CALIBRATION AND DATA PROCESSING STRATEGIES FOR ALMA

Robert Lucas(1), **Brian Glendenning**(2), **Gianni Raffi**(3)

(1)***Institut de RadioAstronomie Millmétrique, 300, Rue de la Piscine, 38406, Saint Martin d'Hères, France (lucas@iram.fr)***

(2)***National Radio Astronomical Observatory, P.O. Box 0 Socorro, NM 87801, USA (bglenden@nrao.edu)***

(3)***European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei Mnchen, Germany (graffi@eso.org)***

ABSTRACT

During the course of the next decade ALMA should become operational as the most powerful and versatile facility for ground-based astronomical observations in the millimeter and submillimeter wavelength ranges. The goal of ALMA is to provide spectral images of the sky that can be directly used by non-technical astronomers. This motivates the design of an automatic system for on-line calibration and data processing. We sketch the instrument itself and its software organization, concentrating on the on-line and off-line calibration and data processing issues.

INTRODUCTION

ALMA (Atacama Large Millimeter Array) is a joint project involving astronomical organizations in Europe and North America. ALMA will consist of 64 12-meter diameter antennas operating in the millimeter and sub-millimeter wavelength range. Each antenna will have a surface accuracy better than 25 μm and point within $.6''$. ALMA will allow precision imaging on sub-arc-second scales at millimeter and sub-millimeter wavelengths. Among its science objectives are studies of the formation and early evolution of galaxies and active galactic nuclei, of the structure and contents of galactic molecular clouds, of the formation processes of stars and planetary systems, of the latest phases of stellar evolution and of solar system phenomena.

GENERAL SCIENCE SOFTWARE REQUIREMENTS

We give here an overview of ALMA Science Software Requirements. They are detailed in an internal memo [1], available on the Web.

The operation of ALMA will have to deal with a larger variety of projects than previous instruments: on one hand at long wavelengths (1-3 mm) due to the high sensitivity and quality of the site, and a long experience with millimeter-wave interferometry, we can predict with reasonable certainty the observing modes that will be used, the relevant observing strategies to schedule the instrument, and the data reduction techniques. On the other hand aperture synthesis at the highest frequencies ($\sim 300\mu$m) has only started very recently; we plan to rely on techniques such as radiometric phase correction, fast phase switching and phase transfer between frequency bands, that have been demonstrated, but not applied with the operational scale that we foresee for ALMA. We thus will have to combine in the software a high level of automation, needed to deal with the large information rate that will be available, with a high level of flexibility at all levels to be able to develop and implement new observing methods and reduction procedures. For simple projects the astronomer with little or no experience of radio techniques should be able to use the instrument and obtain good quality results; however experts should easily be able to perform experiments we do not even foresee today.

The expert user/developer will need to be able to send direct commands to the instrument through a simple, easily editable command language. Atomic commands in a script language will directly send orders to the basic software elements controlling the hardware: antenna motion, instrument setup, or transmitting parameters to the data processing (pipeline). The script language will support loops, structured conditional tests, parametrized procedures, global variables and arrays, and other programming constructs. These scripts, once fully developed and tested, will evolve into the basic observing procedures of the instrument.

The general user will need more user-friendly graphic interfaces to many components of the system. They will be proposed several standard observing modes, and be provided a simple way to pass astronomy parameters to the basic observing process, and to the corresponding data reduction procedures of the pipeline. Input parameters will preferably be expressed in terms of astronomical quantities, which will be translated into technical parameters by sophisticated configuration tools.

Proposal submission will be in two phases, the first before proposal evaluation, the second to provide information needed for the actual scheduling and observation.

Dynamic scheduling is an essential feature of the instrument and should be installed from the very beginning of its operational life. Though the site is undoubtedly one of the best for submillimeter observations, it will only be usable at the highest frequencies for a fraction of the time; to improve the total efficiency we must be able to make the best use of all weather conditions, by selecting in quasi-real time the project most suited to the current weather and to the state of the array. This means we should always be able to observe a given project in appropriate weather conditions. This philosophy can be extended to the point where a given project can change its own observing parameters (such as cycle times for phase and amplitude calibrations) according to variations in observing conditions (such as atmospheric phase rms).

The whole real-time system will be under control of a telescope operator, through a specially designed interface. This must provide an overview of what observation is occurring, the state of the instrument, and observing conditions on the site, and should enable the operator to react to any unexpected event. A general monitoring interface must be also accessible through the network.

The instrument should produce images, aiming to be final for most projects, even when projects are spread over several sessions and configurations, and/or include short/zero spacings. For this purpose an on-line pipeline is required. It will include calibration of the array itself, to reduce measurements of baseline, delay offsets, and determine pointing models during specific sessions. During standard observing sessions reference pointing and focusing measurements will have to be reduced, with fast loop-back of results to the observing process; the phase fluctuations on the phase calibrators must be evaluated, with a feedback to both the real time process and the scheduler. Calibration will be applied on-line and maps/data-cubes will be produced according to data processing parameters input by the observer. Single-dish observing sessions will also be reduced on-line. The pipeline must be able to reduce on line the quasi totality of the data, which is expected to be produced at an average rate of 6 MB/s, with a peak rate of 60 MB/s for some observing sessions.

For most projects the data pipeline will produce results in a form suitable for quality evaluation, and astronomical processing, hopefully leading to fast publication. Uncalibrated uv data will be archived together with the calibrations curves and the resulting images. The archive should enable fast access to the observing parameters and full reprocessing of the data set with improved processing algorithms.

A general requirement is that the various parts of the system should be developed in a highly consistent way, from the very beginning of design; they may however be installed progressively, provided the critical elements are implemented first.

PIPELINE DATA PROCESSING

ALMA's goal is provide a data pipeline that will build up sky spectral images of a large majority of observing projects. The data pipeline will include the following functionalities:

Near-real-time Data Reduction

In order to maintain the array in a fully operational state, several data reductions have to be performed, some on them quasi on-line. For instance the system should monitor the atmosphere, to compute system temperatures and convert correlator data into temperature data; calibrate the water vapor radiometric phase correction process; determine pointing offsets from the reference pointing measurements; determine focus offsets; compute the amplitude and phase r.m.s. fluctuations due to atmospheric seeing, using repeated observations of point sources at a low frequency (phase calibrators). Results of these measurements will be used to improve in quasi real time the observing procedure (reacting on varying weather conditions) and provide quantitative quality data essential for an efficient short-term dynamic scheduling of the array.

Quick Look

We plan to provide at short intervals raw (i.e. not deconvolved) science images in order to provide to humans a feedback on the actual quality of the data being acquired. This is intended for operators and staff astronomers on the operations site, but also for science investigators, by read-only web access.

Science Data Processing

Whenever actually requested by the science investigators, the pipeline will perform a full reduction of the science data obtained so far. This will be done at the end of observations of a given project, but also at predefined break points, where the observer may want to evaluate the need to continue the observations or adjust the observing parameters, while staying within the scope of the approved project.

Science operations performed will include final calibration of the data sets, calculation of images, inclusion of data obtained in previous sessions (e.g. other array configurations and single-dish observations), image deconvolution.

ARCHIVING

The above specified data rate is not set by the actual processing power of the correlator, but estimated from the needs (in integration time, bandwidth and frequency resolution) of typical science projects, and the feasibility and cost of post-processing equipment. At an average date of 6MB/s we expect to have to archive approximately 180 Terabytes a year. These numbers include both visibilities and image pixels. These data rates are expected to be raised during the operational lifetime of ALMA. At first, data sets will contain visibilities both corrected and uncorrected for atmospheric phase fluctuations, as measured in real time using water vapor radiometric receivers (it is felt safer to keep both sets until calibration strategies are well reliable).

ALMA will not only archive visibility data, but also images produced by the automatic pipeline, as well as complete information about the conditions under which the observations were performed (including monitor data) and how the data were reduced.

The archive should include efficient search functionalities to select and retrieve science data, and to enable archival research when the proprietary data rights of the observers have expired. It should be possible, if enough data processing resources are available, to request reprocessing of archived visibility data with improved calibration and reduction algorithms. The archiving system will be built with the concern that it is easily used by global data mining systems, foreseen to be implemented in the coming years (Virtual Observatories).

OFF-LINE DATA PROCESSING

A specific document [2] has been prepared to outline the science requirements of off-line data analysis, as it has always been foreseen that ALMA will be able to re-use existing radio astronomy data analysis packages.

It is essential the all standard observing modes of ALMA are well covered: both single dish and interferometric observations, as well as accurate reduction of mosaics, with inclusion of short and zero spacing data obtained in single-dish mode, as well as other required ALMA processing. For all the standard observing modes user-friendly interfaces will be needed as ALMA must be open to astronomers of various experience; no detailed knowledge of aperture synthesis should be required for ALMA standard mode users.

STATUS

While the software needed to control the ALMA hardware is well into the development stage, having to support the installation and evaluation of prototype antennas on the VLA site, software for data processing is only now entering the design stages, as the construction phase for ALMA is planned to start in mid 2002.

It is foreseen that the data reduction (both pipeline and off-line) will make a large use of the AIPS++ imaging package, developed by the AIPS++ consortium led by NRAO. A detailed auditing and benchmarking process has been started to assess to which degree this software package is indeed suitable for ALMA needs, and identify areas where further development is necessary (e.g., atmospheric calibration).

References

[1] Lucas, R., et al., ALMA Software Memo 11, *ALMA Science Software Requirements and Use Cases*, http://www.alma.nrao.edu/development/computing/docs/memos/computreviewed.html

[2] Myers, S., et al., ALMA Software Memo 18, *ALMA Offline Data Processing Requirements*, http://www.alma.nrao.edu/development/computing/docs/memos/computreviewed.html