Atacama Large Millimeter Array	ALMA-SW-666 Revision:2.2 2001-07-12 *Requirements* Steven T. Myers

ALMA Software Pipeline and Offline Requirements

Requirements

ALMA Pipeline and Offline Software Group:
B. Clark, F. Gueth, M. Momose, S. Myers,
P. Schilke, K. Tatematsu
ALMA Software Science Requirements Committee:
H. Ezawa, T. Handa, R. Lucas, J. Mangum, T. Ohnishi,
J. Richer, S. Scott, F. Viallefond, M. Wright
ALMA Software Use Case Group:
G. Harris, D. Muders, J. Schwarz, R. Warmels

Keywords: Requirements, Pipeline, Offline, Science, Calibration

Author Signature: Steven T. Myers	Date: 2001-07-12
Approved by: B. Glendenning, G. Raffi Institute: NRAO, ESO	Signature: Date:
Released by: Institute:	Signature: Date:

Change Record

Revision	Date	Author	Section/ Page affected	Remarks
1	2001-03-06	S.T. Myers	All	Initial Version
1.2	2001-05-01	S.T. Myers	All	Version for SSR Telecon
2	2001-06-21	S.T. Myers	All	Version to SSR
2.2	2001-07-12	S.T. Myers	All	Comments from SSR

`$Id: myers-report-2.tex,v2.2 2001/07/12 myers Exp myers $`

Contents

1 Introduction **6**
1.1 Nomenclature . . . 6
1.2 General Considerations . . . 7

2 Pipeline Data Processing Requirements **9**
2.1 General . . . 9
2.2 Calibration Pipeline . . . 9
2.2.1 Interferometric Calibration . . . 10
2.2.2 Single-Dish Calibration . . . 10
2.3 Quick Look Pipeline . . . 11
2.4 Science Pipeline . . . 12
2.4.1 Interferometric Data . . . 12
2.4.2 Single Dish Data . . . 13
2.5 Interface with the Archive . . . 13
2.5.1 Data Monitor and Control Quantities . . . 14

3 Offline Data Processing Requirements **15**
3.1 General Requirements and Interaction with Other ALMA Elements . . . 15
3.1.1 Goals of the Offline Package . . . 15
3.1.2 Relation to the Pipeline . . . 15
3.1.3 Operational Issues . . . 15
3.2 Interface . . . 15
3.2.1 General User Interface Requirements . . . 15
3.2.2 Graphical User Interface (GUI) . . . 16
3.2.3 Command Line Interface (CLI) . . . 16
3.2.4 Interface Programming, Parameter Passing and Feedback . . . 17
3.2.5 Documentation and Help Facility . . . 17
3.3 Data Handling . . . 18
3.3.1 General Data Requirements . . . 18
3.3.2 Data Import and Export . . . 19
3.3.3 I/O Speed and Efficiency . . . 19
3.3.4 ALMA Interferometer Data . . . 19
3.3.5 ALMA Single Dish and Interferometer Phased-array Data . . . 19
3.3.6 Images and Other Data Products . . . 19
3.3.7 Foreign Data . . . 20

3.3.8 Interaction with the Archive . . . 20
3.4 Calibration and Editing . . . 20
3.4.1 General Calibration and Editing Requirements . . . 20
3.4.2 Interferometer Data . . . 21
3.4.3 Single Dish Data . . . 21
3.4.4 Mosaicing Considerations . . . 22
3.4.5 Ancillary and Diagnostic Data . . . 22
3.5 Imaging . . . 22
3.5.1 General Imaging Requirements . . . 22
3.5.2 Interferometer Imaging . . . 23
3.5.3 Single Dish Imaging . . . 23
3.5.4 Mosaicing Considerations . . . 23
3.5.5 Inclusion of the ACA . . . 23
3.6 Analysis . . . 23
3.6.1 General Analysis Requirements . . . 23
3.6.2 Visibility Ddata Analysis . . . 24
3.6.3 Single Dish Data Analysis . . . 24
3.6.4 Mosaicing and Combined Array Analysis . . . 24
3.6.5 Image Analysis and Manipulation . . . 24
3.7 Visualization . . . 24
3.7.1 General Visualization and Plotting Requirements . . . 24
3.7.2 Display Appearance and Interactivity . . . 25
3.7.3 Image-cube Manipulation . . . 25
3.8 Simulation . . . 25
3.8.1 General Simulation Requirements . . . 25
3.8.2 Interferometer Simulation . . . 26
3.8.3 Single Dish Simulation . . . 26
3.8.4 Incorporation of Previous or Foreign Data . . . 26
3.8.5 Interaction with Observing Tool . . . 27
3.9 Special Features . . . 27
3.9.1 VLBI . . . 27
3.9.2 Solar Observing . . . 27
3.9.3 Pulasr Observing . . . 27
3.9.4 Other Modes . . . 27

4 Common Algorithms **28**

A The ACA **28**

1 Introduction

The history of this document and correspondence can be found at:

`http://www.aoc.nrao.edu/ smyers/alma/offline-req/`

This document is `olr-2001-07-12-myers.txt`.

This document describes the requirements for data reduction software packages in order to be able to handle the ALMA data output. There are two main sets of requirements: requirements on the ALMA internal data pipelines, and requirements on offline and external data reduction packages. In particular, it is assumed that there will be (at least) one software package used internally by ALMA in order to process the data in quasi-real time (the quick-look and calibration pipelines) and also to fill the archive (the science pipeline). This Pipeline must fulfill the Pipeline Requirements (Section 2). Note that the tools that make up the Pipeline need not be in the same program, and may be from different software packages. However, this suite will be referred to as an integrated entity in this document.

In addition, there must be software available for users to reduce their own data and/or data from the archive offline at their home institution, at an ALMA regional center, or remotely using ALMA center computing. This suite of tools must fulfill the Offline Data Reduction Requirements (Section 3). We refer to the ensemble that fulfills these as the "Package". Again this might be an assortment of different programs from different software packages (e.g. AIPS, GILDAS, MIRIAD, aips++) but in this case it is highly desirable that there be at least one single package that fulfills the Offline requirements or that there be an installation that integrated the necessary parts. Note, however, that is is highly unlikely that disparate applications from different packages will fulfill the requirements on similar "look and feel" and inter-connectivity, and thus it is likely that this will be a single homogeneous suite by default.

I think it is important that the Pipeline be available for installation and use by users, and not be merely an internal ALMA "black-ops" secret weapon. It is desirable, though not necessary, that the Pipeline package be part of the offline Package.

This document assumes requirements already delineated in ALMA-SW MEMO 11 "ALMA Software Science Requirements and Use Cases" at:

`http://www.alma.nrao.edu/development/computing/docs/joint/0011/ssranduc.pdf`

Much of the content of this document is based on the AIPS++ User specifications Memo 115 found at:

`http://aips2.nrao.edu/stable/docs/specs/specs.html`

Although it was intended to be the basis for AIPS++, these are an excellent starting point for our requirements to build upon. I do not think this makes this too specific to AIPS++, as our document should be package-independent.

1.1 Nomenclature

The subject of the Section 2 on Pipeline Requirements is referred to as the *Pipeline*. This may be implemented as disparate tools or programs, or as separate packages provided by different groups, or as a single package, as long as it fulfills the requirements. At its core, the Pipeline is a set of operations, implemented by an underlying software package, which takes a concise description of the way these operations are to be performed and accesses ALMA data, either from the ALMA archive or from local files, and produces a desired data product. There are several pipelines within the Pipeline essential to the efficient operation of ALMA.

The subject of Section 3 on Offline Data Reduction Requirements is referred to as the *Package* or *Offline Package*. This is intended as a set of tools or programs, believed adequate for ALMA reductions, and

used end-users for science and by ALMA staff for reductions upon which the behavior of the system will be judged. It may consist of packages provided by different groups, with transitions provided to integrate them into a single suite. It may be that more than one Package fulfills the requirements of Section 3 and thus can be considered as suitable for ALMA processing. The requirements will state that the Package will be available for installation on the observer's own computer systems as well as present at ALMA centers. Note that the pipeline may or may not be based on the offline Package(s), depending on implementation. However, the functionality of the Pipeline must be available in the offline Package for users.

The user of the package may be referred to as *user* or *observer*, and may be the actual proposer or a staff member. In pipeline mode, the user may actually be another tool or program.

The *archive* refers to the totality of the ALMA data storage and consists of possible different physical archives.

The definitions of various terms, such as *scans*, are given in ALMA-SW MEMO 11.

1.2 General Considerations

A. Two fundamentally new aspects of ALMA are the integrated archive and the pipeline, therefore the impact of requirements on these two areas should be considered. In particular the Pipeline will be the most critical aspect of ALMA given that we envision both an effective dynamically scheduled observatory with prompt user feedback mechanism and a scientifically viable archive. The first substantial section deals with the Pipeline in detail. We have also included topics "Relation to the Pipeline" in the first section of General Requirements for the Offline Reduction, and "Interaction with Archive" to the section on Data Handling.

B. There is a fundamental difference between running the ALMA array in interferometric and single-dish modes. This difference may not be so fundamental for many types of data processing, however. For example, if several single dishes do single-dish together, much of the calibration will be done interferometrically (pointing, focus, beam shape); for interferometric observations, temperature scale is derived from single-dish measurements. Therefore where appropriate we split these two paths, with inclusion of integration of the two where appropriate.

C. There is no fundamental difference between spectral-line and continuum observations, merely number of channels and bandwidths coming out of the correlator. Due to the nature of the ALMA correlator, most of the special calibration needs of traditional spectral line observations (e.g. bandpass calibration) are also applicable to continuum observations (continuum is built up through summation of spectral channels taken at low resolution). We will assume that all data will be effectively taken in spectral line mode therefore.

D. There is no fundamental distinction needed for polarization data, merely consideration of the number of polarization products needed for processing. There may be special modes where polarization (e.g. RR or LL) is synthesized in hardware, but these can be treated as special cases. The model we consider here is the processing of one or more Stokes parameters and thus polarization is integrated into all the topics.

E. Mosaicing will likely be widespread in ALMA data reduction due to the small fields of view. In this trial outline, we choose to include mosaicing as cases (like spectral line) under the interferometric and single-dish modes when the requirements are straightforward, but also under special headings "Mosaicing considerations" where it seems more appropriate (under the Calibration and Imaging headings).

F. We propose a system of prioritizing with codes:

1 = essential
2 = highly desirable,
3 = desirable, but not critical

These codes are enclosed in brackets [] at the end of the items they qualify.

NOTE: No priorities have been assigned in this draft of the document.

G. Specific requirements can/should be broken into sub-cases or instances (e.g. R1 R1.1, R1.2) for clarity.

H. The requirements for the ACA are currently placed in an Appendix A, pending executive decision on whether ACA is within the scope of ALMA. This may be deleted (if ACA is not part of ALMA) or promoted to a Chapter at a later date.

2 Pipeline Data Processing Requirements

Comments to: Frederic Gueth and Peter Schilke

2.1 General

We distinguish three different pipelines, the *Calibration Pipeline*, the *Quick-Look Pipeline*, and the *Science Pipeline*. The Calibration pipeline is intended for processing of array calibration data, usually on short turnaround time-scales, with feedback to the online system and into the archive. The Quick-Look pipeline has the job of providing quasi-real-time (∼minutes) or short turnaround-time (<hours) images, spectra, and data-quality assessment for feedback to the online system and to the observers, and possibly output to the archive. The Science pipeline is the primary data path from the array to the archive and to the observer, usually operating on longer timescales to produce results after breakpoints and after completion of projects (see ALMA-SW MEMO 11).

2–1.0–R1 The Pipelines shall be able to process all data coming from the array in standard modes designated by the project. It must not constitute a bottle-neck in the data flow, meaning that several occurrences of the same pipeline shall be able to run in parallel if necessary.

2–1.0–R1.1 Some projects will require unusually high data rates or processing requirements. These may require processing outside of the ALMA system and should be flagged appropriately so they are not processed by the ALMA pipeline.

2–1.0–R2 The Pipelines shall operate through readable and comprehensible data reduction scripts.

2–1.0–R3 Sufficient recording of corrections applied and/or models used shall be carried out so that any step can be reversed and redone if needed without recourse to repeating an entire series of operations or resorting to a copy of the dataset at the intermediate state.

2–1.0–R4 A manual, interactive mode of operations shall be available for debugging, technical developments, inspection by experts engineers and astronomers on duty, etc.

2–1.0–R5 The Pipelines shall also be run at the Regional Centers. Some of the actions described below are not relevant in that case: interaction with the Dynamic Scheduler, with the Sequencer.

2–1.0–R6 The pipelines should output a comprehensible summary of the operations performed with diagnostic information to allow checking of results and a record of the processing.

2.2 Calibration Pipeline

There are three general categories of calibrations that must be handled:

— The instrumental calibration: pointing, focus, delay, baseline, etc. What is required here is a fast feedback to the control software. In particular, there are critical calibrations (focus, pointing, delay) which must be executed successfully before telescope operations can resume - these are the most time-critical and have highest priority.

— The calibrations that do not require a time interpolation, as the atmospheric or bandpass calibration: each time such a scan is observed, something has to be derived and then stored, to be applied to all the following observations, until a new calibration of that kind is observed. Many of these can be immediately processed (e.g. bandpass) and stored for subsequent use.

— The calibrations that require a time interpolation (e.g. the phase and amplitude calibration, polarization leakage) where a calibration curve has to be fitted using data taken over a range in time. Some of these are local in time, and are applied to target observations that were observed in between calibrations. Others, once determined, are valid over a longer time period.

The first two are clearly handled by the calibration pipeline as described in this section. The third will likely be shared with the other pipelines, the Science Pipeline in particular for end-point calibration. It will probably be sufficient to include only the simple modes in the Calibration and Quick-look Pipelines, with more elaborate and mode-dependent calibrations occurring in the Science Pipeline, especially for calibrations that affect only the data in a given observation locally (e.g. fast-switching gain calibration).

2–2.0–R1 The Calibration Pipeline shall be activated after each scan has been observed.

2–2.0–R2 The Calibration Pipeline may also be re-invoked at any time with updated parameters or improved data. The results should not immediately overwrite old results so comparison is possible before adopting the new calibration. There will need to be a method for validation and acceptance of calibration updates.

2.2.1 Interferometric Calibration

2–2.1–R1 The Calibration Pipeline shall reduce, and store the following results for further use:

2–2.1–R1.1 the receiver sideband ratio calibration
2–2.1–R1.2 the atmospheric calibration

The results of the atmospheric calibration shall be passed to or made available for access by the Dynamic Scheduler (in real-time mode).

2–2.1–R2 For all observations of an astronomical source, the Calibration Pipeline shall:

2–2.1–R2.1 apply the atmospheric calibration to the data
2–2.1–R2.2 store the phase corrected from the atmospheric effect, if required

In subsequent operations, the corrected or uncorrected phase is to be used, depending on the selected mode.

2–2.1–R3 For all observations of a calibrator source, the Calibration Pipeline shall:

2–2.1–R3.1 compute the phase rms on the scan timescale
2–2.1–R3.2 compute the antenna efficiencies, using the averaged amplitudes
2–2.1–R3.3 do the previous operations both with and without the atmospheric phase correction, and deduce from the comparison whether the atmospheric phase correction improves the results or not
2–2.1–R3.4 derive amplitude and phase time-dependent variations by fitting smoothed curves (e.g. polynomials, splines) using all observations of calibrators since the beginning of the session

These results shall be passed to the Dynamic Scheduler, and made available to the Pipeline for gain transfer to target sources.

2–2.1–R4 The Calibration Pipeline shall reduce the following observations:

2–2.1–R4.1 pointing scans (results to be passed to the Sequencer)
2–2.1–R4.2 focus measurements (results to be passed to the Sequencer)
2–2.1–R4.3 delay calibration (results to be passed to the Sequencer)
2–2.1–R4.4 bandpass calibration
2–2.1–R4.5 baseline calibration
2–2.1–R4.6 holography measurement

2.2.2 Single-Dish Calibration

2–2.2–R1 The Calibration Pipeline shall reduce the atmospheric calibration, using sideband ratios determine from most recent interferometric calibration, and pass the results to the dynamic Scheduler.

2–2.2–R2 For all observations of an astronomical source, the Calibration Pipeline shall apply the atmospheric calibration to the data.

2–2.2–R3 The Calibration Pipeline shall reduce and pass the results to the Sequencer:

2–2.2–R3.1 pointing
2–2.2–R3.2 focus

Note that in most cases these determinations will be done interferometrically.

2–2.2–R4 For the pointing and focus measurements, the fitting results should be automatically stored in the telescope parameter file if the fitting error is less than the system specified value. If the error is not less than the specified value, the pipeline will send a message to the alarm system.

2–2.2–R5 The calibration pipeline shall derive the half-power beam size, the main-beam efficiency, and the Moon (fss) efficiency from the calibration scans toward planets and the Moon, and store the successful results in the telescope parameter file. Other derived parameters include the total forward efficiency obtained from skydip measurements, etc.

2.3 Quick Look Pipeline

2–3.0–R1 The Quick Look pipeline shall be activated after the Calibration Pipeline has been completed.

2–3.0–R2 A Monitoring Tool shall be available, plotting and archiving in a log file various results of the Calibration Pipeline, including:

2–3.0–R2.1 the results of the last pointing or focus scan
2–3.0–R2.2 the phase rms computed over the last scan and computed over the current session
2–3.0–R2.3 the corresponding seeing
2–3.0–R2.4 the atmospheric opacity

This tool shall include a variety of options, to control the plot parameters, to plot the variation of these results with time, to allow the operator to monitor one antenna or baseline in particular, etc.

Automatic checks shall be available to detect bad or degrading results, triggering alarms if necessary.

Since it is required that the observer/operator can efficiently check out the status of ongoing observations, all the plots by the monitoring tool should be reasonably simple, and the plot option should be able to quickly be changed by the observer/operator.

2–3.0–R3 A Monitoring Tool shall be available to plot the current properties of the array, such as:

2–3.0–R3.1 the current instantaneous uv coverage
2–3.0–R3.2 the corresponding weight distribution
2–3.0–R3.3 the corresponding dirty beam
2–3.0–R3.4 the previous quantities, integrated since the beginning of the session
2–3.0–R3.5 the thermal noise rms reached since the beginning of the session

2–3.0–R4 Single-Dish data — the current spectra observed on the astronomical target shall be corrected from the emission at a reference position or frequency (depending on the observing mode), and displayed with various options, such as:

2–3.0–R4.1 time integration
2–3.0–R4.2 antenna summation
2–3.0–R4.3 baseline fit, excluding a pre-defined window, or a window defined by the Operator or AoD
2–3.0–R4.4 spectra on a pseudo-grid corresponding to position on a raster (a "stamp" or "profile" plot)

2–3.0–R5 Interferometric data — the visibilities observed on a target source shall be calibrated, using the results of the Calibration Pipeline such as:

- **2–3.0–R5.1** apply the current bandpass calibration
- **2–3.0–R5.2** apply the current amplitude and phase correction
- **2–3.0–R5.3** apply the flux conversion factor based on standard antenna efficiencies

2–3.0–R6 Interferometric data — the current spectra observed on the astronomical target shall be displayed (amplitude and phase) with various options such as:

- **2–3.0–R6.1** time integration
- **2–3.0–R6.2** choice of the baseline(s)
- **2–3.0–R6.3** baselines summation
- **2–3.0–R6.4** intensity (amp or phase) as function of baseline and time (for a frequency), or time and frequency (for a baseline)

2–3.0–R7 Interferometric data — the Quick Look Pipeline shall compute the Fourier Transform of the visibilities, using the fastest algorithm, and display the resulting image. Alternatively, the actual Fourier Transform of each new visibility point can be computed and added to the current image. This shall be done for:

- **2–3.0–R7.1** the continuum data
- **2–3.0–R7.2** the line-averaged spectra, over a pre-defined velocity range, or possibly a velocity range defined by the Operator/AoD

2.4 Science Pipeline

2–4.0–R1 The Science Pipeline shall be activated after reaching a break-point (either user-defined or end-of-session normally).

2–4.0–R2 The Science Pipeline shall find in the Archive all data observed during the session. It shall use the atmospheric-calibrated data (amplitude and phase).

2.4.1 Interferometric Data

2–4.1–R1 The Science Pipeline shall use the calibrator to derive:

- **2–4.1–R1.1** the bandpass calibration
- **2–4.1–R1.2** the best phase and amplitude solution

2–4.1–R2 The Science Pipeline shall calibrate the source observations by applying:

- **2–4.1–R2.1** the bandpass calibration
- **2–4.1–R2.2** the phase calibration
- **2–4.1–R2.3** the amplitude calibration

2–4.1–R3 The Science Pipeline shall check and correct the flux scale by using observations of source of known fluxes. Any effect due to the source being resolved shall be taken into account.

2–4.1–R4 The Science Pipeline shall compute images for each (non-blanked, possibly user-specified) frequency channel, as well as for the continuum emission:

- **2–4.1–R4.1** extract the visibilities with the appropriate frequency resolution plus the continuum measurements
- **2–4.1–R4.2** find in the Archive the previous (calibrated) visibilities, and check whether the data are compatible with the current dataset
- **2–4.1–R4.3** grid the whole data set

2–4.1–R4.4 Fourier transform

2–4.1–R5 The images shall be deconvolved using the most appropriate algorithm. In case of a complex image, it should be possible to have several algorithms running in parallel, the best (according to criteria TBD) image being eventually selected.

2–4.1–R6 Designated modes shall be supported, including:

2–4.1–R6.1 mosaic observations
2–4.1–R6.2 on-the-fly mosaics
2–4.1–R6.3 self calibration projects
2–4.1–R6.4 combination of single-dish + ALMA data (+ACA)

Comment: Careful cross calibration of the flux scales between ALMA interferometric data and single dish data (and ACA) is required for high fidelity imaging. This will require careful coordination with the calibration pipeline, especially as ACA observations may be taken at very different times than the main array data.

2–4.1–R7 Subtraction of continuum level from spectral data is required. This can be done in both Fourier and image domain. In the case of uv-plane subtraction, flexible setting of the frequency channel ranges for the calculation of the continuum level should be available.

Comment: There will likely need to be a way to make trial subtractions and select the "best" in an automated manner for the pipeline to function.

2.4.2 Single Dish Data

2–4.2–R1 The data taken on the astronomical source shall be reduced, depending on the the observing mode. All ALMA modes (as designated by the project) shall be supported, including:

2–4.2–R1.1 on/off
2–4.2–R1.2 nutator switch
2–4.2–R1.3 frequency switch
2–4.2–R1.4 raster maps using one of the above modes
2–4.2–R1.5 OTF maps using one of the above modes

2–4.2–R2 The resulting spectra shall be corrected for a baseline, fitted on all spectra channels but a pre-defined window.

2–4.2–R3 The Science Pipeline shall:

2–4.2–R3.1 find in the Archive previous (calibrated) observations, and check whether the data are compatible with the current dataset
2–4.2–R3.2 grid the whole data set

2–4.2–R4 Provision shall be taken to allow running appropriate algorithms (deconvolution, destriping), if required by the data or by the experience gained using ALMA.

2.5 Interface with the Archive

TO BE DETAILED

2–5.0–R1 The images produced by the Science Pipeline shall be archived, together with the:

2–5.0–R1.1 the script that was used to produce the image
2–5.0–R1.2 the log file of the software

2–5.0–R2 cf 7.0-R3 general SSR document

2.5.1 Data Monitor and Control Quantities

2–5.1–R1 Data quality control:

- **2–5.1–R1.1** estimate of the noise
- **2–5.1–R1.2** seeing
- **2–5.1–R1.3** image fidelity based on model?

2–5.1–R2 observation quality control:

- **2–5.1–R2.1** baseline quality
- **2–5.1–R2.2** calibration quality

2–5.1–R3 telescope state: (possibly in monitor file, but accessible)

- **2–5.1–R3.1** telescope pointing
- **2–5.1–R3.2** subreflector focus
- **2–5.1–R3.3** monitor point (e.g. temperatures) data

Appendix: Barry Clark's list of input parameters needed for each procedure NOW MOVED TO APPENDIX AT END OF DOCUMENT

3 Offline Data Processing Requirements

3.1 General Requirements and Interaction with Other ALMA Elements

3.1.1 Goals of the Offline Package

3–1.1–R1 An ALMA Offline Data Reduction Package (or "the package") is primarily intended to enable end-users of ALMA (e.g. observers or archive users) to produce scientifically viable results that involve ALMA data products. The secondary use is to enable ALMA staff to assess the state of the array and derive calibration parameters for the system.

3–1.1–R2 The package should be able to function (be installed) at the users home institution and at ALMA regional centers (both locally and remotely). It should be portable to a reasonable number of supported platforms, including laptops without network connections.

3–1.1–R3 The performance of the package should be quantifiable and commensurate with the data processing requirements of ALMA output at a given time. This should be benchmarked (e.g. "AIPS-marks") and reproduce accurately results for a fiducial set of reduction tasks.

3.1.2 Relation to the Pipeline

3–1.2–R1 All modules available in the pipeline must be available also as an offline analysis option.

3–1.2–R2 Note that not all offline analysis tools will necessarily be in the pipeline package. For example, one of the important differences between pipeline and offline reduction path is that offline one should have extensive interactive capabilities to merge and compare data with different resolution, coordinate system, data grid, and so on.

3.1.3 Operational Issues

3–1.3–R1 Installation of package must be flexible, and able to be installed on non-specialized hardware by end user, preferably without root user permission (on Unix).

3–1.3–R2 Updates and new versions of the package should be backward compatible where possible, such that user-built and observatory provided scripts and tools should be executable with only minor changes.

3–1.3–R3 User installation of the package should not be restricted by other issues such as expensive or unduly restrictive licenses (ie. the package license should convey all other necessary licenses, such as GNU).

Comment: Although it may be attractive to build upon a commercial package such as IDL, this is likely to be prohibitive unless a blanket license would be bought by the project and distributable free of charge to users.

3.2 Interface

3.2.1 General User Interface Requirements

3–2.1–R1 User must be able to choose from a variety of interface styles, including:

3–2.1–R1.1 A Command Line Interface (CLI) must be provided, with access via both an interactive input and via script

3–2.1–R1.2 A Graphical User Interface (GUI) must be provided for interactive processing. Actions taken under the GUI must be loggable and convertible into scripts executable by the CLI.

3–2.1–R2 The user should be able to interact with the host operating system with command sequences invoked from the UI.

3–2.1–R3 Multitasking for all interfaces should be available where appropriate.

3–2.1–R3.1 It must be possible to run one or more long-running calculations in the "background." While background tasks are running normal interactive activities must be possible.

3–2.1–R4 User-selectable output display devices will be supported.

3–2.1–R5 User-understandable and non-destructive error handling at all levels is highly desirable.

3–2.1–R6 The application of successive stages of calibration, correction, flagging and editing should not be destructive to the data. The package should be able to recover and revert to earlier stages without repeating an entire series of operations.

3–2.1–R7 The interface and package should function without a network connection. Users should be able to (conveniently) run the data processing user interface from a laptop (e.g. on an airplane).

3.2.2 Graphical User Interface (GUI)

3–2.2–R1 The GUI should provide real-time feedback via standard compact displays.

3–2.2–R2 The default look and feel of the GUI should be uniform and familiar through the entire package.

3–2.2–R3 The look and feel of the GUI must be customizable to accommodate both the expert specialist and the novice user, with ability to hide complexity when prudent and the ability to access deeper levels when desired. The default look to the novice should not be overly busy, with functionality easily apparent through labeling and built-in help facility.

3–2.2–R4 The look and feel of the GUI must be customizable to accommodate both the expert specialist and the novice user, with ability to hide complexity when prudent and the ability to access deeper levels when desired. The default look to the novice should not be overly busy, with functionality easily apparent through labeling and built-in help facility.

3–2.2–R5 All functionality of the CLI must also be available in GUI mode.

3–2.2–R6 A graphical data-flow oriented (IDL style) tool assembler would be desirable, perhaps as an advanced GUI for later development.

3.2.3 Command Line Interface (CLI)

3–2.3–R1 The CLI must be usable remotely over low-speed modem lines or network connections, with ACSII terminal emulation.

3–2.3–R2 The interface must have the facility to read in command files for batch processing of a sequence of CLI commands.

3–2.3–R3 The CLI should have command-line recall and editing, name completion and minimum match where appropriate.

3–2.3–R4 All functionality of the GUI must also be available in CLI mode (although posssibly with loss of simplicitity in instances where the graphical selection is important).

3.2.4 Interface Programming, Parameter Passing and Feedback

3–2.4–R1 The UI must have basic programming facilities such as:

- **3–2.4–R1.1** variable assignment and evaluation
- **3–2.4–R1.2** conditional statements
- **3–2.4–R1.3** control loops
- **3–2.4–R1.4** string manipulation
- **3–2.4–R1.5** user-defined functions and procedures
- **3–2.4–R1.6** standard mathematical operations
- **3–2.4–R1.7** vector and matrix manipulation

3–2.4–R2 Commands executed should be logged, with provision to re-execute the session.

3–2.4–R3 Input parameter checking upon parsing with reporting of incorrect, suspicious or dangerous choices should be done before execution where possible.

3–2.4–R4 Parameters should be passable between applications in as transparent a manner as possible. However, "global" variables should not be the default, unless designated specifically by the user-programmer.

3–2.4–R5 Application variables should be named consistently and as clearly as possible indicating their intended use using astronomical terms where possible.

3–2.4–R6 There shall be no hidden parameters, all changeable parameters in all actions must be accessible. In complex cases, this should (must?) be through an hierarchical interface with the most important parameters accessible directly, and the others through sub-menus.

3.2.5 Documentation and Help Facility

3–2.5–R1 The package creators must provide comprehensive and user comprehensible documentation for all parts of the package.

3–2.5–R2 There should be a variety of help levels and documentation formats accessible from the UI and over the Internet, applicable to novices, experts, and technical users. These would include:

- **3–2.5–R2.1** user cookbooks with extensive examples
- **3–2.5–R2.2** application descriptions and reference manual
- **3–2.5–R2.3** programmer references
- **3–2.5–R2.4** data format descriptions
- **3–2.5–R2.5** algorithm descriptions
- **3–2.5–R2.6** online help, FAQ, email contacts
- **3–2.5–R2.7** release history, bug reports and tracking, patch descriptions
- **3–2.5–R2.8** newsletters, email exploders, notes series
  These would be maintained by the package providers, with help from the ALMA project.
- **3–2.5–R2.9** These should be in printer-friendly formats.
- **3–2.5–R2.10** Optional native HTML formats are desirable.

3–2.5–R3 Help should be context-sensitive. In GUI mode, fly-over banners should indicate use of buttons and fields, and clickable help buttons should be available on all pages.

3–2.5–R4 Help functions can direct user to Web page, although in CLI mode, must support in-line text based help also.

3–2.5–R5 Full search capability must be built into documentation library.

3.3 Data Handling

3.3.1 General Data Requirements

3–3.1–R1 The package must support data taken in any of the available ALMA hardware modes in the most appropriate manner

3–3.1–R2 Must be able to handle the integrated data objects corresponding to the observational programs carried out by ALMA. These objects may be implemented in any manner appropriate, though relations between the components of the object must be maintained through some mechanism. These include:

- **3–3.1–R2.1** Program header information
- **3–3.1–R2.2** Observation status information (and schedules themselves)
- **3–3.1–R2.3** Field information
- **3–3.1–R2.4** Coherence function (visibility) data from interferometer in all available polarization products, spectral channels, frequency bands, IFs, including auto-correlations
- **3–3.1–R2.5** Auto-correlations in single-dish total power modes
- **3–3.1–R2.6** Weights and/or data uncertainties
- **3–3.1–R2.7** Flagging data or masks
- **3–3.1–R2.8** Diagnostic data and errors
- **3–3.1–R2.9** A-Priori calibration data (bandpasses, flux densities, polarization leakages, etc.)
- **3–3.1–R2.10** Derived calibration data (gain tables, flux bootstraps, etc.).
- **3–3.1–R2.11** Images and/or models produced from data
- **3–3.1–R2.12** Processing history

3–3.1–R3 Must support data taken in one or more polarization products, spectral channels, frequency bands, IFs. Transformation must be provided to the desired Stokes output parameter(s) with facility for spectral or band averaging.

3–3.1–R4 Multiple pointing centers for mosaics must be supported.

3–3.1–R5 Multiple subarrays must be supported.

3–3.1–R6 Data taken in arbitrary scanning patterns must be dealt with.

3–3.1–R7 The flagging mask must be maintained and associated with the data it refers to during any subsequent operations (such as splitting of data sets).

3–3.1–R8 Calibration and ancillary monitoring data must be preserved.

3–3.1–R9 Comprehensive and understandable processing history information for the data must be maintained and be exportable.

3–3.1–R10 Distinctions between "single-source", "multi-source", single-dish, and interferometer datasets should be avoided with context built into the dataset or header.

3–3.1–R11 When sorting or indexing is desirable for performance enhancement, this should be carried out in a manner transparent to the user.

3–3.1–R12 Concatenation of datasets should be straightforward and robust. Extraction and reinsertion of data subsets should be supported.

3–3.1–R13 Users should have access (at the manipulation level) to all aspects of the data including the header.

3–3.1–R14 The package must support locking data files so that there is no possibility of one process corrupting a file that is also being written to by another process. The model should be: "one writer, multiple readers."

3.3.2 Data Import and Export

3–3.2–R1 Standard data formats (e.g. FITS) must be supported for both input and output without loss of functionality or information, though need not be the native format for both the package and archive. The project will maintain a list of formats which the package must support.

3–3.2–R2 Access to the archive must be supported, including for data from the currently active observing session. Security and integrity of the archive must be ensured during these operations.

3–3.2–R3 Disk and offline data storage (e.g. DAT, DDS, DLT) must be supported. The project will maintain a list of media which the package must support.

3–3.2–R4 The ability to drop flagged data on export should be included.

3.3.3 I/O Speed and Efficiency

3–3.3–R1 I/O of data must not be a bottleneck for processing, especially for pipeline use. The definition of what constitutes a "bottleneck" and what I/O throughput rate is acceptable must be defined at each stage of ALMA operations (e.g. interim science, full stand-alone ALMA, ALMA + ACA) and in each mode (e.g. quick-look pipeline, offline use).

3–3.3–R2 I/O failure (e.g. due to full disks, device offline) should not result in catastrophic failure. Error recovery should be as robust as possible.

3.3.4 ALMA Interferometer Data

3–3.4–R1 Correlation products accumulated at multiple bit depths (16-bit,32-bit) must be supported transparently.

3–3.4–R2 On-line gain correction data must be carried along with data, unless specified.

3–3.4–R3 Calibration tables and editing information must be associated with the data and preserved on output.

3.3.5 ALMA Single Dish and Interferometer Phased-array Data

3–3.5–R1 Data taken with nutating secondary must be supported, as a function of nutation phase.

3–3.5–R2 Total power and phased array data sequences with scanning pattern preserved.

3.3.6 Images and Other Data Products

3–3.6–R1 Standard multi-dimensional images must be supported, such as:

- **3–3.6–R1.1** Spectra and image slices (1D)
- **3–3.6–R1.2** Planar images (2D)
- **3–3.6–R1.3** Spectral and Time Cubes (3D)

3–3.6–R2 Standard derived data products must be supported, such as:

- **3–3.6–R2.1** Models (e.g. CLEAN models, Gaussian models, wavelets, pixons)

3–3.6–R3 Blanking of pixels (magic-value) must be maintained through the processing of images.

3.3.7 Foreign Data

3–3.7–R1 Data produced by other interferometers and single dishes in similar observing modes should be importable and processable if provided in a standard data exchange format.

3–3.7–R2 Imaging data in standard formats (e.g. FITS) from astronomical instruments at different wavelengths should be importable, with the ability to combine (coadd) these with ALMA data where appropriate. This should be through a set of widely used formats, with a minimal list of supported standards established by the project.

3.3.8 Interaction with the Archive

3–3.8–R1 The interface between the package and archive must be able to provide data access (when such access is granted) without interfering with other access to the archive.

3–3.8–R2 Products of imaging and data analysis should be able to be archived in association with the data if desired. This means output formats supported by the archive must be supported by the package.

3.4 Calibration and Editing

3.4.1 General Calibration and Editing Requirements

3–4.1–R1 The package must be able reliably handle all designated ALMA standard calibration modes, including but not exclusive to temperature controlled loads, semi-transparent vanes, apex calibration systems, WVR data, noise injection, fast-switching calibration transfer, planetary observations.

3–4.1–R2 Calibration, editing, flagging, and correction of data in the off-line and pipeline package should be easily reversible within the process (ie. not requiring re-reading of the data from the archive).

3–4.1–R2.1 Logging of individual editing steps should be clearly marked in some sort of history table (possibly distinct from a more readable history) with individual edit undo desirable.

3–4.1–R3 Data display and editing should be effected through generic tools applicable to both single-dish and interferometer modes. These should present should, as far as possible, present similar interfaces to the user and have the same look-and-feel.

3–4.1–R4 Data editing and flagging should be possible based upon array and environmental monitoring data.

3–4.1–R5 Data calibration, correction and flagging should be possible based upon standard or user-defined models in either functional or tabular form.

3–4.1–R6 Calibration should involve flexible averaging of data and calibration quantities with user-controllable interpolation, filtering, weighting, and application scope.

3–4.1–R7 Interactive data editing, calibration, and display of calibration quantities should be largely graphical and intuitive, with user-definable setups. Displays of greyscale (e.g. TVFLG style) or line plots (Difmap style) should be options. Visualization of the calibration results will be key to obtaining robust and trustworthy calibration with ALMA.

Specialized editing display tools should include:

3–4.1–R7.1 Specification of data by selection of time range, uv range, pointing center
3–4.1–R7.2 Displays of spectra and spectral cubes, with time and or channel averaging
3–4.1–R7.3 For interferometer data, amplitude (phase) vs. time on each baseline (Difmap vplot), time-baseline (TVFLG) with interactive zoom and clipping.

3–4.1–R7.4 Editing based on difference from a running mean, or rms in boxcar, or difference versus model.

3–4.1–R7.5 Cuts through a data plane or cube.

3–4.1–R8 Editing should be incorporated into most visualization tools where data or data-derived quantities are plotted, such as from calibration solutions, amplitude vs. uv-distance plots, or any number of other plots. A "see-it, flag-it" capability should be the standard within the tools.

3–4.1–R9 Automatic editing tools should be available in the package.

3–4.1–R10 Access to time history of calibration information such as source catalogs containing flux density histories, planetary ephemerides, noise tube values, should be built into calibration engines. Output of calibration procedures should be exportable into similar structures.

3.4.2 Interferometer Data

3–4.2–R1 Antenna-based determination of calibration quantities such as gains, polarization leakages, bandpasses, is the primary form of calibration where appropriate.

3–4.2–R2 In addition to antenna-based calibration, baseline dependent corrections will also be supported. For example, coherence loss due to atmospheric phase fluctuation depends on baseline length (this aspect will be more important at higher frequencies) and must be taken into account if some of the WVR corrections are deemed incorrect while others are applied. Also, in general, the bandpasses are baseline dependent and contain non-closing terms.

3–4.2–R3 Gain corrections will be made based on differences between observed and modeled data quantities, possibly with iteration (e.g. self-calibration and determination of gains using calibration sources). Where solutions are discrepant or poor automatic edition should be possible.

3–4.2–R4 Redundancy (e.g. same or crossing baselines) should be used wherever possible to increase accuracy of or check calibration solutions. Editing base on this comparison should be possible.

3–4.2–R5 Calibration quantities (possibly stored in tables or data structures) should be transferable between sources, possibly after interpolation, extrapolation or smoothing. This will be the primary method of phase calibration transfer using fast-switching between source and calibrator.

3–4.2–R6 Determination of, correction for, and examination of closure errors should be straightforward to carry out.

3–4.2–R7 Determination of the complex bandpass using calibration source observations, and transfer to target sources, should be simple and robust.

3–4.2–R8 Interferometric pointing, baseline, and beam response fitting should be available.

3–4.2–R9 Determination of polarization calibration quantities such as leakage (D-term or Jones matrix) and complex gain difference must be an integral part of the package, using both linearized and full matrix calculations.

3–4.2–R10 Incorporation of standard models (e.g. planetary disks, models for HII region structure, known source spectra) should be easy for calibration operations.

3.4.3 Single Dish Data

3–4.3–R1 Processing for pointing, tipping, focusing, beam-fitting data must be available.

3–4.3–R2 Straightforward and flexible fitting of spectral bandpass from calibration source observations is required

3–4.3–R3 Calibration of temperature controlled loads and noise sources from observations of celestial sources should be supported

3–4.3–R4 Final data scaling must be possible in case that the 1calibration fails for unexpected reason. A user may want to make their manual scaling, e. g. by referring to the line intensity of the map reference center, to compensate for the daily/time variation.

3–4.3–R5 For single-dish OTF observations with the two orthogonal scan directions (e.g., N-S and E-W), the intensity scale can be adjusted to minimize the "scan effect" due to slight variation of gain.

3.4.4 Mosaicing Considerations

3–4.4–R1 Individual data points must be associated with pointing center information, and one must have the ability to deal with complex scanning strategies.

3–4.4–R2 Determination of and correction for pointing offsets and the beam shape is critical to the ability to reliably mosaic using ALMA, and thus must be available in several algorithmic forms in the package.

3–4.4–R3 The complex polarization response of the telescope beams must be calibratable (though this is mostly an imaging step).

3–4.4–R4 Careful cross calibration of the flux scales between ALMA interferometric data and single dish data is required for high fidelity imaging (important and more difficult for ACA data). There must be tools to cross-check and correct the relative calibration between mosaics and different component observations.

3.4.5 Ancillary and Diagnostic Data

3–4.5–R1 Environmental data such as weather (e.g. wind speed, temperature, dew point) should be available for editing or calibration procedures, and easily incorporated into user-specified calibration models.

3–4.5–R2 Engineering monitoring information such as temperature sensor readings and tilt-meter outputs, perhaps included as ancillary tables attached to data files with special keywords, should be readable and incorporated into the calibration and editing process.

3–4.5–R3 Output from the atmospheric monitoring (e.g. WVR, FTS) instrumentation should be processed and used by calibration software.

3–4.5–R4 Pointing, focus and subreflector information must be dealt with appropriately.

3.5 Imaging

3.5.1 General Imaging Requirements

3–5.1–R1 Imaging data selection from any combination of ALMA exported data, the ALMA archive, or other instruments supporting common export formats must be provided.

3–5.1–R2 Efficient selection of subsets of the imaging data must be provided.

3–5.1–R3 Provision must be made for the utilization and development of a variety of imaging, deconvolution, and analysis algorithms (e.g. flavors of CLEAN, MEM, linear and non-linear mosaics)

3–5.1–R4 Astrometric accuracy must be preserved over phase-calibration distances of at least 5 degrees.

3–5.1–R5 Images made on the different equinox (e.g. B1950 and J2000) or different coordinate (RA,DEC and l,b) system or different projection (tangent, sinusoidal, ...) can be merged and compared appropriately.

3–5.1–R6 Data cubes using different velocity definition (optical or radio definition for Doppler velocity) must be merged appropriately.

3–5.1–R7 Image pixel blanking should be supported.

3.5.2 Interferometer Imaging

3–5.2–R1 High-fidelity imaging of the entire primary beam in all Stokes parameters is the primary goal — therefore, incorporation of the polarized primary beam response of the array is required.

3–5.2–R2 Imaging must deal seamlessly with mosaiced data, with proper gridding in the uv-plane and compensation for primary beam effects and pointing in such a manner as to mitigate the effects of non-coplanar baselines and sky curvature. A variety of options for gridding and beam correction should be available at user request.

3–5.2–R3 There must be seamless integration of data from multiple epochs and configurations

3–5.2–R4 There must be the ability to include short-spacing data taken in single-dish mode (both ALMA and non-ALMA data)

3–5.2–R5 Subtraction of continuum level from spectral data is required. This can be done in both the Fourier and image domain. In the case of uv-plane subtraction, flexible setting of the frequency channel ranges for the calculation of the continuum level (graphically as well as CLI) should be available.

3–5.2–R6 The creation of 3D images for rotating object (e.g. planets) should be supported.

3.5.3 Single Dish Imaging

3–5.3–R1 The package must be able to produce an image by using data observed at different spacing (or even at random positions) must be spatially interpolated or re-gridded correctly (e.g. 1 arcmin RA-DEC grid observations and 1 arcmin l-b grid observations; two sets of 1 arcmin grid observations using two different map reference centers (0,0) and (0'.2, 0'.6))

3–5.3–R2 The package should have the capability to produce the stamp map (profile map) of spectra at a grid of positions.

3.5.4 Mosaicing Considerations

3–5.4–R1 Combination of interferometer and single-dish data into mosaic imaging is essential.

3–5.4–R2 Careful (polarized) primary beam correction and pointing correction is critical for high fidelity mosaic imaging and must be incorporated into the mosaicing algorithms.

3.5.5 Inclusion of the ACA

3.6 Analysis

3.6.1 General Analysis Requirements

3–6.1–R1 The astronomer must have the capability to develop their own tools or tasks, with easy access to data and images, and straightforward interface with the package

3–6.1–R2 Translation between various astronomical quantities and units (e.g. Jy and K, MHz and km/s) should be straightforward and user selectable.

3.6.2 Visibility Ddata Analysis

3–6.2–R1 UV-plane analysis based on goodness-of-fit to a model will be required.

3.6.3 Single Dish Data Analysis

3–6.3–R1 Automatic measurement of line parameters (line intensity, integrated intensity, Gaussian-fit line width, rms noise level, ...) for user specified velocity (frequency) window must be made, and to be stored in a text-format list file that can be output by the user if desired.

3–6.3–R2 Effective, robust and precise spectral baseline removal facility is required. Fourier analysis of standing waves and their removal will be a critical task.

3–6.3–R3 Fourier transform to a pseudo-uv plane "image" will be needed.

3.6.4 Mosaicing and Combined Array Analysis

3–6.4–R1 Seamless transformation between image-plane and uv-plane analysis is necessary.

3.6.5 Image Analysis and Manipulation

3–6.5–R1 The ability to extract lower-dimensional "slices" from n-dimensional data cubes efficiently is required.

3–6.5–R2 The ability to collapse or integrate over sub-dimensions of data cubes in order to form "moments" is required.

3–6.5–R3 Blanking of pixels (magic-value) must be maintained through the analysis process. It is desirable that blanks not be destructive (the original pixel value is retained), and it be possible to turn on and off different blanking ("mask") levels.

3–6.5–R3.1 Interactive and automatic facilities for setting of blanking parameters (e.g. windowing, S/N based blanking) to avoid degrading S/N in the analysis must be provided.

3.7 Visualization

3.7.1 General Visualization and Plotting Requirements

3–7.1–R1 Plotting and display capabilities should be integrated into the GUI tools throughout the package. Where possible the displays should have similar look and feel to reduce the plotting learning curve.

3–7.1–R2 Plotting of data and calibration quantities as a function of time (LST,UT,etc.) in both standard X-Y and image cube (with additional axes) must be available.

3–7.1–R3 Where appropriate, edition and flagging capability should be incorporated into all plots. Basically, if you "see-it" you should be able to "flag-it".

3–7.1–R4 The output of the display should be possible in many different designated formats, including fits, postscript, pdf, gif and jpeg.

3.7.2 Display Appearance and Interactivity

3–7.2–R1 User should be able to produce overlays of different data sets of standard formats. It should be possible to place these data sets in layers which can be switched on and off separately. The different images should be editable, and it should be possible to declare certain colors transparent. It must be possible to shift, rotate and scale the images at will.

3–7.2–R2 It must be possible to display and overlay data with different coordinate systems, i.e. the coordinate system of the display can be chosen independent of the system the data were observed in.

3–7.2–R3 Both contour plots with variously colored and styled lines and false color maps should be possible, it should also be possible to produce RGB overlays (i.e. one layer gets assigned intensity scales of red, another one of green, and one of blue), or Hue/Saturation/Value.

3–7.2–R4 User should be able to manipulate intensity and color scales easily and graphically, the setup achieved should be saveable and reloadable.

3–7.2–R5 User should be able to add annotation, both interactively and through scripts, for publication quality plots, i.e. text with various fonts (including Greek letters), symbols (e.g. all the symbols provided by the LaTeX package with AMSTeX extension), arrows, geometrical figures like boxes and circles etc. Different line styles, sizes, thicknesses and colors for all those should be available. The various elements should be editable and removable separately, and it should be possible to put them in a separate layer.

3–7.2–R6 The display tool should have astrometry facilities, i.e. based on catalogs and by assigning sources in the maps it should be possible to calculate the coordinate system.

3–7.2–R7 The display for spectra must be linked to molecular data bases which make identification of the lines possible.

3.7.3 Image-cube Manipulation

3–7.3–R1 Interactive display of spectra corresponding to a displayed image should be supported. For example, the display of spectrum by clicking the the image map on the display, for the position nearest the cursor position. Also, dragging the line on the map to bring up a position-velocity diagram.

3–7.3–R2 Plotting of spectra on a pseudo-grid corresponding to position on a raster (e.g. a "stamp map" or "profile map", basically thumbnail spectra in panels corresponding to position) should be possible.

3–7.3–R3 Data cubes should be viewable as movies with varying frame rates.

3–7.3–R4 It should be possible to view arbitrary subsets or slices of data cubes.

3.8 Simulation

3.8.1 General Simulation Requirements

3–8.1–R1 There must be simulation capability for interferometer and single dish observation with ALMA in all modes, for planning (with the ObsTool) and comparison of data with models (for editing and correction). Various simulator components with different levels of complexity and execution speed will be necessary to carry out desired tasks, such as:

3–8.1–R1.1 Level 1 — Simple expected sensitivity levels given integration time, configuration, mosaicing strategy, atmospheric quality limits, used for proposal and schedule preparation in

the Observing Tool, though also useful offline to check basic data properties. Timescale for execution should be 0.1-5 minutes and should not require significant computational resources.

3–8.1–R1.2 Level 2 — Basic dataset simulation tools to generate fake data given observing parameters and simple models of the instrument and atmosphere. These should include error generation for thermal noise, pointing, primary beam, atmosphere temperature and mean structure function, antenna, optics, receiver, and correlator efficiencies, etc. This is primarily for use in the pipelines and also for testing of other software components. Timescale for execution should be 1-30 minutes and may require significant cpu and memory resources.

3–8.1–R1.3 Level 3 — More complex instrumental modeling, most useful for project staff and engineers. Timescale of execution is indefinite and will likely require special resources (e.g. parallel computing).

3–8.1–R2 The output of the simulator must be compatible with the rest of the offline package, and with the ALMA pipeline. It should be available in all ALMA data format(s).

3–8.1–R3 The speed of the simulator must be commensurate with the desired feedback time. For instance, if used with the real-time-system to assess quality the simulator must respond in minutes, if used for proposer feedback for ObsTool application it should feedback also on minute timescales for most simple experiments, while for complicated engineering simulations it may be allowed to take correspondingly longer.

3–8.1–R4 Relevant parts of the simulator (e.g. simple single field and mosaic dataset generation with thermal noise and pointing errors) should be available early in the software production cycle in order to use it to test other components of the package.

3.8.2 Interferometer Simulation

3–8.2–R1 All baseline correlator modes should be supported in the simulation.

3–8.2–R1.1 If an advanced correlator is constructed, appropriate simulation modes should be added to the package.

3–8.2–R2 A primary use of simulation capability of the package is to compute expected visibilities for trial models during the calibration, imaging, or analysis stages, and thus must be integrated to this extent with those tools.

3–8.2–R3 Realistic inclusion of telescope primary beam response (with polarization), gain fluctuations, bandpasses, atmospheric effects, correlator errors, and other effects must be supported. It must be relatively easy for the user to modify the simulator to include new error terms.

3.8.3 Single Dish Simulation

3–8.3–R1 All scanning and subreflector nutation modes must be supported.

3–8.3–R2 Baseline drift and standing wave effects must be included in spectra.

3.8.4 Incorporation of Previous or Foreign Data

3–8.4–R1 Simulation software must have the capability to incorporate ALMA data taken previously, assuming the observer has been granted access to this data.

3–8.4–R2 Simulation software must have the capability to incorporate user-supplied input models and/or data in ALMA-supported form.

3.8.5 Interaction with Observing Tool

3–8.5–R1 A primary use of the ALMA simulator is to provide guidance during the proposal and schedule preparation phases. As such, it is critical that the relevant parts of the simulation software be compatible with the Observing Tool, preferably integrated seamlessly into its interface.

3.9 Special Features

3.9.1 VLBI

3–9.1–R1 It is assumed that the major processing of VLBI data from ALMA will be outside the ALMA package, and only parts of the processing necessary to produce usable single-dish or phased array data and the associated calibration must be supported by ALMA software.

3–9.1–R2 ALMA must provide the export of VLBI data in a suitable standard format.

3.9.2 Solar Observing

3.9.3 Pulasr Observing

3.9.4 Other Modes

4 Common Algorithms

Do we want a section here to point to in the previous sections? Who will write this part?

A The ACA

to be written