ASTRONOMY AND SPACE SCIENCE
www.csiro.au

ASKAP Pilot Survey Plan

Aidan Hotan
ASKAP lead scientist
21 September 2018

A description of the ASKAP operations model for 2018 – 2020
Version 1.0

CSIRO Astronomy and Space Science
PO Box 1130, Bentley Delivery Centre, WA 6983, Australia

Copyright and disclaimer

© 2009 CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

1 Background

The Australian Square Kilometre Array Pathfinder (ASKAP) is the latest CSIRO Astronomy and Space Science (CASS) national facility radio telescope. It was designed to perform rapid surveys using a combination of Phased Array Feed (PAF) technology to achieve a large field of view, a large number of small antennas to provide excellent spatial frequency sampling, and antenna mounts with three axes of rotation to keep the orientation of the reflecting structure fixed on the sky, leading to high dynamic range.

ASKAP's digital systems were designed to process a wide bandwidth, currently 288 MHz[1], which can be tuned within the range 700–1800 MHz. This, combined with the large number of beams and antennas, ensures that ASKAP produces what is currently considered to be a very large amount of raw visibility data. The incoming rate is 2.5 GB per second when using a 5 second cycle time, which results in 108 TB of data for a 12 hour observation.

ASKAP's original design concept dealt with this data rate by producing images in real-time, using on-the-fly calibration and an automated single-pass imaging pipeline [1]. In this model, the software would have one chance to read the visibilities and one chance to write the resulting image, with no opportunity to iterate.

For various reasons [2] this approach was not compatible with commissioning or early operation of the telescope. In addition, the Cray XC30 supercomputer (known as Galaxy) that was tasked with running the imaging pipeline has proven to have insufficient resources to process full-resolution data from ASKAP in the way originally intended.

Funding has been secured to purchase a replacement for Galaxy, but the new machine will not be commissioned until 2020. Installation and commissioning of ASKAP's receivers and digital electronics will be completed in 2018, with the expectation that full-scale pilot surveys will begin in early 2019. It is therefore necessary to use existing batch processing techniques for pilot surveys. This will provide more flexibility than originally envisioned, at the cost of being able to keep up with the full data rate in all modes.

Commissioning activities have focused software development efforts on image quality, meta-data correctness and optimisation of tools that were not expected to be used in the final system. This increases the complexity of the software, but has the advantage that it allows further experimentation and provides experience to inform a revised final design in advance of the Pawsey hardware upgrade.

The pipeline that will be available for pilot surveys should allow tests of algorithms and imaging techniques on full-scale data sets. We will temporarily store visibility data (allowing for a small number of iterative passes to refine processing techniques) and manage the pipeline using human operators, with additional quality control provided by external science teams.

2 Introduction

This document begins by describing key aspects of the telescope, including the unique wide-field capabilities that give it a high survey speed. We then discuss the operations model and data processing work-flow, including key components and interfaces. Next, we outline the pilot survey

[1] The original specification of 300 MHz involves a partial block of correlator hardware and increases system complexity. For pilot surveys we will be operating with a 6-block, 288 MHz correlator

process that will test full-scale operations for the first time. This is followed by a description of several observing modes that will be offered, each with different resource requirements. The final section of the document outlines the scheduling process and how it will be linked to the goals of the pilot survey program.

Note that detailed performance specifications of ASKAP will be provided in a different document to facilitate planning and preparation of pilot survey proposals. Details of specific observatory-led projects (a new addition to the plan) will also be circulated [3].

2.1 Operating a survey telescope

ASKAP is unusual in that it was designed to spend most of its time conducting large-scale surveys. This mode of operation differs from traditional radio astronomy work-flows where individuals or small teams carry out their own processing of a single object or region. Large-scale surveys involve much more data and require careful planning in advance.

Most radio telescopes conduct large scale surveys only after they have been operating for many years and have well-established procedures. It is important to understand the characteristics of the instrument (and its limitations) when conducting surveys as large-area data sets are particularly vulnerable to systematic errors and biases. Since ASKAP will begin survey operations very early in its life, we must accelerate the process of learning key characteristics that will allow optimal survey strategies to be devised.

Due to the above, the operations model for ASKAP is different to other CASS national facilities. ASKAP will be operated in service mode by a team of CASS staff who will schedule and conduct observations, calibrate and image the resulting visibilities and finally upload the results to the science data archive (CASDA) [4] which will be the primary interface to the telescope for all external users. This model centralises expertise and resource allocation, which should allow for optimal processing efficiency.

As methods and procedures mature, we expect that more of the processing will be automated. However, while we are still learning the system, it is crucial to have humans driving the processing software and assessing data quality.

3 Pilot survey work-flow and goals

ASKAP is expected to detect tens of millions of new radio sources, making carefully curated catalogues one of its primary outputs. The pilot survey program is a joint exercise in which CASS and its partner science teams learn how to conduct large-scale surveys using the full capabilities of this new telescope.

A key aspect of the pilot program is availability of the full array – all antennas, beams and as many spectral channels (with appropriate frequency resolution) as required to meet the science goals. This will ensure that pilot surveys reflect the full data rate and capability of the telescope.

In order to ensure timely processing, analysis and feedback, the scope of each pilot survey will be limited in observing time with 100 hours being the benchmark allocation for all science teams.

Each pilot will have a well-defined end point and will culminate with an assessment of readiness for further observations. Additional pilot surveys may be conducted after the conclusion of the first, if the assessment process deems that the strategy is not yet suitable for a full-scale survey.

The assessment of each pilot survey will occur once processing has reached a conclusive outcome. There will be several stages to this process, including completion of observations, completion of operations processing and availability of level 5 data on CASDA, completion of science team analysis and release of level 6 data and the availability of level 7 data products (if required).

In addition to the pilot survey assessment process, we will conduct an overall review of the proposed survey science projects in mid–late 2019 in order to set scheduling priorities for full surveys. This is the review that was postponed in 2018 after discussions with the science team Principal Investigators (PIs).

Pilot survey progress will be displayed on the ASKAP Observation Management Portal (OMP) [5] for transparency. The goals of each pilot survey should be displayed online (most likely via a confluence template) along with a summary of progress through the stages described above.

It is highly likely that the pilot survey assessment process will deem some strategies to be suitable sooner than others. In this case, pilot surveys may co-exist with full operations and telescope time will be scheduled to ensure that pilot surveys retain a high priority.

3.1 Pilot survey plan outline

The baseline plan involves an initial allocation of 100 hours duration for each team. This small allocation is designed to explicitly encourage rapid analysis and keep the emphasis on assessing technical feasibility. We expect that processing times may be significantly longer than observing times, so a small allocation of time should help avoid lengthy processing backlogs.

Telescope time will be shared between pilot surveys and observatory projects, alongside operational overheads, capability improvement testing and maintenance. Additional use cases will be considered if emerging priorities present a compelling science case. Major changes in direction require approval from the ASKAP project director (John Reynolds at the time of writing).

Survey teams will be required to submit short proposals outlining their planned pilot surveys in early 2019: these should explicitly address the key goal of establishing readiness for full survey operations. This includes determining and testing both the observing and processing strategies, through to validation and data release on CASDA.

Each pilot survey proposal should specify a list of field directions, footprints, dwell times, scheduling constraints (e.g. night only), modes of operation and also a processing strategy. Total observing time should not exceed the values indicated above. Requests for intensive or highly customised processing strategies should be accompanied by compelling reasons why the default strategy will not yield suitable image quality.

Proposals must also estimate resource requirements in terms of disk storage - this will be factored in to the scheduling process. Pilot surveys operating in resource-heavy modes (e.g. all frequency channels) will be more difficult to schedule and will be conducted a few scheduling blocks at a time.

Each science team will be allocated an appropriate fraction of the total available disk space for raw visibilities. We will attempt to adaptively schedule observations so as to keep data flowing to all science teams in a timely manner. Disk-intensive pilot surveys will be scheduled until their space allocation is full and must then wait until processing is complete and older data can be replaced.

3.2 Pilot survey proposal process

CASS will distribute a pilot survey proposal template to all science team PIs following the workshop in October 2018 (see timeline below). This will help to ensure that all necessary information is captured. Survey proposals should be developed by the science teams after the workshop and after technical feasibility assessment observations using array release three.

Each team should have a rough idea of their survey concept so that initial technical feasibility discussions can be had at the October 2018 workshop. We expect to receive pilot survey proposals in early 2019. These proposals will not be assessed competitively, rather it is assumed that every pilot survey gets its share of observing time. Instead the proposals will be assessed for technical feasibility and used as documentation for the operations team to develop a detailed observing plan. This will be done in consultation with science team representatives.

4 ASKAP science data processing software development

Due to ASKAP's high data rates, a new imaging software package was designed and constructed specifically with parallel architectures in mind. This package is known as ASKAPsoft. Despite using a custom package, great care has been taken to ensure that ASKAP's raw data output is compatible with CASA. Visibilities are written to a CASA measurement set with comprehensive and correct metadata, so it should be possible to image these data using standard CASA tools. However, due to the fact that few other telescopes use the multi-feed aspect of the measurement set, existing CASA tools may not always treat ASKAP data in the way the user expects. It may also be difficult to process more than a small subset of data on a typical PC due to large file sizes.

ASKAPsoft was designed to be less flexible but more efficient than other software packages. It is based on casacore libraries, but implements its own gridding and imaging algorithms that can be distributed over multiple processors using MPI. In addition, we have developed a scripted pipeline that handles job distribution via the slurm batch queue system used at the Pawsey super-computing centre.

ASKAPsoft will be used for default operations processing in order to make the most efficient use of super-computing resources. However, if science teams require complex processing algorithms that ASKAPsoft does not implement, there are two possible solutions. If small changes to ASKAPsoft code could deliver the requested feature, the ASKAPsoft development team can add the work to its monthly sprints. Priorities will be decided by the ASKAP lead scientist (Aidan Hotan at the time of writing) in discussion with the developers and the science teams.

Alternatively, if the requested features are extensive or poorly suited to efficient parallel execution, science teams should investigate extracting a subset of the raw data for processing in another environment using CASA or other tools. In these cases, it will still be necessary to remove the raw visibility data from disk in a timely manner so that other observations may continue.

ASKAPsoft is available for download in source code form [6] and can potentially be installed at other super-computing facilities to help spread processing load. We are currently exploring options to employ a full-time member of staff to support remote installations of ASKAPsoft and assist with data transfer between institutions.

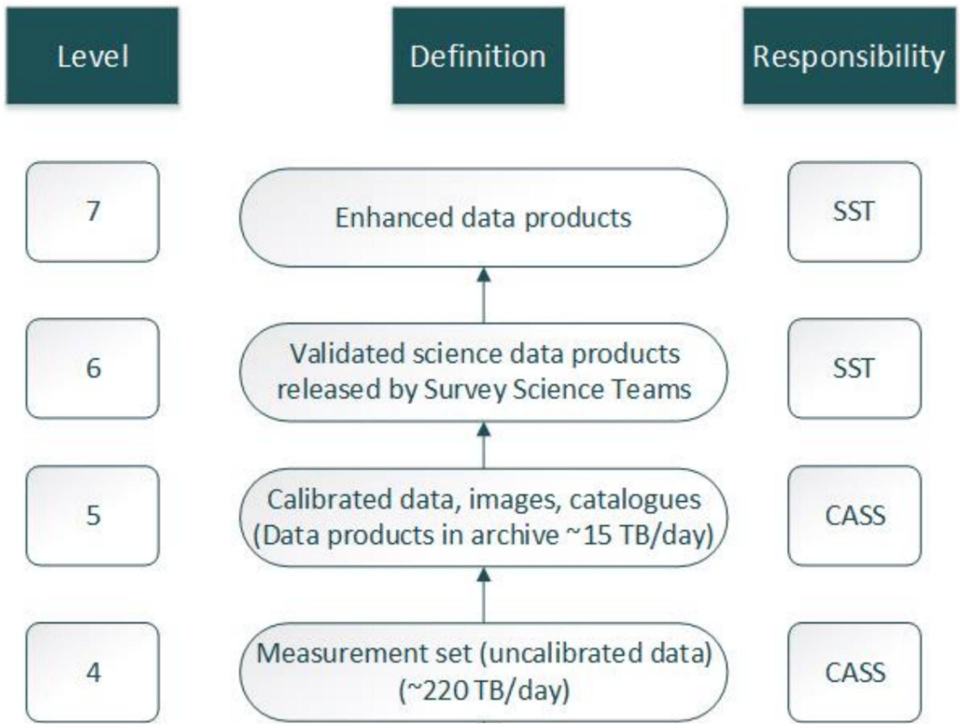

Figure 5.1: Detailed description of CASDA data levels

5 ASKAP operations work-flow

As described above, ASKAP observing will be run by an operations team internal to CASS. This team will be responsible for converting survey descriptions into observable scheduling blocks and managing the life cycle of each block through to deposit on CASDA [7] at level 5. Further details regarding the default imaging parameters and modes of operation can be found below.

For the duration of pilot surveys, images will be made in batch processing mode on Galaxy at Pawsey, from visibilities stored on disk. This provides some amount of flexibility to perform two or more passes in order for the operations team to assess and improve image quality before the raw visibilities are removed. This places some hard limits on the telescope's duty cycle and acknowledges that learning about the system is more important than maximising total throughput.

Survey science teams will be responsible for final quality assessment of each level 5 product using the CASDA web interface, after which the data will be released to the public by CASS. Each science team will then perform further processing, including multi-block combination, mosaicking of large survey fields over several footprints, creation of value-added catalogues and so on. These higher-level data products should be uploaded to CASDA as level 7 products.

The default strategy will be that each individual scheduling block is imaged and uploaded to CASDA. If there are concerns over ASKAP's ability to deliver a science-ready data product from a single scheduling block, they should be raised in the pilot survey proposal. One such example is a lack of sensitivity in individual blocks due to the amount of integration time required. Another is the situation where uniform sensitivity on the sky is achieved using two or more interleaved footprints. If all interleaved positions are observed in the same scheduling block, no additional effort is required. However, if the interleaved positions are observed in different blocks, some additional mosaicking is required and it may even be necessary to jointly de-convolve all beams.

We expect that dedicated astronomy data processing resources at Pawsey (Galaxy) will be fully utilised by routine operations processing and internal software development activities. Creation of level 7 data products may require significant computing resources which the science teams will need to arrange on other machines at Pawsey (e.g. Magnus), or other facilities. In some cases where advanced processing can be made relatively routine, the ASKAP operations team may be able to assist or incorporate the additional steps into the standard work-flow.

The ASKAP computing group will submit a proposal for Magnus time dedicated to exploratory processing and post-processing, and developing ways to construct science-ready data products from level 6 data available through CASDA. This project will be managed and supported by CSIRO as an alternative to Galaxy, but hours will be shared amongst all science teams and will not be sufficient for bulk processing. Limits to the number of concurrent jobs will mean that it is not possible to run full-scale pipeline instances on Magnus.

Science teams may wish to submit their own proposals for Magnus time, though it would be good to differentiate the associated goals from the CSIRO-led proposal which is managed by Matthew Whiting at the time of writing.

5.1 Scheduling and observing

Observations will be scheduled in an adaptive manner by the operations team. Priority will be set by the desire to ensure that all science teams have data to work with at any given time. This involves communication with the science teams and we expect that the weekly ASKAP Commissioning and Early Science (ACES) meetings will evolve to include an operations planning component when pilot surveys commence. The operations team will manage bandpass calibration overheads, beam formation and updates, scheduled maintenance and so on (time for these activities does not count towards survey allocations).

When proposing pilot surveys, science teams should specify any additional scheduling considerations required. This could include avoiding Solar interference by observing only at night, or specifying a minimum number of available antennas (and spatial resolution given UV coverage). For safety and data quality reasons, ASKAP will not automatically observe any position within 10 degrees of the Sun, which will impact the completeness of shallow all-sky surveys at any given time. Another consideration is shadowing of nearby antennas, which can occur on the shortest baselines between antennas 1, 2 and 3 in the core of the array. Shadowed antennas will be flagged, but if the resulting loss of short baselines will impact the science project, the operations team can schedule observations to occur only when the target is at higher elevations. This will be done automatically for beam-forming and bandpass calibration observations.

5.2 Calibration

The default calibration scheme will be traditional bandpass calibration, done by placing each beam on a calibrator source for a few minutes once per day. Bandpass tables will be maintained by the operations team and applied to science observations. Note that this incurs an operational overhead of roughly 3 hours in every 24 hour period. This overhead may be reduced if it can be shown that the on-dish calibration system is able to track gain changes and extend the validity of any given bandpass measurement.

Phase calibration can be done in a number of ways. The default mode will be to perform phase-only self-calibration on the brightest sources in any given beam, based on source-finding on the output of the first major cycle. Amplitude self-calibration can also be done at this stage but it is likely to invalidate flux calibration and is not recommended.

Beams will routinely be calibrated to have zero X-Y phase using the on-dish calibration system. Any additional polarisation calibration requirements should be clearly specified in pilot survey proposals. Ionospheric rotation measure calibration will be supported before frequency averaging providing the required correction factors are available within a few days of observing. Note that if such corrections are necessary, it will not be possible to use online channel averaging, which significantly increases the disk usage of each observation.

For certain well-studied fields and after creation of an initial sky model, it will also be possible to perform phase calibration using model components corresponding to selected bright sources in the target field. This is likely to be less resource intensive than self-calibration based on CLEAN components or source-finding and we expect to rely more on this approach as sky models improve. Eventually, this step could be done online as per the original pipeline processing model presented in [1].

5.3 Imaging

By default, imaging will be done on each beam individually. Visibility data will be gridded using W-projection (with optional snapshots to decrease memory requirements) and de-convolution will be done using multi-scale, multi-frequency CLEAN with thresholds chosen appropriately for the science goals. Parameters (such as CLEAN scale sizes and visibility weighting) will be set in consultation with the science teams, likely after some initial exploratory processing on representative single-beam data. The full set of available parameters is documented online [8].

An image of the full field will be created by linearly mosaicking all beams in the image domain. Alternatives include joint de-convolution of all beams, though this will only be possible in special circumstances (for fields containing broad, diffuse emission) for a subset of frequency channels.

Continuum imaging will be done using a Taylor term algorithm with two terms from which spectral indices will be derived. Images will be produced for all Stokes parameters for the full band by default. Science teams can request continuum cubes (with 1 MHz channel resolution) without significantly increasing resource requirements. If preferred, spectral indices can be derived from these continuum cubes instead of the Taylor term images.

Spectral line imaging will be considered a secondary stage on top of continuum imaging, since high-quality images are required for continuum subtraction. By default, only Stokes I images will be made at full spectral resolution. Identification of a limited velocity range to process will reduce resource requirements and is highly recommended. We acknowledge that this is not possible for non-targeted surveys over the full frequency range and such projects will likely be the most technically challenging to conduct.

As previously mentioned, one image will be made from each scheduling block. For projects such as transient searching, it may be desirable to make several images (e.g. one every few minutes). This requires additional processing load and complexity, but should be possible. Initial assessment of advanced mode feasibility will be a topic for discussion at the October 2018 workshop.

Requests for non-standard imaging strategies should be included in pilot survey proposals and will be assessed for operational feasibility. Where possible, tests of advanced modes will be conducted between November 2018–March 2019 in advance of pilot survey commencement.

5.4 Commensal observing

For simplicity, we assume that pilot surveys will be conducted independently for each science team in order to test specific modes of operation. However, another aspect that could be tested is whether two or more science teams can make use of the same observations.

While we continue to store visibility data for offline processing, it is straight-forward to produce both continuum and spectral line images from the same data with different processing strategies. For example, it would be possible to make continuum-averaged images at the full spatial resolution provided by the longest baselines, and then impose a UV distance limit on a second round of spectral line imaging to reduce memory requirements and better suit the goal of detecting diffuse emission.

We encourage all science teams to discuss the feasibility of using common fields. Should a commensal strategy prove viable, two or more teams may combine their allocation of time.

5.5 Quality control

The operations team will take responsibility for providing images that are, within reason, free from obvious artifacts. We do not guarantee that images will reach thermal noise limits or any specific dynamic range target. Quality assessment at this level will be an exercise for the science teams, who will be responsible for validating level 5 data using the CASDA web interface. The observatory will then release validated data to the public as level 6 data. It is expected that this quality assessment will happen in a timely manner (within a few days of the data being made available at level 5). If no action is taken within two weeks, CASS staff responsible for the CASDA platform will perform the quality assessment task in conjunction with the operations team.

Image products uploaded to CASDA by the operations team will represent a reasonable, best-effort attempt to apply correct calibration and appropriate imaging parameters. If the resulting image quality does not meet science team requirements, it can be rejected during the CASDA approval process. However, unless the problem is isolated to a small number of images, the fields will not automatically be re-observed without consulting the relevant science team and deciding how the observing strategy should be modified so that future images do meet the requirements.

In most cases, it is expected that images will be released at level 6 if they are free from artifacts or systematic errors, even if they do not meet all science requirements. This is to ensure wide availability of data that could be useful for science goals other than those originally intended. Data released at level 6 through CASDA can have notes added at the time of validation to describe issues with data quality in more detail.

Quality control will be a key part of assessing the viability of a survey strategy. If an initial attempt does not meet requirements, the pilot process will continue until a suitable strategy has been developed, or it is decided that ASKAP cannot meet the requirements specified.

When additional processing resources become available during full operations (as a result of upgrades or for other reasons), survey strategies may need to be adjusted. Additional pilot surveys may be conducted during full operations to test any such changes.

5.6 Primary data products uploaded to CASDA

Each scheduling block processed by the operations team will be uploaded to CASDA if it meets the initial quality control criteria described above. There will be a standard set of data products in the upload package, consisting of the following for continuum mode:

- Continuum-averaged, calibrated visibilities in measurement set format
- Full-field, restored continuum images (all beams mosaicked) for all Stokes parameters
- A continuum source catalogue produced by the Selavy source-finding software
- Output of the continuum validation pipeline to assist with quality control
- Packaged parameter set and log files for a full record of all processing done

Data will be available through a web interface [4] and catalogues will also be available via the Virtual Observatory table access protocol [7] for ease of access. If requested by the science teams, additional data products (within reasonable storage limitations) can be added to the list above. These may include a continuum image cube (retaining 1 MHz channels), individual beam images prior to mosaicking, CLEAN component images, noise maps, etc.

In addition, spectral line data products will be uploaded when appropriate. These include:

- Full-field, continuum-subtracted, restored Stokes I image cube
- A spectral line source catalogue produced by the Selavy source-finding software
- Output of the spectral line validation pipeline to assist with quality control
- Packaged parameter set and log files for a full record of all processing done

Validation pipelines will be constructed with the assistance of the science teams. These will typically consist of `python-casacore` routines that exist within or will be added to the module environment on Galaxy. Metrics should be as simple and robust as possible. Statistical information obtained as part of the imaging process (e.g. RMS in each channel) will be made available by the imaging software in a format to be documented online [8]. Validation metrics will be summarised in XML format for upload to CASDA so that they may be used as search criteria.

Pilot survey proposals should clearly list the data products that need to be uploaded to CASDA if they are not part of the default package described above.

5.7 Value-added data products uploaded to CASDA

Experience with early science observations suggests that most of the science from ASKAP will be derived from so-called value-added data products. These are typically formed by combining individual scheduling blocks in various ways, including the creation of new catalogues and cross-matching with other surveys.

CASDA supports upload of these high-level data products as level 7 items and we encourage all science teams to make use of this facility. To support internal sharing of value-added data, items uploaded at level 7 may have a proprietary period of up to 12 months, during which time only members of the originating science team will be granted access. This allows time for publications to be prepared while still providing the Digital Object Identifiers (DOIs) required to reference these objects. It also enables collaboration with other observatories that may have policy requirements regarding non-public access to pre-release data.

6 Modes of operation

As previously stated, we do not have the resources required to process full-resolution ASKAP data in the way originally intended. Eventually, we expect that computing capability will evolve to make current data rates much more tractable. Future processing pipelines and algorithms will benefit from the experience gained during the pilot survey process.

We will work with the ASKAP science teams to define a processing strategy for each pilot survey project that produces acceptable results. In the beginning, this may not be the optimal strategy. In future years it will be possible to improve data quality and we expect to re-survey the sky on a regular basis, both to search for transient sources and improve existing data products.

In order to make best use of resources, we have established several specific modes of operation that are linked to different science goals. These will be described in the following sections.

Note that the original plan called for 5 second correlator cycle times. So far, we have been unable to reliably achieve this cadence and have most recently been recording with 10 second cycle times. There is some concern that the fringe tracking parameters need to be updated more frequently than this to avoid de-correlation on the longest baselines, but decreasing the cycle time also increases the data rate in direct proportion. This will be investigated further before pilot surveys commence.

6.1 Online frequency averaging mode

The default mode of operation will be to average in frequency space to 1 MHz. Basic online flagging will be done at full spectral resolution, but subsequent flagging can only reject whole 1 MHz channels. However, this mode uses 54 times less disk space than recording at the full rate and can also be processed relatively easily on Galaxy. In this mode we expect that processing could keep up with observing.

However, it should be noted that image processing time does not scale linearly with observing time. Imaging 10 minutes of data cannot currently be done more than a factor of a few times faster than imaging 10 hours. Shallow surveys will therefore suffer from disproportionate data processing overheads. Early tests show that as integration time decreases, most of the processing time is spent in the minor cycle CLEAN, which is currently a serial task. This could be made more efficient, but the required development would have to be balanced against other priorities.

6.2 Full data rate mode

In this mode, all possible data products will be recorded, including 15,552 spectral channels. With 10 second cycles, this is roughly 1.25 GB per second, or 54 TB in 12 hours.

One way to mitigate the large data rate is to specify a subset of channels likely to contain the spectral line emission of interest. After creation of the initial continuum data products, the operations team can extract the desired frequency channels and remove the original measurement set to conserve space. It may be possible to do this online, but further development is required.

6.3 Partial array mode

This mode can be combined with any of the others to reduce the number of baselines processed. One application would be excluding the outer 6 antennas from the array to reduce spatial resolution when studying diffuse spectral line emission. Note that this mode is currently implemented through meta-data rather than reconfiguration (excluded antennas are present in the measurement set, but flagged) and does not save significantly on disk space. It does however reduce memory requirements at the gridding and imaging stages. This means it can also be implemented at the imaging stage in case commensal projects would like to use different sets of antennas from the same observation.

Note that routine maintenance and other activities will cause the total number of available antennas to be less than the full 36 some of the time. It will not be uncommon to operate with one or two antennas flagged and this should be factored in to sensitivity calculations if reaching a specific noise limit is of utmost importance.

6.4 Enhanced frequency resolution modes

The default ASKAP channel width is 18.52 kHz. If required, this can be decreased at the expense of overall bandwidth. During pilot surveys, all correlator modes will produce the same number of frequency channels (15,552) covering at most 288 MHz. It will be possible to operate in the following zoom modes with the same number of total channels:

- 9.26 kHz, 144 MHz bandwidth
- 4.63 kHz, 72 MHz bandwidth
- 2.31 kHz, 36 MHz bandwidth
- 1.16 kHz, 18 MHz bandwidth
- 0.58 kHz, 9 MHz bandwidth

These zoom modes will be a single continuous band which can be placed anywhere within ASKAP's frequency coverage of 700–1800 MHz. Multiple zooms or split-band modes will not be available for pilot surveys.

6.5 Fast transient searching

This mode of operation is unique in that it is targeted at one science goal specifically. However, the methods developed may be of use for other science cases and creative thinking is encouraged.

Searching for fast transients requires downloading power spectra at a rate that is not possible from the normal imaging data path through the correlator. It is however possible to download data more rapidly from the beamformers, both in detected and raw voltage form, with the former being continuously streamed and the later being read from a memory buffer after an event trigger. This allows offline correlation of the voltage data.

In both cases, data are streamed to dedicated servers at the observatory that are not part of the normal imaging system. This means that Pawsey data ingest and imaging can occur in parallel with fast transient searching. This is desirable to assist with delay calibration for offline correlation of voltage data, and can be combined with online frequency averaging to keep data rates low.

Operating the fast filter-bank download in fly's eye mode is an alternative that provides increased effective field of view at the expense of sensitivity and does not use visibility storage resources at all. However, events detected in this mode cannot be precisely located.

7 Scheduling and priority

High spectral resolution is required for many ASKAP science projects, but we do not expect to operate at full duty cycle when recording the maximum data rate. Therefore, we will use online averaging mode to reduce resource usage between full-rate observations. By scheduling in an adaptive manner, we hope to make the most efficient use of available telescope time. In general, telescope time is likely to be plentiful when compared to the CPU time required to process observations.

As experience is gained, we will adjust the ratio of averaged to full resolution observations in an attempt to utilise roughly 70% of the available telescope time. This may mean we spend several days observing in averaged mode or fly's eye mode for every day spent observing at full resolution. Within this practical limitation, observations will be scheduled in such a way as to ensure that all science teams have data to work on. This will involve jumping between different observing programs on a regular basis. If there are observations that must be done in close proximity, this should be made clear in pilot survey proposals.

As far as possible, data will be processed in the order it was obtained, to keep results flowing through the system. This may present difficulties where complicated processing is required, especially if there is simpler processing that could share system resources. The operations team will endeavour to schedule processing jobs as efficiently as possible.

In some cases it may be necessary for a small number of science teams to directly process visibility data on Galaxy, particularly when new observing strategies are being tested. Provision will be made to support this, but only on limited data sets (e.g. single beams) and with lower batch queue priority so as not to interfere with operations processing.

7.1 Observatory-led projects

In addition to the time allocated for survey science operations, ASKAP will spend some of the early science and pilot survey period conducting observatory-led projects. The goals of these will be complementary to (not competing with) the goals of the survey science teams and aim to provide data that can be used to improve the telescope or that will be of broad interest.

The first such project will be a shallow all-sky survey designed to provide a sky model that can be used for calibration purposes. This sky model must cover ASKAP's full frequency range and will likely be extremely useful for other science goals. All data from these observatory projects will be made available on CASDA for the entire community.

Observatory-led projects that are crucial to the development of new features (such as the sky model) will be scheduled in advance of pilot surveys. The time used will be in addition to the 100 hours initially allocated to each science team for individual pilot surveys.

8 Early science and pre-pilot observations

During 2016–2018, ASKAP conducted an early science program, with an array of nominally 12 antennas. Several of the science teams were able to obtain publishable results from these observations, but we acknowledge that the original scope of the program could not be fulfilled in all cases.

Now that the size of the array has expanded significantly, this first phase of early science is drawing to a close. To clearly mark the end-point and make way for the pilot survey process to begin, we will be deleting all raw visibility data from early science phase one from the online storage disk `/astro` on November 6th 2018.

Data from early science phase one will remain in the commissioning tape archive, but will only be restored upon special request for compelling reasons. It is assumed that all processing of these data will be complete by November 6th, or that all data still needed for processing will have been copied elsewhere. The exception to this is operations-led processing of spectral line data from NGC7232, Dorado, Fornax and M83, for upload to CASDA. These observations will be processed from tape one block at a time to minimise the amount of space used.

8.1 Early science phase two

In the lead-up to pilot surveys, we expect to make an unspecified amount of observing time available for preliminary tests of survey modes with 28 or more antennas. Time will be allocated on a shared-risk best-effort basis after November 6th 2018 until the start of pilot surveys. Requests for time should be kept small (especially for resource-intensive modes) and targeted at understanding the system rather than achieving publishable results.

During this period, the telescope will primarily be used for commissioning work and early science observations must fit in around other activities. The main priority will be delivering the full ASKAP system in time for pilot surveys to commence in March 2019 and we expect that only a small fraction of time will be available for early science observations.

Disk space for early science will be limited to existing resources (roughly 1 PB of visibility data). We have requested that Pawsey increase the available disk space to better support pilot surveys and while this has been positively received, the detailed outcome of this proposal is not known at the time of writing.

9 Pilot survey planning considerations

The phased array feed technology that ASKAP uses to achieve a wide field of view is both a source of complexity and an advantage. This new technology gives ASKAP flexibility that other radio telescopes do not possess and with careful use may increase the dynamic range that can be reached.

9.1 Phased array feeds: complexity vs control

ASKAP's primary beams are determined by electronic weights that can be varied on short timescales. This gives us the opportunity to alter the antenna illumination patterns to achieve desirable beam characteristics, then keep the system stable even when underlying factors (such as individual element gains) change. However, since ASKAP is one of the first telescopes to employ such technology, we must learn how to best achieve these goals and develop associated operational procedures.

To begin with, we will treat PAF beams much like traditional telescope feeds and keep fixed beam weights for as long as possible in order to minimise the number of free parameters in the system. Over the life of ASKAP, we expect to greatly improve on this initial starting point.

Areas for improvement include (in order of increasing complexity) monitoring of PAF element gains to adjust beam weights in response to changes in the system, improved models of aperture illumination for accurate mosaicking of individual beams, development of new beam formation algorithms that optimise shape rather than sensitivity, adaptive interference mitigation and eventually joint optimisation of beam formation and imaging using Eigen-beams.

9.2 Beam footprints, shapes and correlated noise

Because ASKAP's 36 primary beams can be steered anywhere within the field of view, surveys can potentially achieve Nyquist spatial sampling of a field in a single pointing. However, this requires closely spaced beams that inevitably share some of their receiver elements, which in turn leads to correlated noise that raises the system temperature [9] [10].

An alternative is to use a more spread-out beam footprint and observe closely spaced interleaved positions at different times. This mitigates the effect of correlated noise, but requires integration time to be spent on one or more interleaved positions. It is not clear which of the above strategies provides the optimal survey speed and in fact the best strategy may depend on the science goals and the scale of the survey.

Primary beam correction is currently done assuming a circular Gaussian beam profile. Holography observations show that this approximation is not strictly valid and is increasingly inaccurate towards the edge of the field. We are exploring ways to better model the antenna illumination pattern and incorporate holography measurements into the linear mosaicking software. Alternative beam-forming methods that optimise shape rather than sensitivity are also under development, but these improvements may not be available by the start of pilot surveys. Using the circular Gaussian approximation could lead to flux errors of up to 10% in some parts of the field.

9.3 Footprint tessellation

When planning a wide-area survey, it is also important to consider how the beam footprint will tesselate to cover larger areas. Due to the placement of elements in the focal plane, the natural field of view of an ASKAP antenna is rectangular, not circular. However, due to the circular nature of individual beams it is not possible to create a uniform field with straight edges. Surveys for which a uniform noise map is key to the science goals must consider these issues carefully.

The curved nature of the sky means that tesselation of rectangular footprints requires rotation of the footprint about the optical axis for different field centres. Fortunately, this can be accomplished mechanically using ASKAP's third axis of rotation (known as the roll axis). By default, this will be configured to track the position angle of the field centre. Travel is limited to roughly 170 degrees either side of centre (when pointing corrections are taken into account), so care should be taken not to cross limits when large offsets are required.

At present, it is not possible to cross roll axis limit boundaries as there is no mechanical overlap. An automatic unwinding procedure will pause a long track for approximately 15 minutes while the source transits this boundary, but requesting a new source outside of these limits will cause the block to fail. This should be considered when tracking an object North of –26.7 degrees Declination through transit, whether using long tracks or many short observations.

9.4 Radio frequency interference

The Murchison Radio-astronomy Observatory (MRO) at which ASKAP is located has a favourable RFI environment compared to most other observatories, but there are still some significant sources of RFI to consider when planning the frequency coverage of a survey. Note that it is not possible to split the observing band in order to avoid regions of RFI (though this may be implemented as a future upgrade).

Interference from satellites and aircraft navigation is difficult to avoid anywhere on Earth. In particular, global navigation system bands around 1220–1250 MHz and 1600 MHz cause significant interference, as do aircraft ADS-B transponders at 1090 MHz.

The 700–1000 MHz band is largely free from interference, except for times when atmospheric conditions cause ducting of mobile phone signals from over the horizon. During these ducting events we see significant interference at 3G bands around 850 and 900 MHz. These events can be predicted from meteorological data and ASKAP operators will attempt to schedule observations at other bands during these times.

9.5 Software tools for survey planning

Addressing the considerations above requires detailed knowledge of supported ASKAP beam configurations. Although beams can be placed arbitrarily within the field of view, we have developed several standard footprints based on sensible packing arrangements. These are parameterised by a name (which gives an indication of the packing method) and pitch angle, which determines how far apart the beams are placed within the basic structure. From these two parameters, it is possible to determine the offset angle and distance of each beam from the field centre.

Software to perform the above calculation is available as a python package [11] on bitbucket. This package includes a number of other useful tools that can compute tiling positions for wide areas, plot baselines between antennas and so on.

There is also an online sensitivity calculator [12], though users should be wary of assuming that thermal noise limits will be reached, especially during the pilot survey phase. The System Equivalent Flux Density (SEFD) of the array can vary depending on the specifics of beam formation, but measurements yield a value of around 1800 Jy on average. The noise temperature of the PAFs is roughly 50 K, but the efficiency of the antennas is about 0.7, giving $\frac{T_{\mathrm{sys}}}{\eta} \approx 70$ K. Please see [13] for updated information and additional details.

10 Future developments and roadmap

Scaling up to full operations with ASKAP will be done through a series of incremental pilot surveys with opportunities for assessment at key stages. The first such point will be the conclusion of processing for the first 100 hours allocated to each team. Pending successful assessment of pilot survey results, full-scale survey programs may begin (subject to the outcome of the time allocation review), or it may be necessary to conduct further pilot surveys if significant changes are required.

Provision of new super-computing hardware at Pawsey towards the end of 2020 is expected to enhance the capabilities of the telescope, particularly for data rate and processing intensive modes of operation. There will be a 6 month overlap period during which established operational procedures can continue while the new super-computer is tested in parallel. We will conduct a review of capabilities and provide opportunities to change operational modes once performance data for the new system become available.

10.1 Timeline

- September 2018: Publication of this plan and observatory projects memo
- October 2018: Publication of ASKAP performance summary
- October 2018: Technical workshop to study pilot survey feasibility
- November 2018: Early science phase two commences
- January 2019: Pilot survey proposals received
- February 2019: Sky model survey commences
- March 2019: Pilot surveys commence
- Mid–late 2019: Pilot survey assessment process
- Mid–late 2019: Full survey time allocation review
- Late 2019: Next round of pilot surveys or full operations commence
- Late 2020: Pawsey upgrade completed, capability review

11 Summary and conclusions

This document outlines the plan to progress from ASKAP's commissioning and early science phase through to full operations, by way of pilot surveys. It is expected that ASKAP's capabilities will continue to improve as we develop new ways to make best use of phased array feed technology. Pilot surveys will therefore not necessarily represent the most optimal output from the telescope, but they will be a significant advance over existing capabilities and should comprehensively demonstrate the viability of full-scale surveys with ASKAP.

References

[1] ASKAP Science Processing. *Cornwell, T., Humphreys, B. and Lenc, E. et al.* http://www.atnf.csiro.au/projects/askap/ASKAP-SW-0020.pdf

[2] Barriers to Real-time System Deployment in Radio Telescopes. *Hurley-Walker, N. and Hotan, A.* In preparation

[3] ASKAP observatory projects memo. *McConnell, D. et al.*. https://confluence.csiro.au/display/askapsst/Pilot+Surveys

[4] CSIRO ASKAP Science Data Archive (CASDA). https://data.csiro.au/dap/public/casda/casdaSearch.zul

[5] ASKAP Observation Management Portal (OMP). https://apps.atnf.csiro.au/OMP/

[6] ASKAP Science Data Processing (SDP) software repository (Bitbucket). https://bitbucket.csiro.au/projects/CASSSOFT/repos/askapsoft/browse

[7] CASDA User's Guide. https://www.atnf.csiro.au/observers/data/casdaguide.html

[8] Science Data Processing (SDP) software documentation (ASKAPsoft). http://www.atnf.csiro.au/computing/software/askapsoft/sdp/docs/current/

[9] Quantifying and mitigating correlated noise between formed beams on the ASKAP Phased Array Feeds. *Heywood, I.* http://www.atnf.csiro.au/projects/askap/interleaving_memo.pdf

[10] Estimating ASKAP beam to beam correlation. *McConnell, D.* http://www.atnf.csiro.au/projects/askap/memo014.pdf

[11] ASKAP survey planning software (aces-packages) Bitbucket repository. https://bitbucket.csiro.au/projects/ACES/repos/aces-packages/

[12] ASKAP online sensitivity calculator. http://www.atnf.csiro.au/people/Keith.Bannister/senscalc/

[13] ASKAP measured performance parameters. *McConnell, D. et al.* In preparation

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.csiro.au

WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today for our customers, all Australians and the world.
Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.
With more than 5,000 experts and a burning desire to get things done, we are Australias catalyst for innovation.
WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION

CSIRO Astronomy and Space Science
Aidan Hotan
t +61 8 6436 8543
e Aidan.Hotan@csiro.au
w Astronomy and Space Science

CSIRO Astronomy and Space Science
David McConnell
t +61 2 9372 4132
e David.McConnell@csiro.au
w Astronomy and Space Science