraday rotation. We derive these by fitting to the residual image (using *e.g.* Duchamp), by deconvolution (*e.g.* by multi-frequency multi-scale CLEAN), or by shapelet analysis, *etc.*.
- Image cutouts, including position, spectral indices, polarisation, and Faraday rotation. These are derived from previous observing, or from single dish observations. The cutouts have no special limit on angular size or frequency range.

Prior to the initiation of observations of a given field, a Local Sky Model (LSM) is constructed by selecting from the GSM. Sources inside the field of view, and bright sources outside the field of view are included.

The steps required to calculate the model visibility arising from the LSM are as follows:

- The LSM is converted into a set of LSM images.
- There may be multiple LSM images per field, to allow for the processing of outlier fields.

- Each LSM image is transformed to the Fourier plane and model visibilities are derived using a de-gridding approach (more in section 3.8.2).

Accurate processing of the LSM is essential for the scientific performance of the telescope. As such there are a number of possible effects yet to be fully investigated. One of the most important concerns the use of pixellated images to represent compact objects. This allows fast and accurate calculation of image plane based effects such as the w-term and the phased array feed primary beams. However, strong compact sources are problematical for pixel based images [7]. When the centroid of a compact source does not lie on a pixel, the signal is actually spread over a number of pixels in a sinc-like pattern. This occurs because the residual image is a representation of the visibility data, not of the sky directly (see section 6.3 in [7]). In the source fitting step, a compact point source is located by fitting in the neighbourhood of the peak. For the forward step to be accurate, the pattern (roughly a sinc function) for a compact source must be reproduced at least in some approximation. This does not require a finely sampled or carefully centred pixel grid. The other advantage of this approach is that it allows application of the wide-field effects during the forward step.

3.5 Continuum subtraction

As it was mentioned in section 2.4, every ASKAP field will contain many compact continuum sources with the combined flux density of about 50 Jy. Although vital for calibration, this has a downside that these sources will also be present in the spectral line dataset. Therefore continuum subtraction procedure is required prior to imaging of the spectral line data. There are a number of approaches to continuum subtraction (see e.g. [15, 16]), which can be summarised as follows:

- Accurate prediction of the model visibilities using a model (uvsub). Traditionally, line-free channels of the same dataset are used to construct the model. However, ASKAP can benefit from using the sky model (potentially updated following the continuum imaging of the same field). As described above, the sky model shall be subtracted prior to any imaging, effectively implementing this approach.
- Visibility-based subtraction (uvlin). This approach is quite popular because it is quite robust to calibration errors and does not require any source model (and is quite cheap computationally). It is based on polynomial fit into real and imaginary parts of each measured visibility spectrum. Such a fit is a linear process (unlike a fit to phase and amplitude), which has a well understood behavior and noise properties [16].
- Image-based subtraction (imlin). The polynomial is fitted into and subtracted from every slice along the frequency axis of the dirty image. Note that the subtraction has to be done from the dirty or residual image to preserve linearity. This approach results in a smaller number of operations than uvlin for long integrations (longer than 5.6 hours, assuming image cell size of 2 arcsec and 1 degree field of view per beam). In addition, it can be executed off-line, and therefore can have a better tuned spectral window to reject channels with the known line emission. It also has different performance properties from uvlin: the performance degrades with the distance from bright source rather than with the distance from delay centre [15].

Although the model contains sources down to 1 mJy, the fitted flux for the components will include everything down to zero flux – that is, there will be little or no bias. In this case, the imperfections of the model-based continuum subtraction (i.e. uvsub) are expected to be below thermal noise level of the high spectral resolution data for observations shorter than 12 hours. If this is insufficient, for instance, for deep integrations, an additional imlin-based step will be used.

The advantage of imlin over uvlin in this circumstance is the direction-independence of the effectiveness of the former. Both these polynomial fit-based algorithms are only effective for a small field of view. According to [16], for uvlin the effectiveness is a function of parameter η

$$\eta = \frac{\Delta\nu}{2\nu}\frac{\theta_{FOV}}{\theta_{res}}, \tag{1}$$

where $\Delta\nu$ is the total bandwidth and ν is a frequency of interest, while θ_{FOV} and θ_{res} are the field of view and the resolution. For a linear fit, η should be much less than 1. Higher order polynomials may be useful for slightly higher η. However, the approach fails for $\eta \gg 1$ because the response to a point source in the real and imaginary spectra becomes a sine wave spanning multiple periods across the band, which cannot be adequately approximated by a polynomial. Given the large fractional bandwidth of ASKAP, $\eta > 17$ even if only 2 km array is considered. The performance of imlin degrades with the distance from bright sources in the image rather than with the distance from the delay centre, but is also a function of η. Given the good uv-coverage of ASKAP and, therefore, a low sidelobe level of the resulting PSF, the hybrid uvsub+imlin approach should deliver the residual error of continuum subtraction better than 10 μJy.

It is noted that there are particular cases where continuum subtraction would not be desirable, for instance spectral-line imaging in the Galactic plane where many continuum sources would exhibit strong HI absorption. While imaging without continuum subtraction should be possible, the implications for the processing have not been fully worked through. It is likely that experience with real data will be the best way to evaluate this.

3.6 Approach to Calibration And Imaging

3.6.1 Normal equations

ASKAP calibration and imaging is formulated in terms of the relevant normal equations for a fit to the observed visibility data. A typical synthesis measurement equation (negelecting, for the moment wide-field effects that are discussed in Section 3.8.3) is:

$$\begin{aligned} V_{i,j}\left(u,v,t,f;P\right) = \quad & g_i(t,f;P_G)g_j^*(t,f;P_G)\times \\ & \int J_{i,j}\left(l-l_0,m-m_0,t,f;P_J\right) I\left(l,m,f;P_I\right) e^{j2\pi\left(\frac{f}{f_0}\right)(ul+vm)}\, dldm \end{aligned} \tag{2}$$

In an observation, the left hand side is measured, and the right hand side parameters are to be estimated. The terms g are the visibility-based gains, and the terms J are the image plane effects such as antenna primary beams. The term I is the sky brightness. We use parametrised versions of all these terms. For calibration, we can use the visibilities as constraints directly. Expanding the model visibility via a Taylor series, we have:

$$V_{i,j}\left(u,v,t,f;P\right) = V_{i,j}\left(u,v,t,f;P_0\right) + \nabla_P V^T\left|_{P=P_0}\right. \delta P \tag{3}$$

Using the observed visibility as the constraint, we have condition equations:

$$V_{i,j}^{res}\left(u,v,t,f\right) = V_{i,j}^{obs}\left(u,v,t,f\right) - V_{i,j}\left(u,v,t,f;P\right) = \nabla_P V^T\left|_{P=P_0}\right. \delta P \tag{4}$$

Indexing the visibilities by k and using A for the gradients, we have design equations:

$$\begin{aligned} V_k^{res} &= \sum_l A_{kl}\delta P_l, \text{ or} \\ V^{res} &= A\delta P, \end{aligned} \tag{5}$$

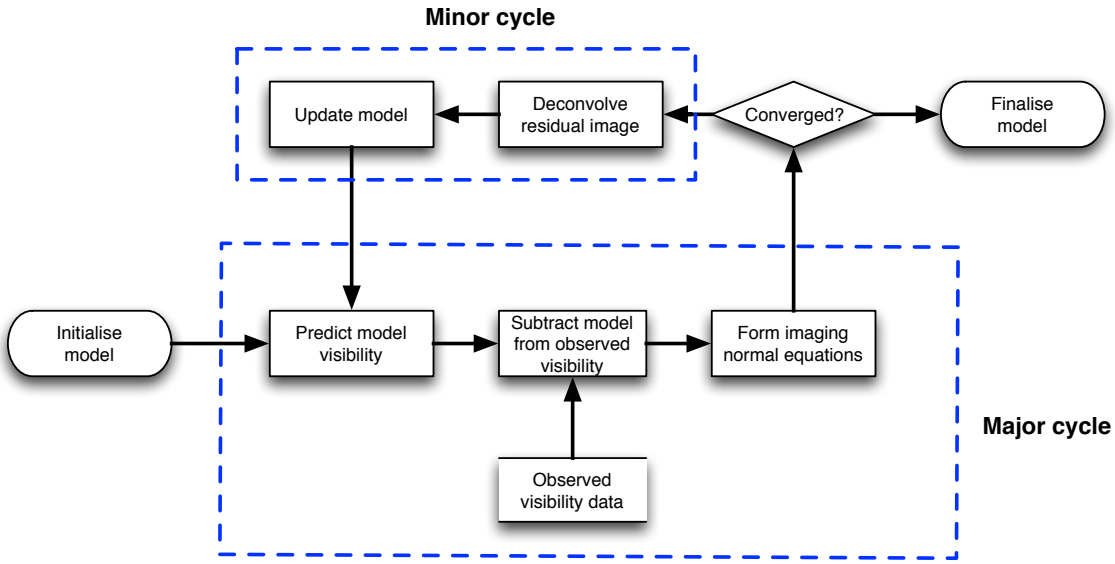


Figure 15: Schematic of imaging using normal equations.

where A is the *design matrix*, having size $M \times N$, where M is the number of condition equations and N is the number of parameters. We form the normal equations by multiplying both sides by the transpose of the design matrix (thus projecting back into parameter space).

$$A^T Q^{-1} V^{res} = \left(A^T Q^{-1} A\right) \delta P, \tag{6}$$

where $\left(A^T Q^{-1} A\right)$ is the normal matrix.

3.6.2 Solution of normal equations

There are two main approaches to solving these equations: via (non-linear) solution of the normal equations and via (non-linear) solution of the design equations. For M observations giving information on N parameters, the design matrix is $O(MN)$ and the normal matrix is $O(N^2)$. For stability and robustness, we want $M \gg N$ in which case the normal equations are preferred. For imaging, the left hand side is essentially the residual image, and the right hand side is a convolution. Thus this equation is essentially the standard convolution relationship between dirty image, point spread function, and true sky. These equations may be summed to produce aggregate normal equations. To distribute our processing steps across multiple processors, we partition the data according to some scheme (e.g. by frequency channels), assign each partition to a node or core, and then sum the resulting normal equations. This approach will be used for both imaging and calibration. For imaging with N pixels on each axis, the full normal matrix is $O(N^4)$. For this reason, we retain only the diagonal terms of the imaging normal equations. This translates into an assumption that the dirty PSF is constant across the field of view. For imaging, the processing splits into two parts, a prediction (or forward) step, and an inverse step

$$\delta I = \left(A^T Q^{-1} A\right)^{-1} A^T Q^{-1} V^{res}. \tag{7}$$

The inverse matrix usually is singular and consequently a non-linear procedure must be used. In addition, the matrix is too large to calculate and so some approximation is necessary. For ASKAP, a diagonal approx-

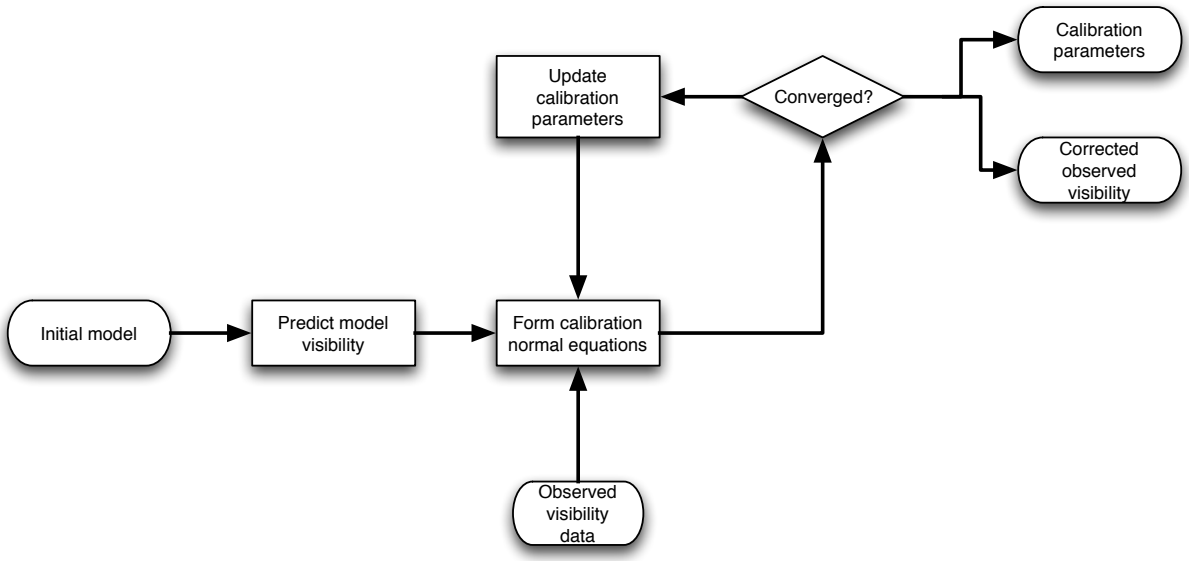


Figure 16: Schematic of calibration using normal equations.

imation is sufficient for the weighting, and a shift-invariant point spread function is adequate. Any small errors in this approximation are corrected in the forward step

$$V^{res} = V^{obs} - V^{mod}(P), \tag{8}$$

which can in principle be calculated with high precision. The model visibility is a function of multiple parameters. The deconvolution is performed using a major/minor cycle algorithm. In the major cycle, a given model for the sky is used to predict model visibilities – the forward step (Equation 8). These are subtracted from the observed visibilities, and a residual image calculated. In the minor cycle, an approximate deconvolution is performed using the residual image, and an approximate point spread function (Equation 7). As long as the major cycle prediction is sufficiently accurate, this approach will almost always converge. Divergence is detected and flagged.

Estimation of the calibration parameters proceeds by iteratively solving the normal equations for parameter updates and applying these. Self-calibration may be inserted inside the major/minor cycle (see Figure 17) at relatively little cost since the model and observed visibilities are already calculated.

3.7 Calibration

3.7.1 Effects requiring calibration

There are multiple physical effects to be calibrated in ASKAP (see Table 2). All effects are calibrated by astronomical observations, save for the PAF element gains and the ionospheric Faraday rotation correction. The PAF element gains are calibrated with respect to noise radiators on the surface of each antenna. The autocorrelation function of signals from the PAF elements is measured, and the weights at the beamformer adjusted appropriately. The ionospheric correction is obtained from GPS data and models of the Earth's magnetic field.

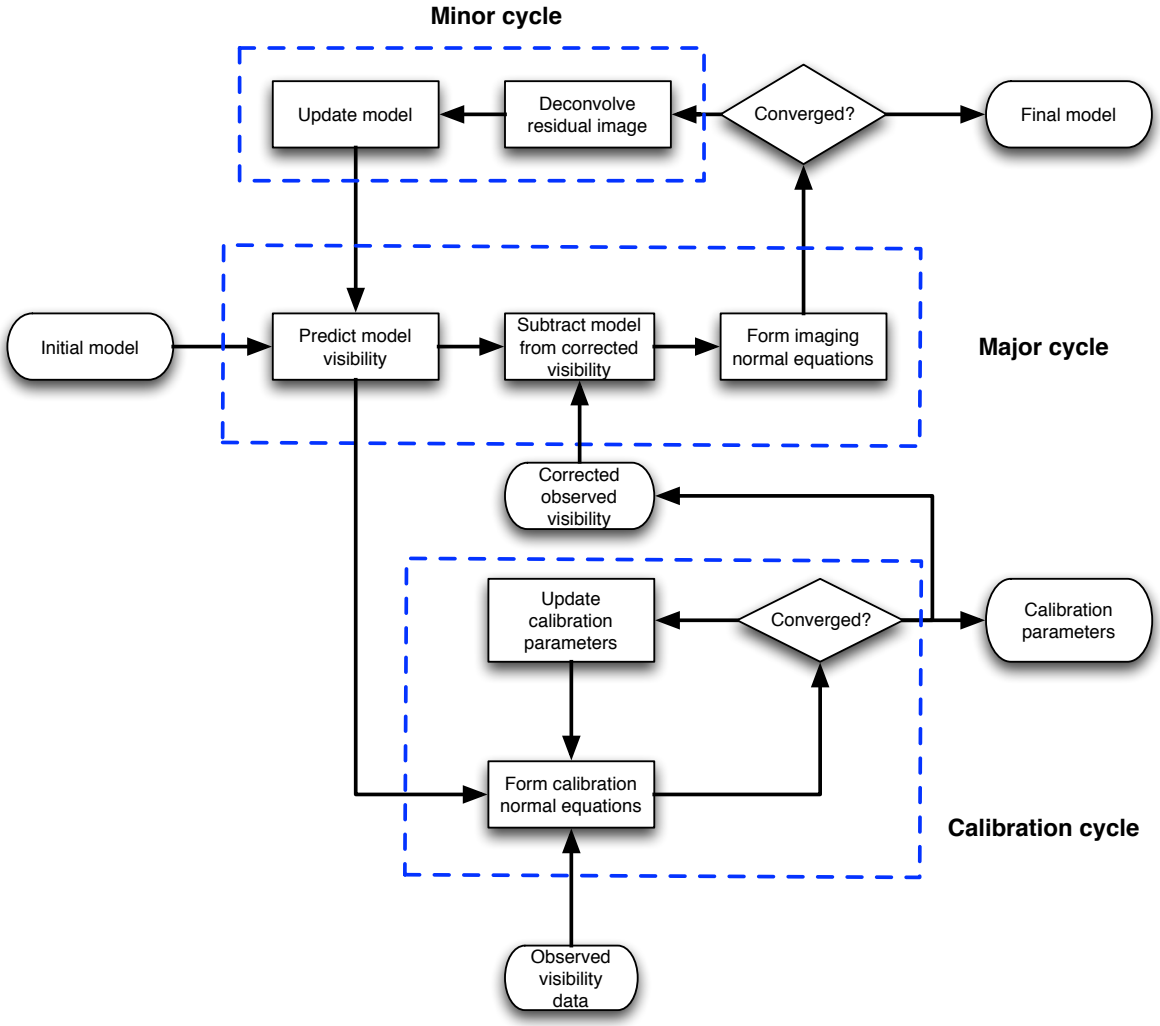


Figure 17: Schematic of major/minor cycle iteration, including self-calibration.

Table 2: Physical effects requiring calibration

Effect	Scales	Solution method
PAF Element gain changes	Minutes; 1 MHz	Noise radiators on antenna surface; Auto-correlation at beam former
PAF beam electronic gain changes	Minutes; 300 MHz	Local Sky Model; previous few minutes of observation
PAF beam coarse bandpass changes	Hours; 1 MHz	Local Sky Model; previous hours of observation
PAF beam fine bandpass changes	10 hours; 18 kHz	Local Sky Model; previous long observation
Ionospheric Faraday rotation	30sec – minutes	GPS data; Earth's magnetic field models
Distant sidelobes unknown or variable	Minutes; 1 MHz?	Global Sky Model; peeling
Absolute flux scale varies	Days; 300 MHz?	1934-638 + Local Sky Model; occasional dedicated observing
Primary beams unknown or variable	Weeks; 10 MHz?	Holography
Antenna locations unknown	Weeks; 300 MHz?	Global Sky Model; hours of observing on multiple fields
Pointing offsets unknown or variable	Tens of Minutes/ Hour; 300 MHz?	Local Sky Model; previous hour of observation

3.7.2 Solution for calibration parameters

Calibration proceeds from the normal equations similarly to imaging. The application of the least-square fit formalism is intuitively more clear in the case of the calibration than for the imaging, because P in Equations 3 to 6 represents the actual unknown parameters (such as antenna and beam-dependent gains) rather than image pixels. In general, the calibration parameters are complex and the fit proceeds independently for the real and imaginary parts, which are treated as separate parameters. Therefore, without a loss of generality, the parameters P can be considered real-valued throughout the following discussion. In principle, Equation 6 can be used directly to form an update for calibration parameters. However, in general non-linear least-square fitting demands multiple iterations (keeping the instrument well calibrated means that a small number of iterations, possibly even one, may be sufficient). Each such iteration involves building normal equations (6) via an iteration over the data, which is expensive. To improve this situation, (6) has been refactored such that the parameters and associated gradients are separate from the visibility data and models. This allows any visibilities that share the same unknown parameters to be averaged together before the fitting process begins, and multiple iterations to improve the parameters can be performed on this reduced dataset. The *pre-averaging* technique is easier to explain in the scalar case (like equation 2). Suppose that the model visibility is

$$V_k = \alpha_k(P)\tilde{V}_k, \tag{9}$$

where all dependence on the calibration parameters is in the factor $\alpha_k(P)$, which is assumed to be a scalar for the moment, and $\tilde{V}_k$ is the perfect model visibility ignoring calibration effects. Then, following the same

approach as used in the derivation of the design equations (5) we get

$$V^{res} = \begin{pmatrix} \tilde{V}_1 & 0 & \cdots & 0 \\ 0 & \tilde{V}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{V}_M \end{pmatrix} A\delta P, \tag{10}$$

where matrix A is composed with gradients of individual α_k with respect to parameters P. The diagonal matrix composed with perfect model visibilities is absorbed into the modified weight matrix $\tilde{Q}^{-1}$:

$$\tilde{Q}^{-1} = Q^{-1} \begin{pmatrix} \tilde{V}_1 & 0 & \cdots & 0 \\ 0 & \tilde{V}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{V}_M \end{pmatrix}, \tag{11}$$

where Q^{-1} is the original weight matrix. This leads to the following normal equations

$$A^T \tilde{Q}^{-1} V^{res} = \left(A^T \tilde{Q}^{-1} A \right) \delta P. \tag{12}$$

The main difference between equation (12) and the original equation (6) is that the normal matrix is only dependent on the unknown parameters, not the visibility model. Therefore the length of the data vector (and the dimensions of matrices A and $\tilde{Q}^{-1}$) can be reduced from M to $\tilde{M}$ by summing individual rows (i.e. summing individual condition equations) without a loss of much information. These $\tilde{M}$ pre-averaged data points will be buffered and iterated over during the major cycle of the calibration without a need to read the original dataset. The application of the outlined scalar approach to the gain calibration is quite straight forward. However, generalisation to the arbitrary measurement equation has the following issues which still require some investigation

- Different calibration effects absorbed in $\alpha(P)$ may not always commute with the averaging in time. One of the examples is the polarisation calibration which may require longer observations and/or special scans observing with a number of different rotations with respect to parallactic axis. Our approach to address this is as follows
  - Polarisation calibration is performed by a separate pipeline which combines the data in chunks and works with a larger $\tilde{M}$
  - We keep the instrument well calibrated. Provided the coupling between polarisation leakages and gains is small, the predict forward approach is expected to converge.
- The vector case results in a block-diagonal weight matrix (with 8×8 blocks due to 4 polarisation products with real and imaginary parts). The individual blocks may not always be invertible, so a pseudoinverse has to be constructed to project the residual visibilities into an appropriate space (an analog of the division by the model visibilities in the scalar case). The limitations to the overall calibration procedure imposed by this operation are not clear at the moment.

The overall performance of calibration cannot be known without experience with the final telescope. To the extent that the effects are well understood, the relevant tolerances have been set with respect to the scientific goals such as dynamic range and noise. However, many calibration issues will not be well understood until the start of observing.

Table 3: Correspondence between normal equation and imaging terminology

Normal equations	Imaging
δI	Update to model
$Q_{ij,p}^{-1}$	Visibility weight $w_{ij,p}$
V^{res}	Residual visibility $V_{ij,p}$
$diag\left(A^T Q^{-1} A\right)$	Weights image $\sum_p w_p A_p^2(l,m)$
$slice\left(A^T Q^{-1} A\right)$	PSF $B^{LM}(l,m)$

3.8 Imaging

This section explains the way in which ASKAP forms images from calibrated visibilities. Note to simplify some of the discussion the notation used here is different from that used above. The correspondence is shown in Table 3.

3.8.1 Properties of linear mosaics

Imaging with ASKAP relies on the construction of linear mosaics of the residuals. If deconvolution is not needed, an image is formed by suitable scaling of the residuals. The measurement equation for an interferometer measuring the sky brightness I is that the visibility function V is given by a Fourier transform. Indexing by p for the p'th set of measurements, and i, j as the antenna indices:

$$V_{ij,p} = \int A_{ij,p}(l,m) I(l,m) e^{j2\pi(ul+vm)}\, dldm. \tag{13}$$

The dirty images are given by the following equation:

$$I_p^D(l,m) = \frac{\sum_{ij} w_{ij,p} V_{ij,p}\; e^{-j2\pi(u_{ij}l+v_{ij}m)}}{\sum_{ij} w_{ij,p}}, \tag{14}$$

The weights w can be chosen according to a number of criteria. See [7] for a very detailed discussion of the possible criteria. It usually requires an additional pass through the entire data set. Using the Fourier convolution function, the dirty image can be shown to be related to the true sky by convolution with a point spread function:

$$\begin{aligned} I_p^D(l,m) &= B_p \otimes I(l,m), \text{ where} \\ B_p(l,m) &= \Re \frac{\sum_{ij} w_{ij,p} e^{-j2\pi\left(u_{ij}l+v_{ij}m\right)}}{\sum_{ij} w_{ij,p}} \end{aligned} \tag{15}$$

In the case of multiple pointings, a *linear mosaic image* can be formed using a least squares approach

$$I^{LM}(l,m) = \frac{\sum_p w_p A_p(l,m)\, I_p^D(l,m)}{\sum_p w_p A_p^2(l,m)}. \tag{16}$$

If the primary beams A are identical for all antennas, this corresponds to a weighted sum of the dirty images formed by a Fourier sum of the observed visibilities. In terms of the dirty images, the linear mosaic is a sum over pointings p:

$$\begin{aligned} I^{LM} &= \frac{1}{\eta(l,m)} \sum_p A_p(l,m)\, I_p^D(l,m) \\ \eta(l,m) &= \frac{\sum_p w_p A_p^2(l,m)}{\sum_q w_q} \\ w_p &= \sum_{ij,p} w_{ij,p} \end{aligned} \tag{17}$$

A heuristic explanation of this equation reveals how point sources appear. The dirty image is a collection of point sources, each convolved with the relevant synthesized PSF, B. Thus the linear mosaic of the dirty images is also a collection of point sources. Any given point source is visible from multiple pointings.

$$I^{LM}(l,m) = \frac{1}{\eta(l,m)} \sum_p A_p(l,m) \left(\sum_q A_p(l_q,m_q)\, B_p(l, m-l_q, m_q)\, S_q \right). \tag{18}$$

In words, for a given pointing, each point source is:

- multiplied by the primary beam at the point source,
- propagated to the evaluation point by the synthesized PSF for the relevant offset
- multiplied by the primary beam at the evaluation point.

This effect is then summed over all pointings and all sources. The most important aspect of this equation is that the footprint of a given source is the synthesized PSF nearby, and is the synthesized PSF times the primary beam far away. The dividing line for the distance is roughly the antenna full-width-to-zero. Far away from the pointing centre, the antenna gain is essentially that of an isotropic antenna ~ -30 dB. The distant sidelobes of the synthesised beam are about -30 dB. Hence a bright source in the sidelobes would have to be 60 dB above the noise floor to be significant. The linear mosaic PSF is manifestly shift variant (see Equation 19)

$$B^{LM}(l,m,l_q,m_q) = \frac{1}{\eta(l,m)} \sum_p A_p(l,m) \left(\sum_q A_p(l_q,m_q)\, B_p(l, m-l_q, m_q) \right). \tag{19}$$

For ASKAP, the point spread functions will be very close to identical. A good approximation can be had by assuming that all the primary beams are shift invariant:

$$B^{LM}(l,m,l_q,m_q) \approx \frac{1}{\eta(l,m)} A(l-l_q, m-m_q)\, B(l-l_q, m-m_q) \sum_p A(l, m-l_q, m_q) \tag{20}$$

gives approximate linear mosaic point spread function for identical synthesised beams B and primary beams A.

Rau [5] notes that the definition of the linear mosaic in Equation 16 is not suitable for deconvolution of a single field – a necessary condition. Instead the deconvolution should proceed on a modified version:

$$I^{LM}(l,m) \sqrt{\sum_p w_p A_p^2(l,m)} = \frac{\sum_p w_p A_p(l,m)\, I_p^D(l,m)}{\sqrt{\sum_p w_p A_p^2(l,m)}}. \tag{21}$$

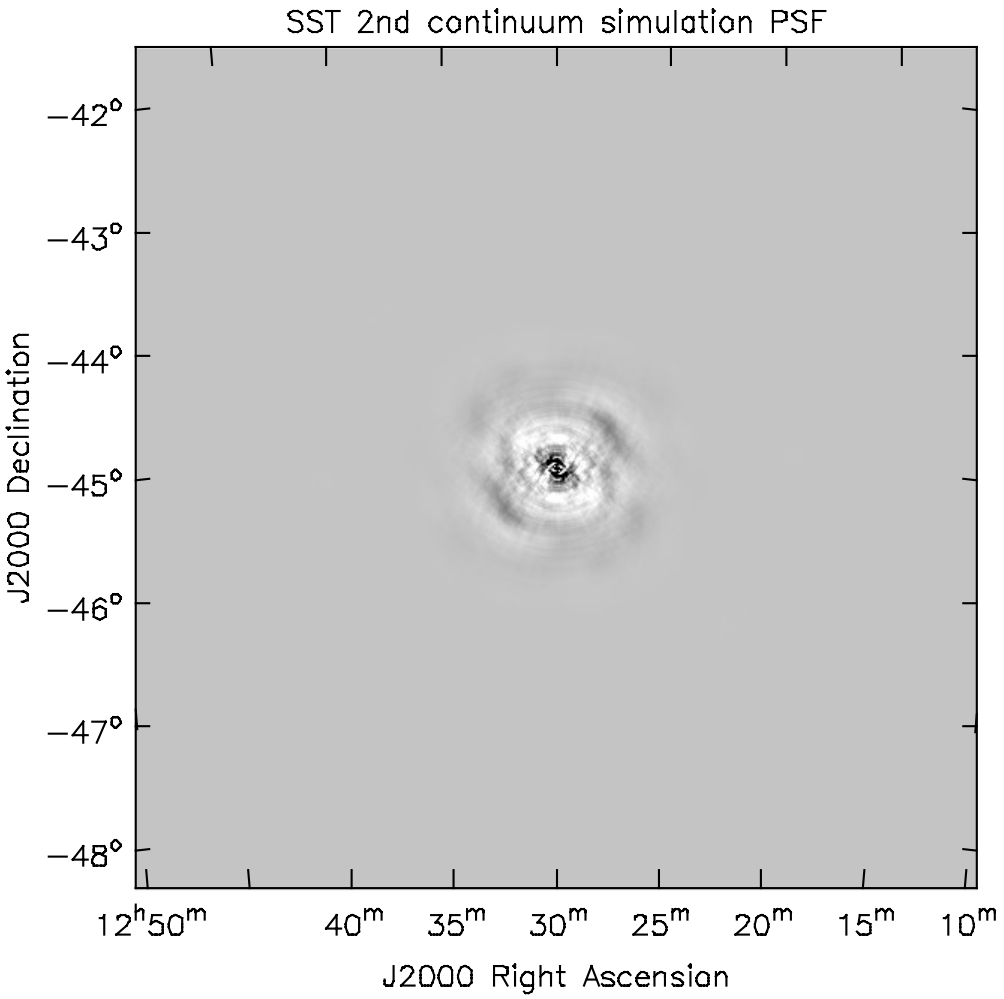


Figure 18: Point Spread Function for long (8h) continuum observation, processing with Wiener preconditioning including an image plane taper.

.

The right hand side of Equation 21 can be calculated, deconvolved using the approximate PSF in Equation 19, and then rescaled by dividing out the second part of the LHS. Alternatively, the right hand side of Equation 16 is calculated, and then a SNR mask is applied during the deconvolution. The SNR mask is:

$$M\left(l,m\right)=\sqrt{\sum_{p}w_{p}A_{p}^{2}\left(l,m\right)}. \tag{22}$$

This means that bright sources outside the main sensitivity can be deconvolved. This is the approach used in ASKAP. In CLEAN, the search for the peak residual is converted to a search for the peak in the SNR.

In figures 18 to 21, we illustrate ASKAP imaging using a simulation of the SKADS model sky. Note that the PAF beams are densely packed, at $\lambda/(2D)$, in this example thus representing a quarter of a full field image.

Finally it is necessary to connect the terminology in this section and that of the normal equations (see

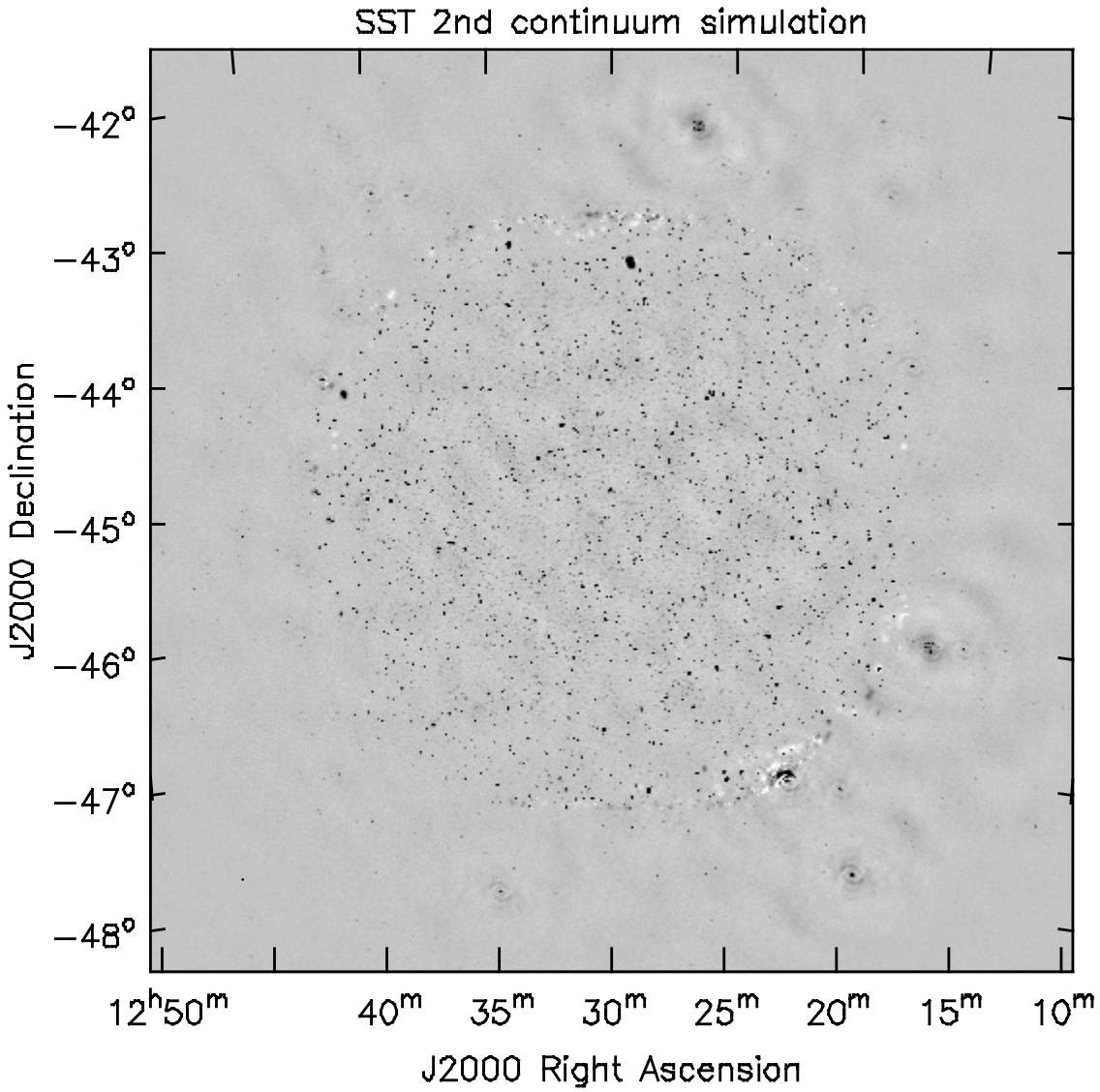


Figure 19: Simulation for long (8h) continuum observation, processing with Wiener preconditioning including an image plane taper.

.

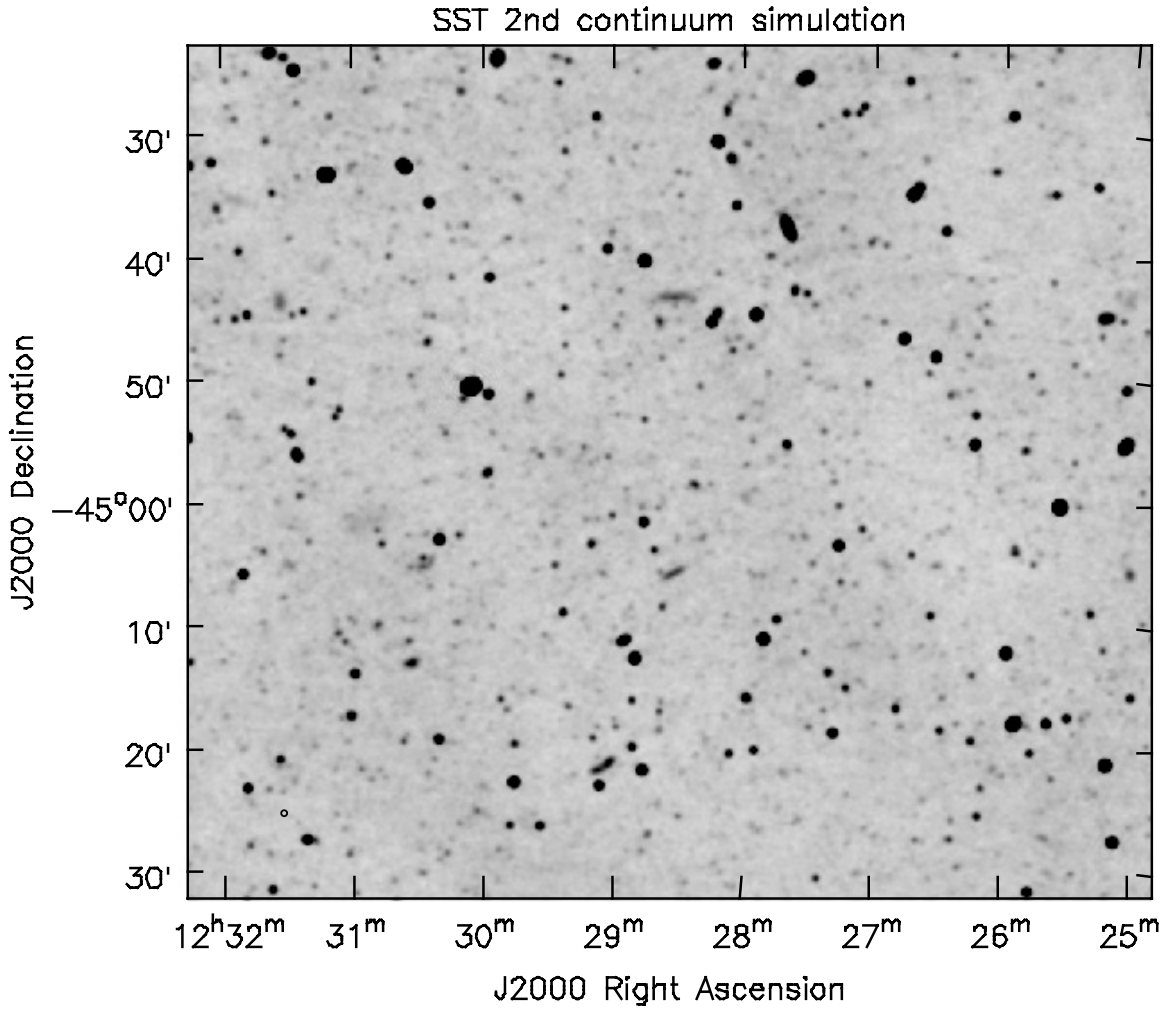


Figure 20: Zoom of simulation for long (8h) continuum observation, processing with Wiener preconditioning including an image plane taper.

.

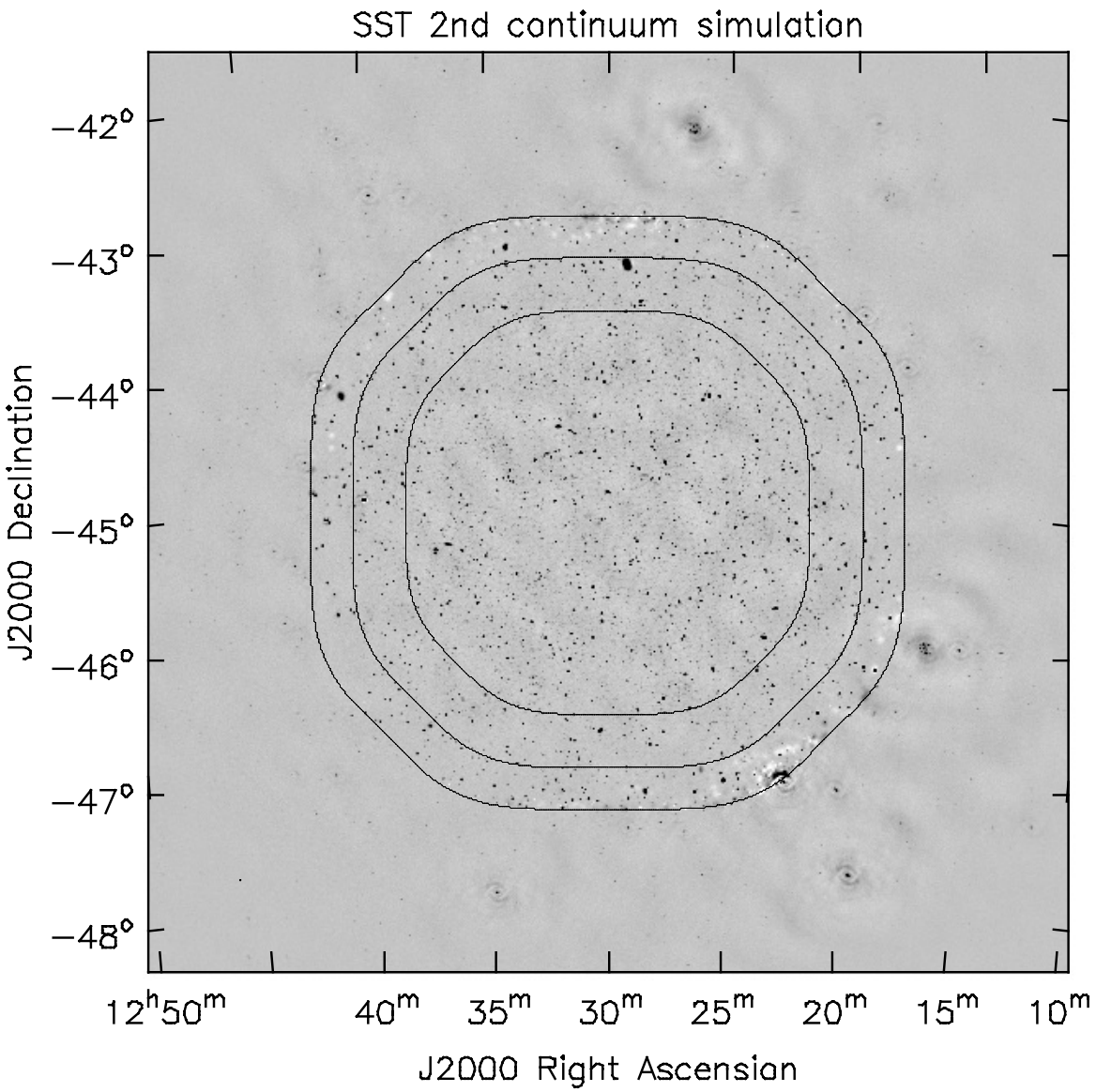


Figure 21: Weight image for simulation for long (8h) continuum observation, processing with Wiener preconditioning including an image plane taper.

.

Table 3). The matrix Q is the covariance matrix of the visibility errors. It is usually assumed that Q is diagonal with values equal to the visibility weights.

3.8.2 Gridding

In order to make use of the Fast Fourier Transform to calculate the normal equations, the visibility data must be transformed to a regular grid. Standard, small field processing uses an approach called convolutional resampling. The visibility data are sampled at discrete locations in u,v space. This forms a sampled visibility function that is composed of a collection of Dirac δ-functions. By convolving this with a finite-size kernel, this is converted into a smooth function, which can then be resampled at the grid points. Following the FFT, the effects of the smoothing kernel may be corrected to some approximation by dividing out the transform of the smoothing kernel. The smoothing kernel is usually a prolate spheroidal wave function. This is chosen for the property of minimising aliasing for a fixed support size in u,v. The processing in this approach scales as the number of visibility samples, and the size of the convolutional kernel in pixels (typically in the range 50–100).

3.8.3 Wide field imaging

ASKAP is subject to a number of wide field effects:

- Image plane effects such as the w term or non-coplanar baselines effect ([2])
- Primary beam variations, including time and frequency variations due to changes in the phased array feed and associated electronics and signal path.
- Antenna time-variable pointing errors

First we describe how the non-coplanar baselines effect is countered. There are two approaches in use in ASKAPsoft - wprojection and snapshots.

W Projection Image plane effects may be corrected in the gridding step. Recall that for small fields of view, the visibility obeys:

$$V(u,v) = \int I(l,m) e^{j2\pi(ul+vm)}\, dldm. \tag{23}$$

For larger fields of view, the w-term becomes important

$$V(u,v) = \int \frac{I(l,m)}{\sqrt{1-l^2-m^2}}\, e^{j2\pi\left(ul+vm+w\left(\sqrt{1-l^2-m^2}-1\right)\right)}\, dldm. \tag{24}$$

The new term w is the component of antenna-antenna vector towards the phase centre of the PAF field of view. The physical origin of this phase term is straightforward – it comes from the need to refer the electric field to the same physical plane. This requires a Fresnel term. The w-term effect is significant if the field of view is comparable to or greater than the square root of the resolution (both measured in radians):

$$\theta_{FOV} \geq \sqrt{\theta_{resolution}} \tag{25}$$

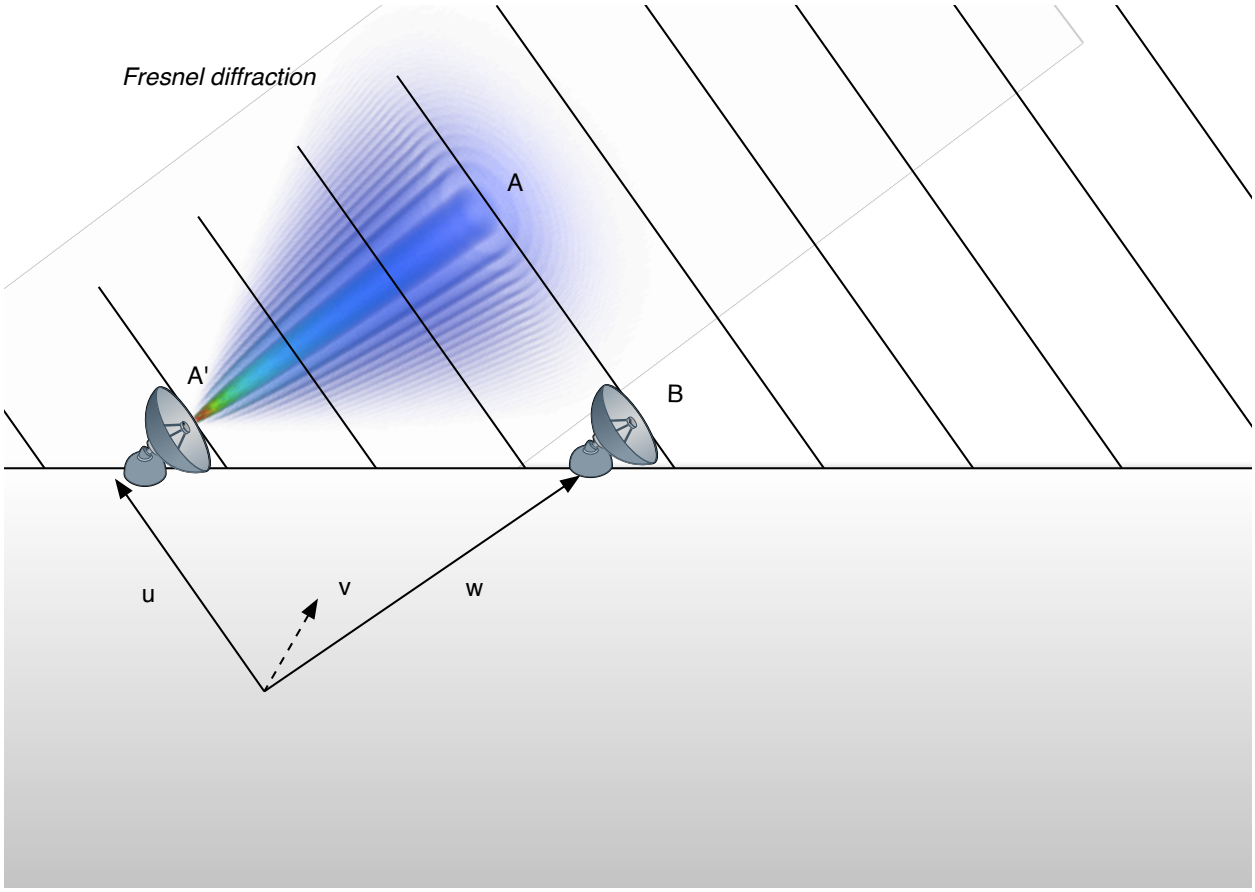


Figure 22: The visibility measured between points A and B is the two dimensional Fourier transform of the brightness. A measurement at A'B must be corrected to the plane AB. A Fresnel transform is required for this propagation.

For ASKAP, the maximum allowed half-width of an antenna beam is $\sim$ 0.7 degrees. The actual maximum depends upon the accuracy required in the peak fluxes, and will be typically smaller. The w term can be expressed as a multiplicative effect in image space, or a convolution in Fourier space. In both cases, the w variable acts as a parameter. The convolution relationship between visibility on $w = 0$ and an arbitrary w plane is

$$\begin{aligned} V(u,v,w) &= \tilde{G}(u,v,w) \otimes V(u,v) \\ \tilde{G}(u,v,w) &= \int \frac{e^{\,j2\pi w(\sqrt{1-l^2-m^2}-1)}}{\sqrt{1-l^2-m^2}}\, e^{\,j2\pi(ul+vm)}\, dldm \\ \tilde{G}(u,v,w) &\approx e^{\,j\pi w\left(u^2+v^2\right)} \end{aligned} \tag{26}$$

In conventional imaging, an Anti-Aliasing Filter (AAF) is used as the GCF. An AAF is factored in as follows:

- In calculating the GCF $\tilde{G}(u,v,w)$, multiply the GCF transform in image space by the IFFT of the AAF.
- In the prediction step before the FFT, multiply the image and the convolution function transform by the IFFT of the AAF.
- In the reverse step, divide the residual image, weights image and PSF by the FFT of the AAF.

For imaging in the presence of the w term, two approaches are supported:

- W stacking: The data are partitioned in w, and the AAF is applied in image space to each w plane.
- W projection: use $\tilde{G}(u,v,w)$ as a gridding function.

The primary beam can be treated in a similar way since it too is a multiplicative term in image space (see Equation 27).

$$\begin{aligned} V(u,v,w) &= \tilde{A}(u,v,w) \otimes V(u,v) \\ \tilde{A}(u,v,w) &= \int A\,(l,m)\, \frac{e^{\,j2\pi w(\sqrt{1-l^2-m^2}-1)}}{\sqrt{1-l^2-m^2}}\, e^{\,j2\pi(ul+vm)}\, dldm \end{aligned} \tag{27}$$

Mixed versions are also possible in which, for example, the w term is addressed using stacking, and the primary beam via projection. The w term convolution function, phase screen and transform are shown in Figures 23, 24 and 25.

Snapshots It is possible to ignore the w term altogether for short periods of time. For a snapshot, neglect of the w-term results in a distorted coordinate system that can be corrected in the image plane by interpolation of the image to the correct coordinate system. For that short period of time, the array will be instantaneously coplanar to good accuracy. This means that the w coordinate is related to (u,v) by a simple relationship:

$$w = au + bv \tag{28}$$

In this case, the relationship between visibility and sky brightness may be rewritten as a two-dimensional Fourier transform by introducing distorted coordinates (l', m') where:

$$\begin{aligned} l' &= l + a\sqrt{1-l^2-m^2} \\ m' &= m + b\sqrt{1-l^2-m^2} \end{aligned} \tag{29}$$

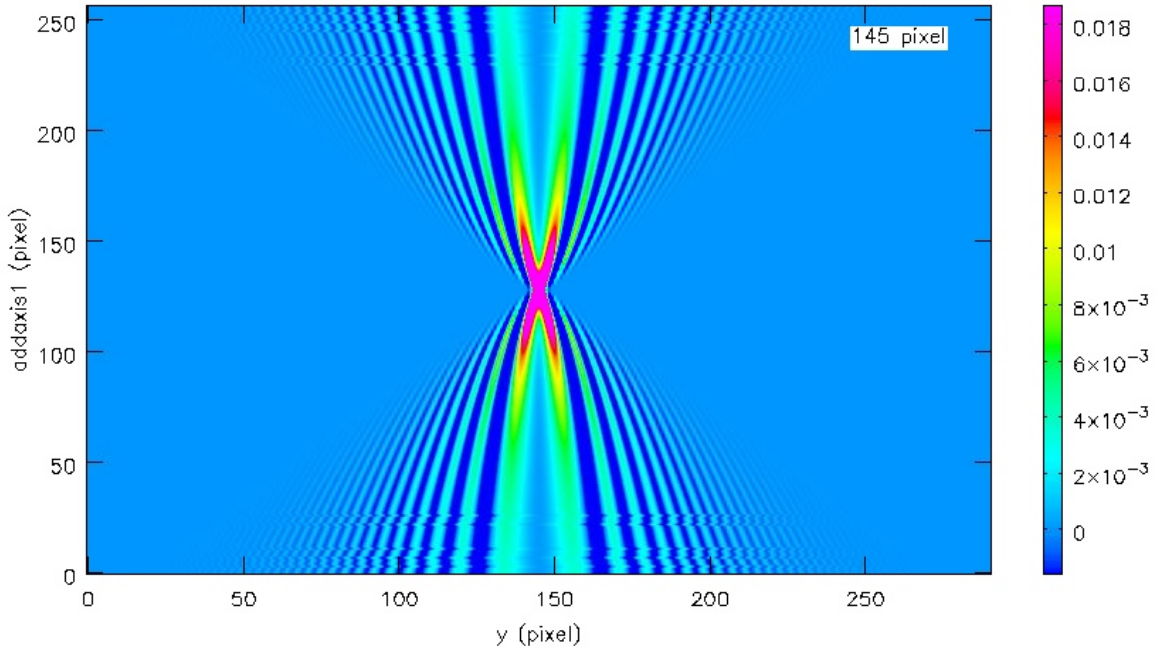


Figure 23: Slice through convolution function. Offset in u (horizontal axis) vs. offset in w (vertical axis). The image shows real part of the w term convolution function. Note that although the prominent bands in the convolution function scale as $\sqrt{w}$, the envelope scales as w.

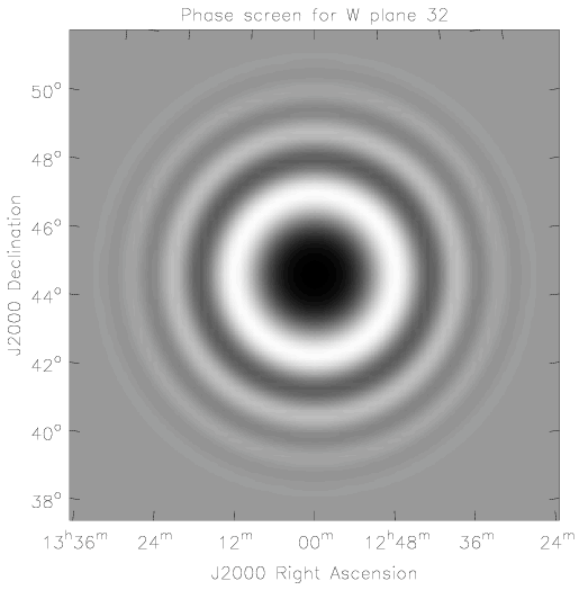


Figure 24: Real part of w-term phase screen

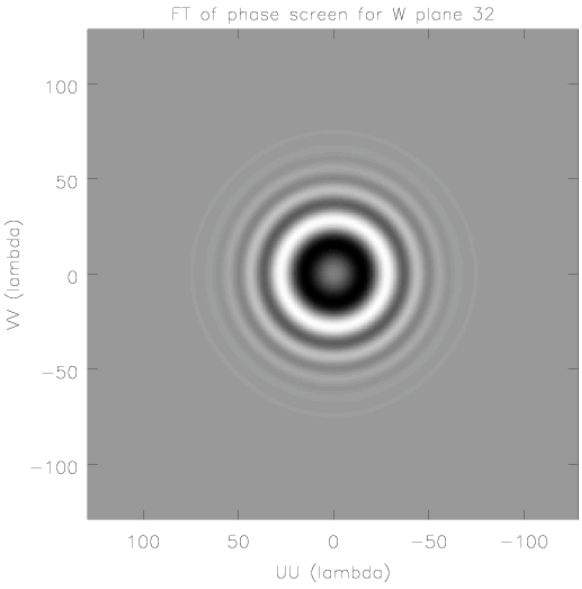


Figure 25: Real part of transform w-term phase screen

Table 4: Strength of source requiring peeling and the residual noise after peeling.

Distance	Inside PAF beam	Outside PAF beam
$\sim$ few synthesised beams	250 mJy	2.5 Jy
$\sim$ many synthesized beams	10 Jy	100 Jy

Parameters a, b may be estimated by linear fitting to the u, v, w coordinates. For a telescope observing at zenith angle Z and parallactic angle χ, the parameters a and b are given by:

$$\begin{aligned} a &= \tan Z \sin \chi \\ b &= -\tan Z \cos \chi \end{aligned} \tag{30}$$

The optimum solution depends on the context. For ASKAP, a combination of w projection and snapshot imaging is used to provide an optimum use of CPU and memory resources.

3.8.4 Peeling

Bright sources cause problems for high dynamic range imaging. Calibration uncertainties or changes can prevent accurate removal of the sidelobes arising from a bright source. The third axis of the ASKAP antennas removes one cause of such changes - the rotation of a source in the beam sidelobes. However, time variable sources or calibration changes still cause problems. In this case, it is necessary to resort to an adaptive approach whereby the apparent strength of each sufficiently bright source is monitored and the effect of that source removed from the measurements. This approach is known as *peeling*.

To calculate the magnitude of this effect and when peeling is required, consider two cases – a bright source outside the field of view and a medium brightness source inside the field of view. The footprint of a given source is the synthesized PSF nearby, and the synthesized PSF times the primary beam far away. The dividing line is distance is roughly the antenna full-width-to-zero.

- Far away from the pointing centre, the antenna gain is essentially that of an isotropic antenna ~ -30 dB. The distant sidelobes of the synthesised beam are about -30 dB. Hence a bright source in the sidelobes would have to be ~ 60 dB above the noise floor to be significant.

- Close to a source in the field of view, the maximum sidelobe of the synthesised beam is about a few percent, say -17 dB. The calibration of the PAF is then the determining factor. If the PAF beams are calibrated to e.g. 1%, then the total dynamic range will be about -37 dB. Moving away from the source, the synthesised beam sidelobes diminish to about -30 dB, and the dynamic range improves to about -50 dB.

Deconvolution is not sufficient to correct for time-variable beams. In this case, it is necessary track and remove sources -50 to -60 dB above the noise floor. For long continuum integration, the noise level is about 10 μJy. The residual noise figures after peeling are given in Table 4.

Peeling of each source proceeds in the self-calibration loop of the imager. The local sky model is examined for all sources that warrant peeling. For each of these, the amplitude of the gain is determined from a least squares fit to the residual visibility.

3.8.5 Time variable sources

Time-variable sources can corrupt the image in a manner similar to sources required peeling. The transient pipeline running in real time identifies variable sources and stores a light curve for each source. This light curve is then used to correct the observed data during the subsequent imaging.

3.8.6 Pre-conditioning

Re-weighting of visibilities is usually performed before or during gridding, with the goal of improving PSF properties such as the resolution and sidelobe level. A number of different reweighting schemes are available. Robust weighting [7] is the most powerful technique currently used. It allows a trade-off between sensitivity and sidelobe suppression.

If the imaging is considered as solution of linear equations relating the dirty image to the true sky by convolution with the point spread function, then the usual re-weighting can be considered to be a form of preconditioning of the data to ensure a point spread function with suitable properties, such as high resolution and low sidelobe levels. Preconditioning can also be performed after the construction of the dirty image and PSF [5]. It is necessary to use this approach to avoid delays in the processing arising from the distribution of data across the Central Processor. Two main forms of preconditioning are supported: a Wiener filter similar to robust weighting, and a Gaussian taper chosen to emphasise a given angular scale. The Wiener filter is implemented via a Fourier filter,

$$W_f = \frac{\widetilde{B}^*}{\widetilde{B}^*\widetilde{B} + \sigma_{uv}^2 / P_{signal}}, \tag{31}$$

where B is the synthesised beam that the filter is aiming to deconvolve, tilde represents a Fourier transform to uv, σ_{uv}^2 is the noise variance in each uv pixel, and P_{signal} is the spatial power spectrum of the signal that the filter is being tuned for. In general all of these terms are variable across the uv plane, and we tune for an unresolved point source with flux density, S, by setting $P_{signal} = S^2$. To get a feel for this filter, suppose that the visibilities have undergone nearest-neightbour gridding with natural weighting, i.e. with weights that are proportional to the inverse variance of the visibility noise (in practice the finite support of the kernels will smear the gridded data across nearby uv pixels). Furthermore, suppose that the noise is approximately equal for all visiblities, with variance σ^2. The noise variance for a uv pixel containing n visibilities will be $n\Delta S_{uv}^2$, where ΔS_{uv} is the RMS noise of a visibility with unit weight, and the uv sampling function $\widetilde{B}$ for the pixel will be $n\Delta S_{uv}^2 \sigma^{-2}$. The filter can be written

$$W_f \propto \frac{1}{S^2\widetilde{B} + \sigma^2}. \tag{32}$$

Equation 32 has the form of the robust weighting equation described in [7]. The robustness parameter, R, is used to further tune the filter between minimising PSF sidelobes ($S \to \infty$) and minimising noise ($S \to 0$). In line with robust weighting in packages such as *casa*, we define the Wiener preconditioner as

$$W_f = \begin{cases} \frac{1}{\overline{w}^{-1}(5\times10^{-R})^2\widetilde{B}+1}; & \widetilde{B} > \varepsilon \\ 0; & \widetilde{B} \leq \varepsilon \end{cases} \tag{33}$$

where $\overline{w}^{-1}$ is the average uv pixel weight-sum across visibilities. We use the approximation suggested in [7] (also used in *casa*) that the average weight-sum reduces to $\sum \widetilde{B}^2 / \sum \widetilde{B}$, which holds when the visibilities have equal noise variance. The threshold ε is required to suppress numerical artefacts arising from processes such as image regridding and anti-aliasing function correction. It is estimated from rumble in the zero-padding part of the uv plane.

Specifying a filter as in (33) is fine when using nearest-neighbour gridding, but for more complicated convolutional gridding approaches it risks damaging the associated apodizations and projections. To avoid this, $\widetilde{B}$ is replaced with its nearest-neighbour counterpart and smoothed with a top-hat function of a width that follows the maximum size of gridding kernels used in different parts of the uv plane. In this way uv density fluctuations are tracked, but not at scales that are finer than the convolutions.

The Fourier plane coverage for a long integration of the 2 km configuration has been designed to produce a Gaussian point spread function. However, there are some minor fluctuations in Fourier plane sampling density that lead to side lobes at just above 1%. It is possible treat these using a combination of Wiener filtering (to remove the fluctuations in sampling density) and Gaussian filtering (to restore the overall Gaussian shape). The effect of these parameters is shown in the ASKAP PSF simulator page [9] and Figures 26 and 27.

3.8.7 Deconvolution: Continuum Imaging

For continuum imaging, the minor cycle deconvolution of the point spread function is performed using the Multi-Frequency-Synthesis, Multi-Scale algorithm developed by Urvashi Rau. The algorithm is described in great detail in Rau's PhD thesis [5]. Only an overview is given here.

In single image Multi Frequency Synthesis, all visibility points are combined into one image, ignoring any spectral index effects. In multiple image MFS, the brightness is modelled explicitly as a power law in frequency, and the spectral effects estimated as images. More specifically, the brightness is expanded as a Taylor series in frequency, and the Taylor term images estimated.

$$I_p\left(l,m\right) = \sum_t \left(\frac{\nu-\nu_0}{\nu_0}\right)^t I_{p,t}\left(l,m\right) \tag{34}$$

where the terms $I_{p,t}$ are the Taylor terms for the sky brightness.

The estimation proceeds as a joint deconvolution of the brightness at a reference frequency and the Taylor terms. Ignoring for the moment the multi-scale aspect, expanding the brightness I around a reference frequency ν_0 gives a set of linear equations:

$$I_p^D\left(l,m\right) = \sum_t B_{p,t}\left(l,m\right) \otimes I_{p,t}\left(l,m\right) \tag{35}$$

$$I_{p,t}^D\left(l,m\right) = \frac{\sum\limits_{ij} w_{ij,p} V_{ij,p} \left(\frac{\nu-\nu_0}{\nu_0}\right)^t e^{-j2\pi\left(\frac{\nu}{\nu_0}\right)\left(u_{ij}l+v_{ij}m\right)}}{\sum\limits_{ij} w_{ij,p}}, \tag{36}$$

$$B_{p,t}\left(l,m\right) = \Re\frac{\sum\limits_{ij} w_{ij,p}\left(\frac{\nu-\nu_0}{\nu_0}\right)^t e^{-j2\pi\left(\frac{\nu}{\nu_0}\right)\left(u_{ij}l+v_{ij}m\right)}}{\sum\limits_{ij} w_{ij,p}} \tag{37}$$

A search for the optimum location for a component is conducted in all the dirty Taylor images using one of a set of heuristics (such as the peak in Taylor term 0). For the optimum location, the peak value is taken from the peak in the set of dirty Taylor images. These components are added to the model, and the effects removed from the dirty Taylor image. The process iterates until the remaining peak is below some threshold or a maximum number of iterations has been performed.

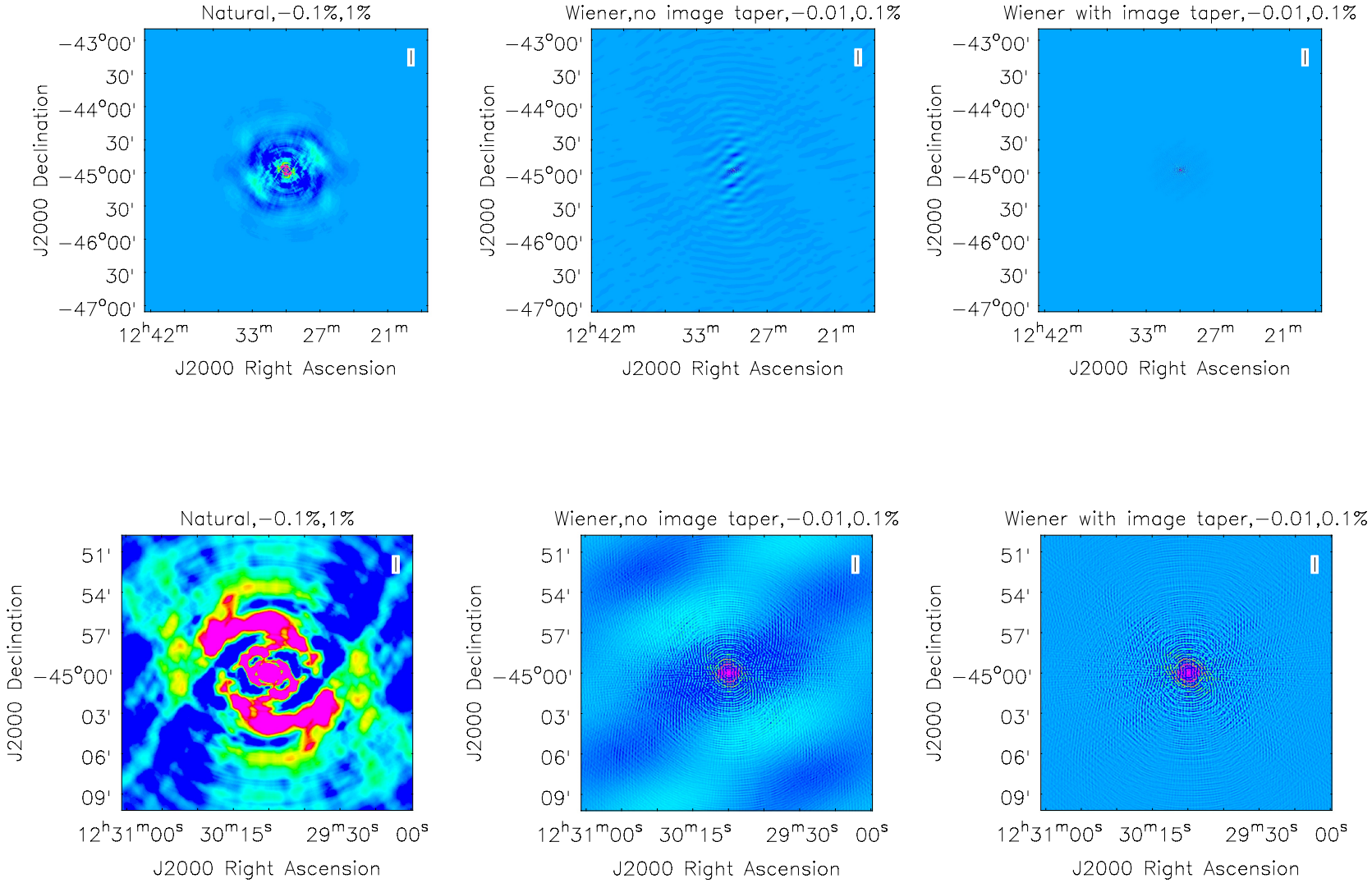


Figure 26: Point spread function before and after preconditioning. The bottom row shows the zoomed in version of the same plots. The left column is naturally weighted, the centre is preconditioned without the primary beam weighting, and the right column is with primary beam weighting. Note that the brightness range is -0.1% to 1% for the naturally weighted image, and -0.01% to $+0.1\%$ for the other two images.

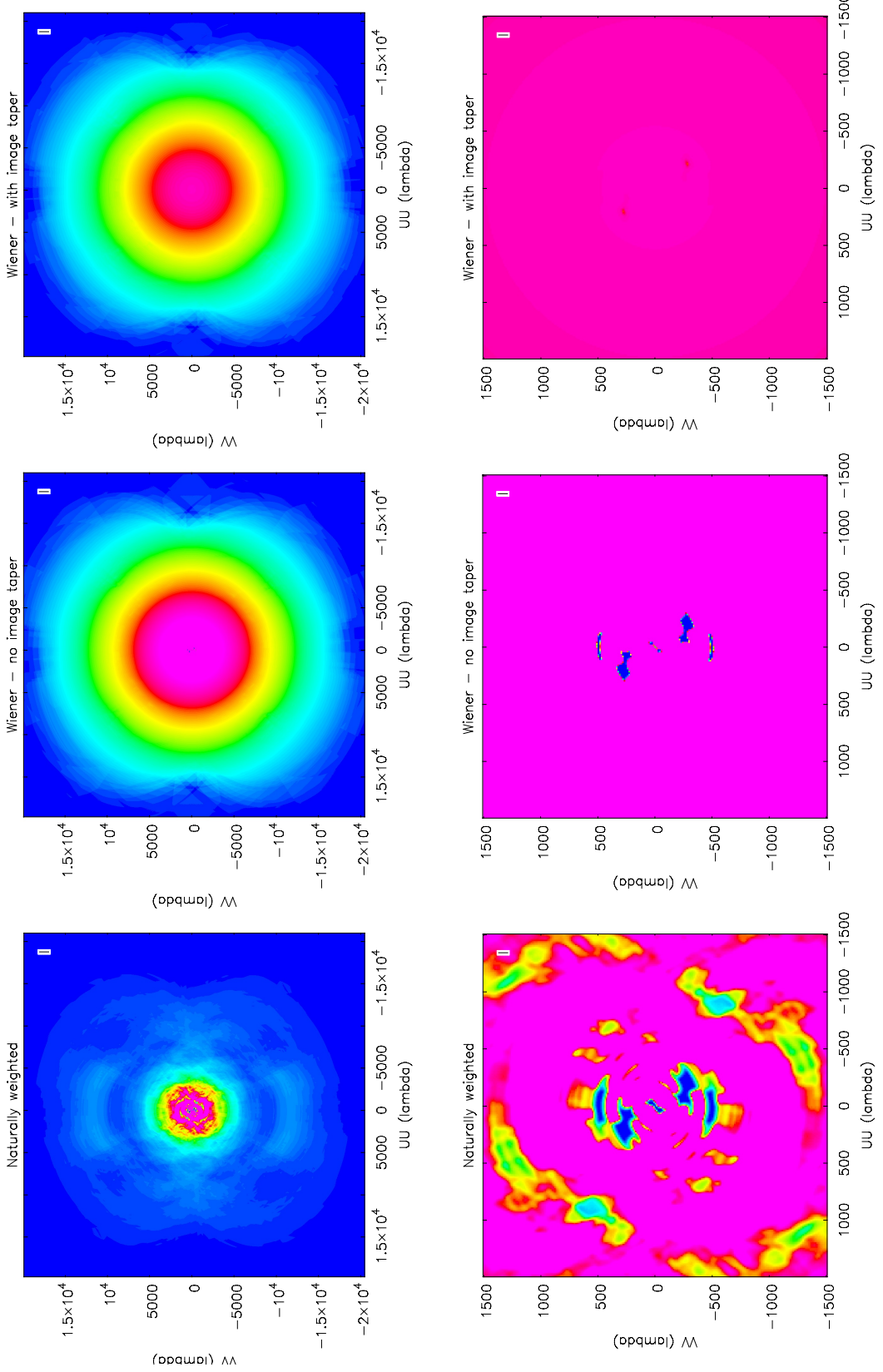


Figure 27: Transform of point spread function before and after preconditioning. The bottom row shows the zoomed in version of the same plots. The left column is naturally weight, the centre is preconditioned without the primary beam weighting, and the right column is with primary beam weighting. The colour scale is the same for each panel, and chosen to highlight the inner structure.

Turning now to the multi-scale aspect, each of these images is decomposed using a multi-scale basis. Specifically, the image is represented by a set of blobs of various scale sizes. The blobs are parabolic in radius with a prolate spheroidal wavefunction multiplied in to taper the edge of the parabola. This is the same multi-scale basis as Multi-Scale Clean [3]. The set of Taylor dirty images and point spread function is expanded by convolution with blobs of given scales. The search is then conducted over all the scale images. A standard set of scales is used: [0, 10, 30, 90] pixels.

The spectral index can then be estimated from the Taylor term images by straightforward algebra.

As discussed in Section 2.11, there is an alternative continuum imaging mode where the multi-frequency synthesis is not used. Instead, a number of channels (between $\sim$ 30 and 300) are kept and deconvolved separately using just the multi-scale algorithm.

3.8.8 Deconvolution: Spectral Line Imaging at 30" resolution

For spectral line imaging at 30" resolution, the minor cycle deconvolution only proceeds once, and on the brightest sources only. The Multi-Scale Clean [3] is used for this deconvolution. A standard set of scales is used: [0, 10, 30, 90] pixels.

3.8.9 Deconvolution: Spectral Line Imaging at 10" resolution

For the large 6km configuration, imaging of the entire field with all spectral channels is out of scope and is therefore not possible. However, imaging of numerous small regions as *postage stamps* is in scope and will be possible. The limitations are first that the 10" and 30" resolution images cannot be made simultaneously. and second that imaging is limited to a moderate number of cubes, between 100 and 200, of size typically 256x256x512. The channel range must be centered on a particular channel which must be known beforehand.

It may be possible to raise these limits in future as we gain experience with the telescope and computing. We are particularly aware that processing the full spectral range is required for some science projects.

3.8.10 Deconvolution: Transient Imaging

For transient imaging, the major challenge is to meet the latency requirements - an image within 10s (required) or 5s (desirable). Deconvolution is not possible within this time scale but the initial sky model is removed from the visibility data before imaging, thus reducing the peak brightness and required dynamic range. Using the usual w projection gridding algorithm is not possible at this level of latency. Instead the snapshot approach is used.

3.9 Specific imaging issues

3.9.1 Doppler correction

Observations take place in a topocentric reference frame, which is a non-inertial reference frame fixed with the Earth. Astronomical analysis normally assumes a barycentric or local standard of rest (LSR) reference frames, which can be considered inertial for practical purposes. The main consequence of non-inertiality is the need for a Doppler correction in frequencies and velocities if the spectral resolution is fine enough.

This correction is time- and position-dependent. The diurnal rotation of the Earth amounts to a 0.45 km s^{-1} velocity drift (with the period of 1 day) at the ASKAP site (an equatorial field is the worst case). The Earth's orbital motion gives the annual period. The velocity drift amplitude is about 30 km s^{-1}, which amounts to a 0.52 km s^{-1} drift per day (a field in the ecliptic plane is the worst case). These figures characterise the time variation. For sufficiently short observations, there is no need to do the correction on-the-fly. Instead, it can be done through either regridding of the final image cube or adjusting the coordinate system (the regridding is still necessary if two such cubes taken at different times are combined together). The direction dependence of the effect gives an additional issue specific to the wide-field of view regime: differential velocity shift. The worst-case scenario is observations of the field towards the Sun or in the opposite direction. In this case, the 30 deg^2 field of view of ASKAP translates to the error of 1.3 km s^{-1} near the image edges.

The standard spectral line setup of ASKAP implies the spectral resolution of around 18 kHz, which is equivalent to 3.9 km s^{-1} at 1.4 GHz. Therefore, provided we do not integrate longer than a few days in one go (which makes sense because most fields rise and set), no Doppler correction of any form is required during imaging. The spectral coordinate of the resulting image cube will be adjusted to simplify merging several such integrations together and cataloguing of detections; the differential velocity shift can be ignored. This approach can only be used if the spectral resolution is worse than approximately 6 kHz (or 2.4 kHz if a single regridding of the final cube is done to take care of the differential shift). However, the zoom mode of the correlator will provide much higher spectral resolution well below this limit. Therefore, some form of the on-the-fly Doppler correction will be required during imaging. There are two possible approaches, the relative merits of each are still the subject of investigation:

1. Add the spectral domain to the convolution function used for imaging and combine spectral regridding with the normal convolutional gridding. This is the most accurate approach, but it results in a significant increase of the number of gridding operations and, therefore, most likely is not affordable.

2. Regrid the visibility spectrum using some from of polynomial interpolation prior to gridding. This is the option which is most likely going to be implemented. Initially, the correction is likely to be as simple as the nearest match (i.e. nearest neighbour gridding) because it has virtually no perfromance penalty and does not couple individual spectral planes together.

3.9.2 Treatment of polarisation

The array acquires data in a linear polarisation frame specific for this particular instrument due to the feed design. The instrumental polarisation frame is the natural frame for calibration. However, astronomical analysis is normally done in the Stokes frame (I, Q, U and V), and so is the definition of the sky model. Therefore, the synthesis code requires polarisation conversion capability. We do this conversion per visibility, which implies it is the same for the whole field of view of the given beam. The conversion formula follows [17]

$$\begin{pmatrix} V_I \\ V_Q \\ V_U \\ V_V \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \end{pmatrix} \begin{pmatrix} V_{xx} \\ V_{xy} \\ V_{yx} \\ V_{yy} \end{pmatrix} \tag{38}$$

There will be an additional factor of 0.5 (accessible via a configurable option of the pipeline) to match the stokes-I definition used in *casa* and *miriad* (see [18] for a review of existing controversies).

The conversion formula of Equation 38 can be modified to correct for effects of ionospheric Faraday rotation on the fly (see [17] for an explicit definition of the appropriate Müller matrix). Therefore, provided

a model of the ionosphere is available at the time of imaging (e.g. in the form of a time series of Rotation Measures per beam) it can be taken into account at a relatively little cost provided the required time resolution is not better than the correlator integration time of 5 seconds. The limitation of a single Rotation Measure per beam (roughly 1 degree in the sky) is imposed by the fact that we correct per visibility. A finer spatial scale correction would require a matrix multiplication similar to Equation 38 to be merged with the convolutional gridding, or pixel-wise linear combinations of snapshot cross-product images. In general, the former would require a separate convolution function per cross-product and up to 16 times more gridding operations, while the latter would require a higher snapshot rate than otherwise needed and in practice the ionospheric update rate would be limited to several minutes or more. They would, however, relax the strict instrumental polarisation purity requirements implicit in Equation 38. Both options would be very challenging given the current estimates of the available processing resources. It is worth noting that a similar functionality is needed to correct for the PAF's response accurately and to achieve the best polarisation purity across the wide field of view. Some further research is needed in this area to find appropriate approximations and make such corrections affordable.

3.9.3 Rotation Measure Synthesis

Once a Stokes cube has been made, it is possible to then extract full-Stokes spectra and perform Faraday Rotation Measure Synthesis [19] to estimate the rotation measure imparted on the spectrum. The pipeline will perform RM Synthesis on compact components identified from the Stokes-I continuum image. The full-Stokes spectrum for each component will be extracted by summing (in each channel) over an $n \times n$ pixel box and normalising the result by the sum of the synthesised beam over the same area.

The Faraday dispersion function (FDF), is given by Eq. 25 of [19]:

$$\tilde{F}(\phi) = K \int_{-\infty}^{+\infty} \tilde{P}(\lambda^2) e^{-2i\phi(\lambda^2-\lambda_0^2)} d\lambda^2 \tag{39}$$

where ϕ is the Faraday depth, λ the wavelength of observation, and $\tilde{P}$ is the observed (complex) polarised intensity, that has been multiplied by the windowing function $W(\lambda^2)$. This windowing function generates the normalisation factor K via

$$K = (\int_{-\infty}^{+\infty} W(\lambda^2) d\lambda^2)^{-1}. \tag{40}$$

Since Faraday rotation depends on the square of the wavelenth, the sampling in the data used in calculating the FDF is irregular and often incomplete. The rotation measure spread function (RMSF), which is analogous to the point spread function in interferometric imaging, can be calculated as follows:

$$R(\phi) = K \int_{-\infty}^{+\infty} W(\lambda^2) e^{-2i\phi(\lambda^2-\lambda_0^2)} d\lambda^2 \tag{41}$$

The pipeline approach is to perform RM Synthesis only on the spectra of identified Stokes-I components and produce a dirty FDF function and RMSF, along with identification of the peak Faraday depth and measures of its quality (i.e. is it resolved in Faraday depth, is the spectrum more complex than a single peak, and so on). These measurements will be recorded in a polarisation catalogue that will be stored in CASDA. See POSSUM report [20] for further details. Further processing of the FDF, through techniques such as RM Clean, is out of scope for the pipeline, as is RM Synthesis for all spatial pixels in the imaged area (as opposed to identified components).

3.9.4 Combination with single-dish observations

For fields with a large amount of large-scale diffuse emission, recovering all the signal is only possible with the inclusion of data from single-dish observations. This is particularly the case for observations of diffuse Galactic emission. There are two principle ways this can be done:

- Use the single-dish image as the starting point for the deconvolution (i.e. the initial model).
- Combine the cleaned image with the single-dish image after deconvolution

GASKAP has been examining the optimal approaches in the context of diffuse Galactic HI emission [21], and find that multi-scale clean followed by combination with single-dish data gives the best results, and that using the single-dish image as the starting model is not necessary. This post-deconvolution approach is the one that we plan to follow, unless there are situations where it does not perform adequately. We note that the case studied by the GASKAP team, featuring a large amount of diffuse Galactic HI emission, is perhaps the hardest situation to deal with, giving us confidence that the approach will work in situations with less diffuse emission.

3.9.5 Combination of multiple images

Once the imaging pipeline has finished, there are cases where the images will need to be combined with earlier images.

The first is for northern fields, where a single day does not provide enough integration time to reach a survey's requirement. In this case, the field will need to be observed over multiple days until the required integration time is reached. In this case, provided the observations are done close in time (for instance, on successive days), storage of the UV data will be possible and regular imaging can be done as normal once all data has been taken.

The second case is for deep integrations. These are fields that require perhaps 100s of hours, built up of many observations, potentially spanning months of actual time. In this case, storage of the UV data is unfeasible (at least, in the spectral-line case), and so combination of the data will need to done using the images themselves. The exact details of how this combination is done are still an area of research, specifically to determine the optimal combination technique, and how to deal with RFI removal and continuum subtraction at fainter levels than necessary in typical 8-12 hr integrations.

3.10 Source Finding

3.10.1 Overview

The images and cubes produced by the processing pipeline are passed through source finding as a final step before transmission to the ASKAP Science Data Archive. As discussed in Section 2.3, the data sizes make pipeline-based source extraction the most efficient way of generating catalogues, as the images are already in memory or on the scratch disks, and so accessible to the distributed processing pipeline.

The aim of the pipeline-based source extraction is to provide catalogues of sources from the image data that are suitable for ingestion into CASDA. The source extraction will be able to search for sources in all the image products produced by the pipeline, creating catalogues of continuum sources, spectral-line (that is, emission-line) sources, and absorption-line components.

The ASKAPsoft source-finder builds on the DUCHAMP 3D source finding package [23] [24] [26], which provides the infrastructure for locating and isolating areas of interest within the image, and performing basic parameterisation of the detected sources. There is additional functionality added to the source-finder (originally described in [25]) that improves the performance, particularly in a distributed computing environment, and provides parameterisation that is appropriate for the specific ASKAP survey requirements. The following sections detail some of the specific approaches used in the source finding.

It is understood that source finding is an area of ongoing research amongst the Survey Science Teams, and so the nature of the algorithms could potentially change. The Survey Science Teams are invited to provide specifications of desired changes, and these will be incorporated into the source finding code by the ASKAP computing team. There will be periodic calls for specifications, followed by periods of implementation and testing.

3.10.2 Distributed processing

Due to the large sizes of the images, the source finding will be done in a distributed manner. The image is split into subimages, using a regular grid in spatial and spectral directions. Each subimage is searched and an object list is produced. The master process then combines these object lists to form the final source catalogue.

To account for sources that lie on the boundaries, the subimages can be defined with configurable overlap. Essentially, the regions of overlap are searched multiple times, and sources that lie (at least partly) therein are passed to the master process to be combined. See Section 3.10.6 for details on the fitting of these sources.

3.10.3 Pre-processing the image

To improve the detectability of sources, the data may be processed prior to searching, to increase the apparent signal-to-noise of real features. This processing can be some form of smoothing, or multi-resolution wavelet reconstruction via the à trous algorithm.

The smoothing can be done in one dimension (that is, in the spectral direction) with a hanning filter, or in two dimensions (spatially) with a 2D Gaussian kernel. This is useful if one wants to highlight a particular scale or if many of the expected features are the same size, for instance the minimum size of a galaxy.

To accurately deal with a range of scales, the multi-resolution reconstruction is used. This constructs wavelet arrays, being the difference between smoothed versions of the data at different scales, then applies a threshold on these, keeping only signal above some level. The motivation of this is to keep real structure in the dataset, but remove the random noise signal. The source-finding is then performed on the reconstructed array. If a signal-to-noise threshold is required, the residual from the reconstruction is used to find the noise properties.

3.10.4 Threshold Determination

The first step in finding sources is to determine a threshold. We use a hard threshold, where pixels with flux values above this level are considered “detected”, and those below are “background”. The threshold in the stand-alone DUCHAMP is a single threshold for the entire dataset, and is specified as either a flux value or a signal-to-noise value.

In the latter case, the signal-to-noise is determined from the image using robust methods to estimate the noise background – measuring the median and then the Median Absolute Deviation From the Median (MADFM), then correcting by a fixed factor to get the equivalent standard deviation for a Normal distribution. The mean and standard deviation can be used as an alternative, although these will be biased by the presence of bright pixels (i.e. the signal that is being sought).

In the distributed case, a flux threshold is easily applied to all subimages. A single signal-to-noise threshold, however, can only be obtained by first combining the mean/median of each subimage, then using that to estimate the overall spread by finding the weighted average of all standard deviations of the subimages.

There are limitations in the single-threshold technique. It is usually the case that the sensitivity varies across the field of view of an image, and so the averaged noise across the image will be a compromise between the large noise at the edges and low noise in the centre. A single threshold will then lead perhaps to more spurious detections at the edges, and more missed detections in the centre. A flexible threshold, that depends on the local noise only, would thus seem preferable (although at the expense of creating a catalogue of varying depth).

The ASKAPsoft implementation has such an option, where a sliding box defines the local region. This box is used to find the median and MADFM of pixels local to a given point. This is currently only implemented for 1D and 2D cases, which correspond to how the searches are done (that is, for 3D data the searching is broken up into a series of 1D or 2D searches).

Alternatively, the pixel values can be scaled by the normalised sensitivity (in practice, the square-root of the normalised weights image), so that a given threshold that produces good results where the sensitivity is good does not get overwhelmed by spurious sources at the edges where the sensitivity is worse. Note that this scaling is **not** taken into account when fluxes etc are calculated for detected objects – it is only used in finding the *locations* of objects (see 3.10.5). Note that in the eventual ASKAP pipeline, the actual sensitivity image will be available, and so this could be used to determine the appropriate threshold at each point (in the case of a signal-to-noise threshold).

3.10.5 Source Detection

The ASKAPsoft implementation takes most of the basic functionality of the stand-alone package, and implements it in a distributed manner under a master-worker scheme. The image may be split up into a user-defined number of subimages, regularly partitioned in both spatial and spectral directions. Each subimage is then searched for sources (and the following descriptions apply for the subimages as well as for a full image), with the final source lists sent to the master where they are combined to form the full list.

Objects are detected in the image on a per-pixel basis. Each pixel is considered separately, a threshold is applied, and pixels brighter than this threshold are assigned to an object. The three-dimensional nature of a spectral cube is dealt with in one of two ways. Objects can be built up in two dimensions through the algorithm of [22], with objects on different channel images being merged according to prescribed rules (typically if they are within a certain distance, or even adjacent). Alternatively, the searching can be done in the one-dimensional spectra, with the same sort of subsequent merging.

Once objects have been found, they can be "grown" to a lower, secondary threshold, to better gauge their full extent. This results in finding all groups of pixels brighter than this secondary threshold, provided they have at least one pixel (or some larger, user-defined number) that is brighter than the primary threshold. This growing is important for the parameter calculations described in 3.10.6, since only the detected pixels are used. A graphical example of the use of growing, as opposed to simply using the lower threshold as the primary, can be seen in Figure 28.

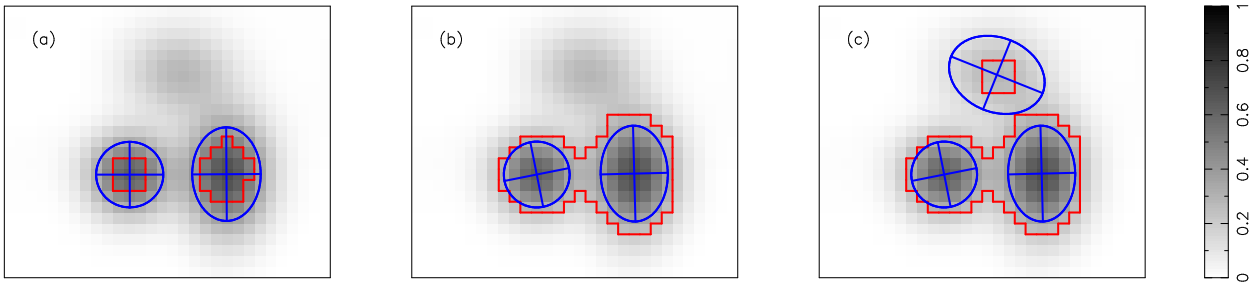


Figure 28: Examples of different methods of thresholding. a) A threshold of 0.5 is used, detecting the peaks of two of the three sources present. b) The same detection threshold, but grown to 0.25. The upper source is missed as it has no pixel above 0.5, while the bottom two have merged. c) A primary threshold of 0.25 is used, detecting all sources (again, with the bottom two merged). In all cases, the detected objects are enclosed by red boundaries, and the results of Gaussian fitting are shown as ellipses.

3.10.6 Source Fitting and Parametrisation

The DUCHAMP library provides a number of basic measurements of source parameters. There is no profile fitting or similar built into DUCHAMP, so these parameters are measured just from the detected pixels. The ASKAPsoft finder extends this to adaptively grow detected sources out to the noise level, so that the best measurement of the entire flux can be made without profile fitting.

Since DUCHAMP was designed with 3D spectral-line data in mind, the parameters include measures of velocity width (full, 20% and 50% widths), as well as integrated & peak fluxes, positions and spatial extents (both in world and pixel coordinates).

For a lot of continuum-source analysis, it is common to decompose sources into a series of 2D Gaussian components. This facility has been incorporated into the ASKAP software. The approach taken is similar to that used in NVSS & FIRST, where multiple Gaussians can be fitted to a source and accepted based on a number of criteria. The component catalogue is one of the key outputs of the continuum source finding. The Gaussian fitting can be a good way to decompose blended objects - for examples, refer to Figure 28.

When the image is distributed (see 3.10.2), the master process first combines any sources that fall into the overlap regions. The fitting to the resulting sources is then done in a distributed manner by farming out sources one at a time to available processors.

When a continuum image has been produced through MSMFS, the first and second Taylor terms can be examined to find the spectral index & curvature of each source. This uses the components from the fit to the total-intensity map, placing them at the same location in the Taylor term map and fitting the height of the Gaussian only (keeping the location and shape the same as that fitted to the total-intensity). The total flux of the resulting Gaussian is then used to calculate the spectral index and/or curvature.

3.11 Quality evaluation

The final step that is performed on the data before they are sent to the archive is a set of automated quality evaluation tasks. These are checks made on the quality of the image data, and the quality of the extracted source catalogues.

The details of this evaluation are, at time of writing, still an area of development. They are likely, however, to include the following things:

- Overall noise level and noise variation within an image.
- Effectiveness of sky-model and continuum subtraction (looking at residuals on and off bright sources).
- Distribution of object sizes and orientations (are there any systematic effects present?).
- Source counts as a function of flux density.
- Correlations between extracted sources and the beam pattern.
- Detected sources close to known bright objects (i.e. sources that are potential artefacts).

In a similar vein, it is expected that catalogue sources will have a set of quality flags associated with them to indicate potential problems or things to be aware of. For instance, sources near the edge of the field, or with a poor-quality fit, or identified as a possible/likely artefact. Transient or variable sources, identified as such by the transient pipeline (Section 2.12), will also be flagged.

4 SCIENCE ARCHIVE

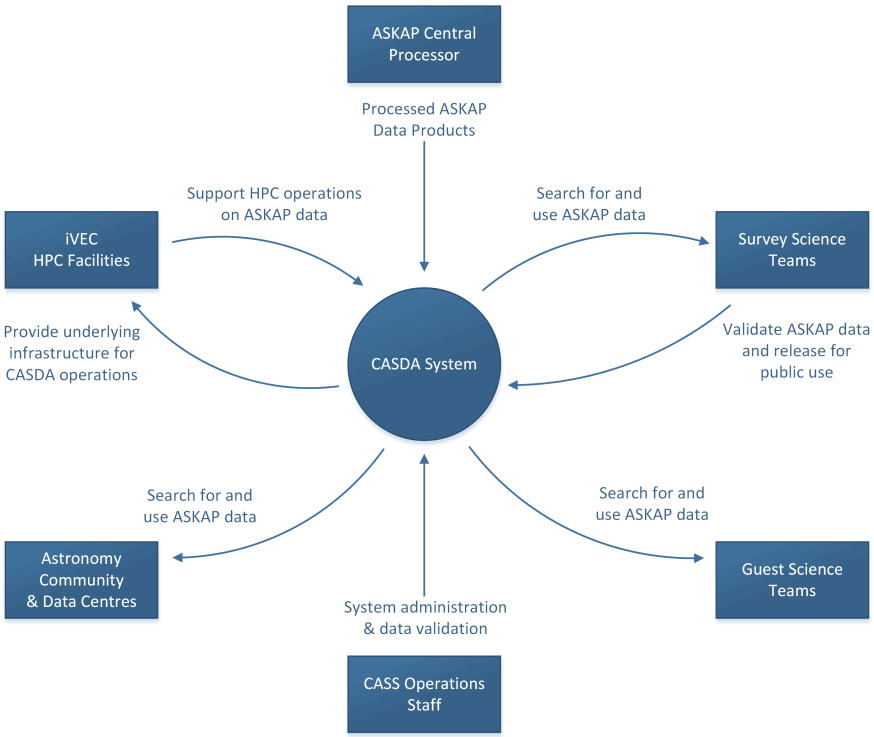
The CSIRO ASKAP Science Data Archive (CASDA) is the prime repository and access point for level 5 and level 6 data products (see Figure 3). These products include calibrated visibility data, images and cubes, and source catalogues. The archive can be queried either via a web page or by accessing the Virtual Observatory (VO) services provided by the archive using a VO-compliant client application.

All data products are accompanied with metadata that describes the data in sufficient detail that searches, queries, and selections may be performed. Examples of metadata are time of observation, field direction, frequency setup, scheduling block identification, *etc.*.

- Catalogues may be queried directly and results returned in a VOTable using either a VO Cone Search service or the VO Table Access Protocol service.
- Images (including cubes) possess metadata that may be queried and thus selected.
- The actual images or image cutouts may be accessed using the VO Simple Image Access Protocol, and will be returned as standard FITS image files.
- Visibility data are stored as compressed Measurement Sets. The metadata for the visibility data may be queried and the data downloaded for local processing.

The capabilities of the science archive are:

- Data ingest from ASKAP as data products become available.
- Facilitating data discovery and access
  - Archive queries via a full-featured web interface catering for data search, data access, administration and management features, using CSIRO's RDS Data Access Portal (DAP) where possible.
  - High performance access to a Hierarchical Storage Management (HSM) system. The HSM keeps frequently accessed data on low-capacity high-speed media e.g. memory and/or disk, and infrequently accessed data on high-capacity low-speed media e.g. tape.
  - A VO Cone Search service to enable spatial searches within a catalogue around a specified sky position.
  - A VO Simple Image Access Protocol (SIAP) service to enable searching for archived images around a user-specified sky position. The service can optionally generate a cut-out image with user-specified dimensions. The service returns a link to the requested image/cut-out.
  - A VO Table Access Protocol (TAP) service to enable more advanced queries of archive data.
  - A VO Registry service to propagate a list of available CASDA VO services.
  - Support transfer of data to Pawsey compute facilities.
- Allow level 7 *catalogue based* data products to be deposited by survey science projects.
- Employ an authentication and authorisation security model which promotes collaboration across research institutions.



CASDA Context Model			
v1	Initial version	9/7/2013	Dan Miller
v2	Minor clarification of data flow directions	11/7/2013	Dan Miller
v3	Updated 'General Astronomy Community'	29/1/2014	Jessica Chapman

Figure 29: CSIRO ASKAP Science Data Archive (CASDA) architecture and data flow. Illustrating the relationship between CASDA, and other entities at a very high level. It also represents inputs and outputs between the system and each entity

Additional science processing may be performed on level 5 and level 6 data products using auxiliary information (e.g. redshifts, multi-wavelength data, morphological classifications, etc.) to create level 7 data products. The implementation, execution and outputs of the additional science processing is the responsibility of the science survey teams.

5 POST-PIPELINE SCIENCE ANALYSIS

All data products that come out of the ASKAP pipeline are classed as level 5 (see Figure 3). These are fully calibrated and suitable for science analysis. Data that forms part of a Survey Science Project then goes through a process of validation, for which the Survey Science Team has responsibility (see Figure. 29).

The validation process involves setting a flag in the metadata associated with the image product or catalogue, moving it from level 5 to level 6, thereby making it publicly available. The details of the validation process are, at time of writing, still the subject of development. The online pipelines will provide as much quality analysis information as practical, but some post-processing may be required on the part of the Survey Science Teams. This would take place on a platform other than the central processor.

It is also expected that the Survey Science Projects will require additional science analysis beyond the pipeline processing. To this end, elements of the processing pipelines can be run in an offline state. An important example of this is stitching of neighbouring fields to produce the final survey-quality image of the sky. Following stitching, source extraction will need to be performed in the overlap regions. The exact way the overlap regions will be processed is still to be determined. Possibilities include:

- No stitching is done in the ASKAP pipeline, and the Survey Science Teams take responsibility for this post-processing. Under this model, the level 5 catalogues produced would be solely from individual fields. Care would need to be taken to ensure that the subsequent source extraction from overlap regions is consistent with the level 5 catalogues (i.e. from the non-overlap regions).
- The stitching and subsequent source-finding is done in the ASKAP central processor, in offline mode. The source-finding would then be done in the same way as for the rest of the catalogue, helping the uniformity of the resulting catalogue. The details of how this would be accomplished still need to be worked out (e.g. when it occurs, how much of the image data is combined).

Work to identify the optimal approach here is currently ongoing. Further examples of post-pipeline analysis include improved or alternative source finding approaches, source finding in deep integrations (which would need to be done in the offline state once the deep cubes are created) or stacking using new sets of positions.

6 DEFINITIONS OF STANDARD DATA PRODUCTS

This section discusses the definitions of standard data products and the rationale for the choices. ASKAP data products are designated by a standard code:

product.mode.type.subtype.baseline.level

Not all of these need be present in which case the two appropriate dot's touch. The allowed values of these fields are given in tables 5, 6, 7. First we give some examples:

MS.Spectral.processed..6km.5 Measurement Set containing spectral line observations including 6km baselines

Image.Continuum.residual.moment0.2km.5 Image containing moment0 of spectral expansion including 2km baselines

Catalogue.Spectral...2km.5 Catalogue containing spectral line catalog including 2km baselines

Table 5: ASKAP visibility products

ID	Description
product	MS
mode	Continuum — Spectral
type	raw — processed
subtype	NA - leave blank
maximum baseline	2km — 6km
level	4 — 5 — 6 — 7

Table 6: ASKAP image products (data products such as PSF and sensitivity images are not shown).

ID	Description
product	Image
mode	Continuum — Spectral
type	model — restored — residual
subtype	moment0 — moment1 — moment2 — cube
maximum baseline	2km — 6km
level	5 — 6 — 7

Table 7: ASKAP catalogue products

ID	Description
product	Catalog — Lightcurve
mode	Continuum — Spectral
type	NA - leave blank
subtype	NA - leave blank
maximum baseline	2km — 6km
level	5 — 6 — 7

Images are constructed with default parameters. These are specified in Figure 30. This figure serves two purposes: one is to indicate the data sizes necessary to do the processing, while the other is to indicate the implications for the survey data products. The former case requires images to be formed that are larger than is useful for scientific purposes – the additional space allows more accurate imaging. The latter case indicates storage averaged over an entire survey. Multiple visits to a single field are included in the number of fields row. The numbers here reflect proposed survey configurations (subject to using 16200/300 as the number of fine/coarse channels respectively), and do not indicate a commitment to produce or store all such data. Note, in particular, that the 10" postage stamp processing is challenging (see Section 3.8.9).

Array Parameters	**DINGO**	**EMU**	**FLASH**	**GASKAP**	**POSSUM**	**VAST**	**WALLABY-2**	**WALLABY-6a**	**WALLABY-6b**	
Number of antennas	36	36	36	36	36	36	36	36	36	Use 6 for BETA, 30 for HI, 36 for continuum/transients
Max baseline length [km]	2	6	6	2	6	6	2	6	6	Use 2 for HI, 6 for continuum/transients
Number of baselines	666	666	666	666	666	666	666	666	666	Including autocorrelations
Number of beams (feeds)					36					
Frequency channels from correlator	16200	16200	16200	16200	16200	16200	16200	16200	16200	Maximum number of channels at correlator output
Number of polarizations	4	4	4	4	4	4	4	4	4	Number of polarizations from correlator
Frequency channels after averaging	16200	300	16200	16200	300	30	16200	16200	16200	
Data sizes and rates										
Bits per complex sample, weight					72					Bits need to store one complex sample plus associated weight
Raw visibility frame [Gbytes]	13.98	13.98	13.98	13.98	13.98	13.98	13.98	13.98	13.98	Fast transients are a special mode
Number of polarizations	4	4	4	4	4	2	4	4	4	2 pol cases are I and V
Integration time [sec]	5	5	5	5	5	5	5	5	5	Typical integration time (to within first sidelobe) - not more than 15s
Data ingest rate [GB/s]					8					
Time to ingest one frame					1.75					
Data rate [Gb/s]	***22.37***	***22.37***	***22.37***	***22.37***	***22.37***	***22.37***	***22.37***	***22.37***	***22.37***	Data rate from correlator to computer, prechannel averaging (note Gbits/s)
Averaged visibility frame [Gbytes]	*13.98*	*0.26*	*13.98*	*13.98*	*0.26*	*0.01*	*13.98*	*13.98*	*13.98*	
Averaged data rate [Gbyte/s]	*2.797*	*0.052*	*2.797*	*2.797*	*0.052*	*0.003*	*2.797*	*2.797*	*2.797*	Rate of data for selected polarizations (note Gbytes/s)
Averaged data rate [Tbyte/h]	*10.07*	*0.19*	*10.07*	*10.07*	*0.19*	*0.01*	*10.07*	*10.07*	*10.07*	
Observation time [hrs]	8	12	2	12.5	8	0	8	8	8	Observation time during which data are to be retained
Averaging time [s]	5	5	5	5	5	5	5	5	5	Averaging time after processing
Averaged visibility data set [Tbytes]	***80.54***	***2.24***	***20.14***	***125.85***	***1.49***	***0.00***	***80.54***	***80.54***	***80.54***	Size of averaged data set at end of observation
Image sizes for processing										
Number of image pols	1	4	1	1	4	1	1	1	1	
Image frequency channels required	16200	1	16200	16200	300	30	16200	512	16200	WALLABY-6a number allows for typical size in frequency
Field of view [degrees]	7.5	7.5	*0.089*	7.5	7.5	7.5	7.5	*0.178*	*0.089*	
Cellsize [arcsec]	7.5	2.5	2.5	7.5	2.5	2.5	7.5	2.5	2.5	4pix/beam, for 6km beam of 10" and 2km beam of 30"
Image size [pixels]	*3600*	*10800*	*128*	*3600*	*10800*	*10800*	*3600*	256	128	Choosen for 7.5 degree FOV, unless in postage stamp mode
Image size [degrees]	*7.5*	*7.5*	*0.089*	*7.5*	*7.5*	*7.5*	*7.5*	*0.178*	*0.089*	
Total image size [Gbytes]	*840*	*1.866*	*1.062*	*840*	*560*	*13.997*	*840*	*0.134*	*1.062*	
Number of images per field per pol	3	13	450	3	3.5	5760	3	1050	3000	See caption for explanation
Image data per field [TB]	**2.52**	**0.024**	**0.48**	**2.52**	**1.96**	**80.62**	**2.52**	**0.14**	**3.19**	
Survey sizes										
Image size [pixels]	2600	7800	40	2600	7800	7800	2600	256	40	Cropped from the processed image if necessary
Total image size [Gbytes]	*438*	*0.973*	*0.104*	*438*	*292*	*7.301*	*438*	*0.134*	*0.104*	
Fields per survey	966	1200	850	644	1200	1200	1200	1200	1200	Number of fields - overlapping at 4.5 degree separation
Survey size (images) [PB]	***1.27***	***0.015***	***0.04***	***0.85***	***1.23***	***50.46***	***1.58***	***0.17***	***0.37***	
Survey size (vis) [PB]	**77.80**	**2.68**	**17.11**	**81.04**	**1.79**	**0.00**	**96.65**	**96.65**	**96.65**	

Figure 30: Parameters for image products for each survey. The high-resolution postage stamp options for WALLABY are also listed explicitly. The number of images for the various surveys are as follows: EMU – restored, residual, model image and PSF for each of three Taylor terms, plus sensitivity image; POSSUM – restored, residual and model images for each of 4 Stokes, plus PSF and sensitivity; DINGO/GASKAP/WALLABY-2 – image, PSF, sensitivity; FLASH/WALLABY-6a/-6b – 150/350/1000 postage stamps per field respectively, each with image, PSF and sensitivity; VAST – one image every 5sec over 8hrs. Note that the quoted data sizes are indicative of the data rate only, and do not represent a commitment to store all such data.

References

[1] Rau, U., Bhatnagar, S., Voronkov, M.A., Cornwell, T.J. (2009). Advances in Calibration and Imaging Techniques in Radio Interferometry. Proceedings of the IEEE, Vol 97(8), 1472-1481.

[2] Cornwell T. J., Perley R. A., (1992), Radio-interferometric imaging of very large fields – The problem of non-coplanar arrays, A&A, 261, 353

[3] Cornwell, T.J. (2008), Multiscale CLEAN Deconvolution of Radio Synthesis Images, IEEE Journal of Selected Topics in Signal Processing, Vol 2(5), 793-801.

[4] Rich, J., et al. (2008), Multi-Scale CLEAN: A Comparison of its Performance Against Classical CLEAN on Galaxies Using THINGS, Astronomical Journal, Volume 136, Issue 6, pp. 2897-2920

[5] Rau. U. (2010), Parameterized Deconvolution for Wide-Band Radio Synthesis Imaging, PhD thesis, Department of Physics, New Mexico Institute of Mining and Technology, Socorro, NM, USA.

[6] `http://code.google.com/p/casacore`

[7] Briggs, D.S. (1995), Deconvolution of moderately resolved Sources, PhD thesis, Department of Physics, New Mexico Institute of Mining and Technology, Socorro, NM, USA.

[8] Bahren, L., (2010), LOFAR Sky Model, LOFAR-USG-ICD-004,
`http://usg.lofar.org/wiki/doku.php?id=documents:icd:lofar-usg-icd-004#icd_lofar-usg-icd-004`

[9] Lenc, E. (2010), `http://www.atnf.csiro.au/people/Emil.Lenc/ASKAP/psf/sim/view.html`

[10] Gupta, N., et al. (2008), The Initial Array Configuration for ASKAP, ASKAP-SUP-0003.

[11] Gupta, N., et al. (2008), ASKAP Array Configurations: Technical Studies, ASKAP-SUP-0004.

[12] Bunton, J., (2009), ASKAP System Architecture, ASKAP-SEIC-0003.

[13] Brown, A.J. et al. (2014), Design and implementation of the 2nd Generation ASKAP Digital Receiver System, 2014 International Conference on Electromagnetics in Advanced Applications (ICEAA), 268

[14] Hampson, G. et al. (2012), ASKAP PAF ADE - advancing an L-band PAF design towards SKA, 2012 International Conference on Electromagnetics in Advanced Applications (ICEAA), 807

[15] Cornwell T.J., Uson J.M., Haddad N., (1992), Radio-interferometric imaging of spectral lines – The problem of continuum subtraction, A&A, 258, 583

[16] Sault R.J., (1994), An analysis of visibility-based continuum subtraction, A&ASS, 107, 55

[17] Hamaker J.P., Bregman J.D., Sault R.J. (1996), Understanding radio polarimetry. I. Mathematical foundations, A&AS, 117, 137

[18] Smirnov O.M. (2011), Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism, A&A, 527, A106

[19] Brentjens M.A. & de Bruyn A.G. (2005), Faraday rotation measure synthesis, A&A, 441, 1217

[20] Purcell, C. (2014), POSSUM Report #62: Integration of the POSSUM analysis pipeline for BETA

[21] McClure-Griffiths, N., McConnell, D., & Dawson, J. (2011), GASKAP WG1 Report #2

[22] Lutz, R.K., (1980), An algorithm for the real time analysis of digitised images, The Computer Journal, 23, 262

[23] Whiting, M.T. (2008), Astronomers! Do you know where your galaxies are?, in "Galaxies in the Local Volume", Eds. B.Koribalski & H.Jerjen, Astrophysics & Space Science Proceedings, Springer, p.343-344

[24] Whiting, M.T. (2012), DUCHAMP: a 3D source finder for spectral-line data, MNRAS, 421, 3242

[25] Whiting, M. & Humphreys, B. (2012), Source-Finding for the Australian Square Kilometre Array Pathfinder, PASA, 29, 371

[26] http://www.atnf.csiro.au/people/Matthew.Whiting/Duchamp