MMA Memo 167: High Level Computing Information Flow for the MMA

M.A. Holdaway
National Radio Astronomy Observatory
949 N. Cherry Ave.
Tucson, AZ 85721-0655
email: mholdawa@nrao.edu

April 16, 1997

Abstract

We map out a straw man concept of the information flow in the computing system for the MMA. We track the information from the proposal stage through refereeing, time allocation, choosing a project from the queue, micro-scheduling the project, flexible queue observing, automated pipeline imaging of preliminary images, automated pipeline imaging of the final images, and archiving. If each of these steps can be accomplished fully electronically, we can design a highly integrated system which is capable of meeting the MMA computing demands which have been laid out by the MMA Computing Working Group (MMA Memo 164).

1 Organizing Principles

The MMA computing system needs to greatly exceed the capabilities of any radio astronomical facility yet built. In order to optimize the scientific throughput of the instrument, we envision flexible scheduling (scheduling a project which can meet its scientific goals in the current weather conditions), flexible micro-scheduling (changing the calibration strategy, integration time, and possibly postponing the project as the weather conditions change during the observation), and automated imaging of the data with the use of an imaging pipeline. These goals sound very challenging, especially in the computing environments of existing telescopes. In order to meet these goals, we must design a computing system from start to finish with these goals in mind. Two guiding principles help us in this design:

- We require that at all steps the information must be in the computer and never primarily as a hard copy.
- We seek whenever possible to automate procedures which can be described algorithmically.

In spite of these principles, some human input, guidance, and oversight is required. The main purpose of the high level information flow we describe here is to show where we need information from humans, where we need information from machines or computer programs, what information will come out of the system, and what algorithmic tools will need to be developed in the future. We merely hint at the types of information and algorithmic tools which will be required, leaving a more detailed look at the problem for a later time.

2 Highest Level Information Flow

At the highest level, the information flow of the MMA is simple: the proposer asks to observe a source, the time allocation committee (with the referees' help) decides how much time to give the project and may possibly modify the scope of the project, the successful project source is observed by the telescope, the telescope produces data, the data are calibrated and imaged, and the images are sent back to the astronomer. Variations of this general scheme involve some amount of input from the astronomer during the observations and the calibration and imaging stages. We intend to design a system which allows for astronomer involvement ranging between the extremes:

- the astronomer sets the specifications for the observations in the proposal and after several months gets images in the mail.
- the astronomer requires control at all points of the process: in project queuing, in micro-scheduling of the project, and in calibration and imaging. (If too successful, this astronomer may be hired by NRAO to upgrade the algorithmic software for automated observing and imaging.)

3 Proposal Handling and Time Allocation

This section deals with the flow of information from the preparation and submission of the proposal to the acceptance of the proposal into the observing queue, as illustrated in Figure 4. (In these diagrams, people or groups of people are represented by ellipses, files or databases are represented by rectangles, and algorithms or machines which generate or eat information are represented by flattened hexagonal shapes.) We describe the steps in the proposal and time allocation process:

1. The proposer generates the observing proposal (OP, which also stands for observation parameters). Current proposals often aim to get most of the information which is required to perform the actual observations, such as source position, observing frequency, correlator setup, and required noise level, but are not in a useful format. If the observation parameters of the proposal are in the correct machine readable format, we can literally give the proposal to the telescope and ask it to observe. The general strategy for the proposals will be for scientific requirements, such as noise level and dynamic range,

rather than observing time required. In addition, environmental requirements, such as maximum acceptable rms residual phase error level or opacity, could be specified. Some tools will assist the proposer in generating the proposal:

- A proposal generating tool (think of a forms interface on a web page for starters) will ensure that the required information is provided in the correct format.
- A proposal viewing tool will convert the native proposal information into a convenient viewing format such as postscript or HTML.
- A time estimation tool, complete with MMA sensitivities and site characteristic statistics, will assist in estimating the time required to meet the desired scientific goals (ie, noise level).

2. The proposer electronically sends the proposal to the NRAO Director's office. NRAO may choose to accept paper proposals. If so, NRAO would need to employ someone to insert the proposal into the appropriate computer readable form.

3. NRAO sets up a database of observing proposals (Set of OP's) resident on an NRAO computer. This database will have limited access.

4. The director's office informs the referees the Set of OP's is complete.

5. Referees view the OP's over the internet.

6. Referees generate comments and suggested changes to the OP's. For example, in addition to comments about the general merit of the observing proposal, a referee might indicate that the desired noise level is unrealistic or not necessary. The general form of the comments will mirror that of the OP, though most items will contain no information unless the referee explicitly thinks a change in the observation parameters is required. A software tool will help the referees input their comments and will automatically input them into the limited access Set of OP's database.

7. When the referees' comments and suggested modifications are in place, the Time Allocation Committee (TAC) reviews all of the proposals.

8. The TAC makes its own comments and modifications to the observation parameters, and....

9. ...assigns each project a queue priority.

10. When the TAC has completed its job, the Observing Queue database is generated from the Set of OP's database. The modifications to the observation parameters decided upon by the TAC are applied to the observer's original observation parameters, and the OP's are ranked by priority in the queue. There will be a software tool which converts the Observing Queue database into an easily viewed form which can be publicly accessed via the internet. Some information may not be available for public viewing, such as the

referee's comments or the text of the scientific justification. The Observing Queue is for use both within and outside NRAO.

11. When the Observing Queue and its easily viewed form are complete, the TAC runs a program that goes through the Set of OP's and sends the appropriate referee comments and the OP's priority to each proposer. The proposer is also instructed to look at the newly generated Observing Queue and is advised about when the OP is likely to be observed.

12. The proposer checks the Observing Queue for herself.

13. The proposer decides that the observation parameters she originally proposed are not quite correct, or that she is unhappy with some modification implemented by the TAC. The proposer can modify her OP in the observing queue, via an NRAO staff scientist who verifies that the change will not significantly alter the estimated observing time. Such changes could occur right up to, or even in the middle of, observations.

4 Projection Scheduling, Observing, and Automated Imaging

At this point, the Observing Queue is all set for dozens or even hundreds of observations. Each OP in the Queue has a set of complete specifications which enable the NRAO to prepare a micro-schedule (ie, the detailed minute by minute schedule). This section goes through the stages of picking an appropriate project from the queue, scheduling, observing, and automated imaging, which will feed back into the queuing and the scheduling. This process is illustrated in Figure 4

1. The physical environment of the MMA will affect observations in many ways, and some conditions will not permit successful completion of certain very demanding observations, but will permit other less demanding OP's without difficulty. In principle, we should be able to determine if a given set of environmental conditions would allow an OP to be observed successfully. To aid in this determination, we will need a series of environmental monitors, such as

    - one or more tippers measuring opacity at various frequencies.
    - one or more phase stability monitors measuring phase fluctuations at various sites about the array.
    - one or more weather stations measuring, among other things, the wind speed and the wind gustiness, which will be used to estimate the pointing errors the antennas would experience.
    - one or more sets of thermal probes on strategic structural members of the antennas to estimate the thermal contribution to the pointing errors and the surface errors.
    - some monitor of the basic health of the array.

The information from these environmental monitors would affect the selection of an OP from the Observing Queue, the scheduling of that OP, and the termination of that OP upon changing environmental conditions.

2. An automated picking algorithm looks at the Observing Queue and selects the highest priority currently observable OP which does not have a hold on it (the observer may observe a small part of the project and then put a hold on that OP until she has verified the observations are proceeding as expected) which can likely be successfully completed with the current environmental conditions.

3. The scheduler uses the output from the environmental monitors and the selected OP to generate the detailed telescope instructions that those familiar with the VLA might call an "observe file". The environmental monitors will affect details such as how often the phase calibrator is observed, or if pointing calibration is performed.

4. The detailed telescope instructions are passed on to a telescope control computer. This stage is much more complicated than shown, but these functions are already operating on the VLA and many other radio astronomical instruments, so we put them into a black box.

5. Astronomical and monitor data gets filled directly onto disk, summarized and placed in a global summary observational database, and archived onto an archiving medium.

6. The imaging pipeline software reads the data from disk and calibrates and images some subset of the data to produce preliminary images. Imaging parameters (ie, weighting scheme, cell size, deconvolution algorithm) are specified in the OP. The data subset (ie, what channels, resolution) and how often the subset is imaged are also specified in the OP, or else sensible default values are used.

7. The preliminary images are evaluated, and the results of the evaluation are fed back into the scheduler and to the proposer. Under some circumstances indicated by the OP, the project will be put "on hold" after a certain condition is met. When an OP is on hold, it is sent back to the queue in an inactive state, and the proposer is informed of the OP's status.

8. The proposer needs to interact with the queue to reactivate the OP.

9. At points specified in the OP (or by default, when the OP has been completed), the imaging pipeline will produce the "final" images using imaging parameters specified in the OP. The "final" images are archived and sent to the proposer. In many cases these "final" images will indeed be limited by thermal noise, and no further data processing will be required. In more demanding cases, the images will require further processing before analysis can proceed. We expect that as the MMA matures, the fraction of projects for which the pipeline produces acceptable final images will increase.

10. Operations marked with the small triangle could be confirmed by a staff scientist to ensure that the system is running smoothly. Human override of the automated system is also possible at this points, and the observer could, in principle, perform these operations herself.

Some kinds of observations, such as VLBI, do not lend themselves to queued observations. We therefore must also maintain the possibility of fixed time observations. An observer who is observing with nonstandard or modes which are not fully supported may also wish to use a fixed time observation to make observer interaction over the internet more convenient.

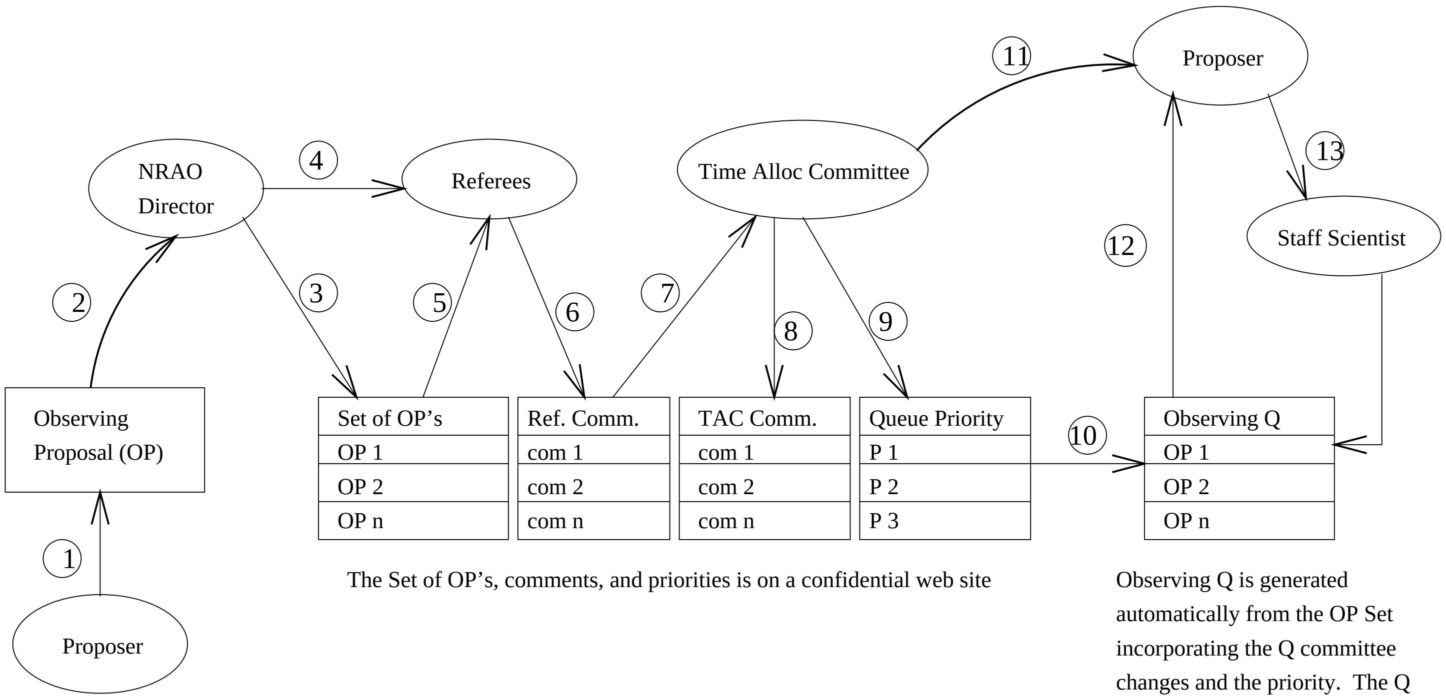
There are a number of pieces of software of varying degrees of complexity which must be written to make this system work. In addition to writing the required software, a lot of work must go into understanding the site and the instrument and how well the observations proceed under various conditions. And finally, the TAC will need to pay close attention to how projects are queued. There may be peculiarities of the logic of the picker which are not immediately obvious. Either the picking algorithm or the priorities assigned by the TAC may need to be modified.

References

Scott, Steve, Emerson, D., Fisher, R., Holdaway. M., Knapp, J., Mundy, L., Tilanus, R., Wright, M., 1996, MMA Memo 164: "MMA Computing Working Group Report".

Figure 1: Initial Proposal Handling and

Generation of the Observing Queue



The Set of OP's, comments, and priorities is on a confidential web site

Observing Q is generated automatically from the OP Set incorporating the Q committee changes and the priority. The Q can be viewed via a public web site.

Figure 2: Observing and Pipeline Imaging