

This paper was accidentally omitted from the printed version, an oversight we deeply regret. Eds.

Dynamic Scheduling in ALMA

Allen Farris, Sohaila Roberts

National Radio Astronomy Observatory

Abstract. The scheduling subsystem of the ALMA software system is described. Since weather will play such an important role in science observations with ALMA, the telescope will operate primarily in a dynamic scheduling mode. Current environmental conditions together with the current state of the telescope itself will be used to algorithmically determine the best scheduling block to execute at any time. This technique implements a micro-scheduler; at each moment in time it answers the question: What is the best thing to do now? The fundamental concepts used to solve this problem are presented as well as the software architecture of this subsystem.

1. The Role of Scheduling in ALMA

The effect of weather conditions on millimeter wave astronomy is well known. Considerable study has been made of environmental conditions at the ALMA site. ALMA Memo 471 contains a study of site properties and stringency. This report includes data on opacity, phase noise, wind and temperature. This and other reports show considerable diurnal variations in these quantities. It is clear that in order to make the best use of ALMA the observing schedule must be changed to reflect the current system status and environmental conditions. It is also clear that these environmental conditions are not fully predictable, although there may be certain times of the year in which some variations do follow patterns that are somewhat predictable. Under these conditions, ALMA has made the decision to implement a dynamic scheduling policy.

ALMA observing projects are represented by a recursive tree structure whose leaves are scheduling blocks. The maximum time associated with a scheduling block may vary considerably but a typical time is thirty minutes. The primary purpose of the scheduling subsystem in ALMA is to manage the execution of approved observing projects. Scheduling is managed in terms of these scheduling blocks, which are regarded as atomic units; they are not normally interrupted. Scheduling blocks may be executed more than once in order to achieve their scientific goals, but, if they are, they begin execution from their beginning.

The sole concern of the ALMA dynamic scheduler is to select the best scheduling block to execute at a particular time, using all available data on

current conditions. Constraints such as long term maintenance periods, special testing times, and VLBI observations are taken into account by producing scheduling blocks representing these events that must be accommodated at fixed times. The ALMA telescope will operate in three fundamental modes: dynamic, interactive and manual. The default and preferred mode is dynamic. Scheduling also supports the concept of sub-arrays that operate independently of one another. The entire life-cycle of an observing project is supported from beginning to end, with support for PI notification, breakpoints, and initiating the science data reduction pipeline. Finally, scheduling also supports a rich simulation mode which is used for short and long term planning as well as tuning scheduling parameters.

2. Architecture

One of the major design goals for the scheduling subsystem is to achieve a flexible architecture that will allow the system to grow with time and experience. There are two distinct but related major functions of the scheduling subsystem. The first is that of selecting work to be done by the ALMA telescope. The second is that of monitoring when certain events in the life of an observing project are completed and taking appropriate action based on those completions. In addition, it is clear that the view the scheduling subsystem must have of the ALMA telescope is that of one or more sub-arrays, with an operational scheduler allocated to each sub-array. These considerations suggest a run-time picture of multiple tasks: a master scheduler, an observing projects manager, and one or more schedulers associated with sub-arrays. Consequently, the scheduling subsystem contains four distinct packages: the Master Scheduler, Project Manager, Scheduler, and Simulator.

There is one Master Scheduler object in the system. It is responsible for starting and stopping the subsystem, creating and configuring the basic objects, maintaining a queue of scheduling blocks to be executed, and dividing the ALMA array into sub-arrays.

The Project Manager object maintains a queue of ready projects that are either in various stages of completion or waiting to be started. Its primary purpose is to monitor events in the lifetime of those projects and take appropriate action in response to those events, including initiating the science data reduction pipeline and monitoring its completion. It also polls the archive for new projects and responses to breakpoints within a project.

A Scheduler object contains its own unique subset of scheduling blocks, made available to it by the Master Scheduler. It also contains the definition of the sub-array on which it is operating. In its dynamic scheduling mode the scheduler uses the dynamic scheduling algorithm to determine the best scheduling block to execute at any point in time. The scheduler initiates the execution of a scheduling block by calling the control subsystem. In the interactive mode, the schedulers actions are under the control of the PI. The PI makes the selection of which scheduling block is to be executed. In all cases a scheduler is created by and operates under the control of the Master Scheduler.

The Scheduling subsystem, running in the simulation mode, is a standalone tool, running off-line and separate from the operational ALMA system. It can

be used in a variety of ways. It can be used to tune parameters associated with the dynamic scheduling algorithm in order to implement effective telescope utilization over long periods. It can be used to develop and test new optimization strategies. It can be used for planning purposes: to identify optimal times for maintenance schedules and testing, for example. The simulator can be configured to run with either actual weather data from the past or with a model of environmental conditions.

3. The Problem

There are two distinct approaches to scheduling problems, which we will call a macro-scheduler and a micro-scheduler. The macro-scheduler takes a global, long-term perspective. It schedules observing projects, maintenance and testing periods over a long period of time, such as weeks or months. It implements telescope operational policies. This is the kind of problem that traditional approaches to scheduling, such as the Spike system, have considered. On the other hand, the micro-scheduler has a far more myopic perspective. Its range of vision is minutes or hours. It is solely concerned with selecting the next scheduling block for execution. These two approaches are obviously related and share much of the same information, but they are distinct nevertheless.

The detailed characteristics of scheduling blocks, their relationships to observing projects, and a high-level knowledge of the characteristics of the telescope, including its constituent antennas, constitute the main body of information that is shared between the macro and micro scheduling approaches. The micro-scheduler must operate in a near real-time mode; it must produce its answer within seconds. From a certain perspective it has many of the characteristics of a decision problem: What should I do next? The micro-scheduler also has access to a body of information that the macro-scheduler does not take into account, viz., changing environmental conditions (temperature, humidity, etc.), and feedback from the actual performance of the telescope (pointing calibrations, etc.). On the other hand, there is no need for the macro-scheduler to operate in a near real-time mode. It may take hours to run, for example, as it tries a wide range of combinatorial possibilities, a traditional approach to structuring such software.

These macro and micro approaches to scheduling are markedly different, because they address different problems and concerns. The long-term macro approach is suitable for telescopes with stable or predictable environmental conditions, or whose scientific goals are not sensitive to variations in those environmental conditions. On the other hand, if the scientific goals are very sensitive to changing environmental conditions that are not predictable, then the macro approach is of little use.

4. The Approach

The dynamic scheduler within ALMA is a micro-scheduler; it operates on a queue of scheduling blocks that it possesses exclusively. No other scheduler has these scheduling blocks under consideration, a fact insured when the Master Scheduler creates the scheduler. The selection process proceeds in four stages.

First, all relevant data concerning the current environmental conditions and the state of the telescope system are gathered, including results of recent calibrations.

Second, a list of candidate scheduling blocks is selected. Members of this list satisfy the following conditions. Targets must be optimally visible throughout the duration of the scheduling block. One must be able to execute the entire scheduling block during the time available. The required number of antennas must be available throughout the scheduling block.

Third, for each scheduling block in this candidate list, we compute a number that answers the following question. Under current conditions, what is the probability that this scheduling block will achieve its scientific goals? Note that this number does not involve comparing one scheduling block with another; it is purely a factual question about the properties of a scheduling block at a particular point in time. This computation takes current environmental conditions and recent calibration accuracies into account, as well as the calibration and setup requirements of the scheduling block. The various factors employed in making this determination are given weights and parameterized into a formula that can be easily tuned.

Fourth, a number is computed for each scheduling block that ranks the scheduling blocks under consideration in order of desirability. This number is computed using such factors as: scientific priority, setup time, resources consumed, cost of having to start a new project, number of scheduling blocks remaining in this project, length of time since executing a block of this project, and stringency. Each factor has a value and a weight, which may be positive or negative. Again, these factors are parameterized into a formula that can be easily tuned.

A final score is computed for each scheduling block in the list by means of a simple formula, for example, by taking the product of its ranking and its probability of success. The top candidates, in ranked order, are submitted for consideration. Unless overridden by the telescope operator, the highest ranked scheduling block is executed.

The actual scheduling algorithm employed in selecting scheduling blocks is referenced by a Scheduling Policy object that is stored in the ALMA archive by ALMA administration. This object includes a name, version, and how the probability of success, ranking, and final score are calculated. It also contains each factor used in the calculations, their weights, and description. The Executive subsystem specifies the scheduling policy used by dynamic scheduling.

References

- The ALMA Software System Schwarz, J. et al. 2004 this volume, 643
- ALMA Memo 471, Site Properties and Stringency, Report of an ASAC Subcommittee, written September 2002, published as memo: 16 July 2003.
- Portable Astronomical Scheduling Tools, Glenn E. Miller, Ashim Bose, in ASP Conf. Ser., Vol. 101, ADASS V, ed. G. H. Jacoby & J. Barnes (San Francisco: ASP). For additional information on Spike, see http://www.stsci.edu/resources/software_hardware/spike