Designing an ngVLA Dynamic Scheduler
ngVLA Computing Memo #6

P. Brandt

September 13, 2021

Abstract

The Next-Generation Very Large Array (ngVLA) is a development project of the National Radio Astronomy Observatory, consisting of 244 antennas spread throughout New Mexico, Arizona, and Texas, with additional stations to provide baselines of up to 5,500 miles (8,000 km). One of the major operational challenges in the ngVLA project will be scheduling observations on these antennas, with the overall goal of maximizing and optimizing the on-sky observing time with every antenna. This memo aims to provide background information on scheduling inputs, look at what other observatories have done with regards to scheduling, and lay out rough requirements for an ngVLA dynamic scheduling system.

1 Introduction

A scheduling system, in the context of an observatory, is responsible for providing a list of the most ideal observation(s) to run in a particular span of time, based on a number of factors discussed later in this memo. Generally the output of a scheduler is, obviously, a schedule. However, it can also be used as a tool to perform simulations, to ensure that frequency and right ascension coverage are distributed fairly across all projects, and to ensure that as many projects as possible are completed across the span of an observing cycle (typically 3-6 months).

2 What is a Scheduler?

Observatory scheduling was first done by expert astronomers by hand, guided by experience and the weather forecast. Now, software performs a similar task with algorithms that attempt to parameterize the knowledge of those expert astronomers. The software-generated schedules are typically passed to human reviewers before being deployed to the telescope. However, it is possible that scheduling could be entirely automatic, with review tools available but not necessary for a schedule's deployment.

A computerized scheduler operates by taking a unit of work (in this case observable science) and slotting it into a time where it can be accomplished. That is, the scheduler has control over a single variable: time. ALMA Memo 282[11] calls this a "level 1" scheduler. More advanced schedulers could adjust the parameters of the units of work in order to accomplish the tasks more efficiently in a given span of time or to better use the available time more efficiently. Memo 282 calls this more sophisticated type of scheduling "level 2".

Relating this to radio astronomy, the level 1 case is what happens with most dynamic schedulers today: for each time slot, a list of potential observations is filtered by a list of constraints (e.g. source visibility, hardware availability) and then scored and sorted via suitability factors (e.g. weather, RFI, project ranking). The scripts that represent the slice of a project that will be run is immutable to the scheduling system. By contrast, a level 2 scheduler would act more like an expert scientist, picking ideal calibrators or changing observing parameters to better fit the current or predicted conditions.

Whether a level 2 scheduler would be feasible, useful, or worth the extra investment in research and development is outside of the scope of this memo; it is mentioned only as a possible avenue to investigate.

2.1 User Insight

A scheduling system should not be a black box. It is important that users understand how the results are produced when seeing the output of a schedule, so that the results can be trusted and viewed as fair and equitable.

A recurring theme in discussions with scheduling staff is a "human factor"[10] of scheduling systems[1], stemming from a perceived need to intervene into automated schedulers. When building the scheduler, sufficient thought should be put into how user tools can provide insight. Providing the proper information to schedule reviewers should allow them to properly tune the scheduler to need less oversight in the future.

2.2 Scheduling Elements

Proposals submitted by scientists, when approved, are transformed into projects. Projects consist of one or more scheduling blocks (SBs), representing a bite-sized piece of the total science, that can each be executed in a time slot. Time slots are not necessarily uniform, and can technically be of any length. In practice, time slices are limited in duration to a few hours to avoid being locked in to a particular observation if the weather degrades[2].

Scheduling blocks specify the parameters needed to schedule the observation, as well as information on how to configure telescope hardware for the intended observation. They can optionally specify calibrators, and the cadence at which they should be observed in tandem with the primary source of interest.

2.3 Scheduling Inputs

Creating a schedule is a multi-objective optimization problem. Defining the optimality of a particular schedule involves selecting how heavily to prioritize each objective. This can be thought of as a point on a Pareto front; the "best" schedule may not represent an absolute optimization with respect to any single parameter or set of parameters, but some trade-off between them.

[1] Mentioned in several private communications with ALMA, GBO, and VLA scheduling staff.

[2] Some telescopes support preempting observations, but even with the relative agility that provides, it still represents a reduction in overall efficiency due to switching overhead and the potential need to re-observe the canceled session.

2.3.1 Weather

The term weather is an umbrella which encompasses a range of atmospheric conditions (e.g. precipitable water vapor, wind, atmospheric stability, frequency suitability) in relation to a given observation. What constitutes good weather varies depending on observing parameters, so there is no one-size-fits-all. Higher frequencies tend to be more sensitive to weather effects (such as antenna surface deformation due to heating, wind, and atmospheric stability) than lower frequencies, meaning there are smaller windows where their constraints can be met.

There is also a distinction between evaluating current or very short term (1-2 hours) conditions versus predicting the weather. For example, the Green Bank Observatory has a system which does weather prediction on timescales of 24 to 48 hours, while the Very Large Array's system is mostly concerned with current conditions to schedule the next observation. The ngVLA will face unique challenges not only in geographically distributed current weather evaluation, but also predicting the weather across the array for future observations.

2.3.2 RFI

Radio frequency interference is increasingly becoming a factor that must be accounted for in telescope operations. Techniques such as RFI excision are well outside the scope of this memo and the scheduling system, but some types of RFI can be easily avoided by properly scheduling observations around them. One example of this strategic scheduling is avoiding areas of the sky where known transmitting satellites are passing through. Another may be avoiding known periodic transmitters that are mainly active during certain times of the day.

Approaching RFI as an input to the scheduling system necessarily involves gathering a database of known time-dependent transmitters in order to be able to avoid them. Satellite constellations for providing internet are becoming increasingly common and prolific, and tracking and avoiding thousands of such satellites may become standard for radio observatories in the future.

2.3.3 Source Position

In order to be observed, the source must be visible to all of the antennas involved. This means that the source must not be below the horizon, as well as not occluded by the sun, moon, or other major solar system objects. For the longest baselines of ngVLA, this may mean some sources will have much shorter visibility windows.

In addition, there are typically minimum and maximum elevations that the antennas are allowed to observe at. The minimum is set so that an antenna's line of sight is not looking through a longer path of atmosphere, and the maximum is set because it is not possible to track something that passes directly overhead with alt-az antennas.

A scheduler must take into account where the source will be located in the sky throughout the period of an observation when determining its suitability.

Efforts are made to ensure that there is equity in observations in different areas of the sky. The galactic center tends to be heavily oversubscribed, so in practice this means ensuring that the remaining observations at the end of a cycle are roughly equal between all right ascensions.

2.3.4 Frequency

The frequency range of interest to an observer determines how several other parameters will be evaluated, especially the weather. Different frequencies can have very different weather constraints, including sensitivity to precipitable water vapor, wind, or atmospheric phase stability. In general, higher frequencies are more difficult to schedule.

A similar approach to equity as with RA can be applied in frequency coverage. Frequency availability due to weather creates an inherent imbalance towards more low-frequency observing time. In cases where oversubscription of high frequency time is allowed, frequency equity "equalizes disappointment" by making sure the remaining hours at the end of a cycle are roughly equal across frequency.

2.3.5 Project Completion

Many observer proposals require more time than can be allocated in a single observing session. Thus proposals (or projects) are split into manageable chunks, called sessions or scheduling blocks. In order to get useful scientific output, all of a project's sessions must be observed[3]. Thus, a weighting factor is added that ensures as many projects are completed as possible in each observing cycle.

2.3.6 Science Rank

When projects are approved by the observatory, they are given a rank or grade. The ranking could be as simple as A, B, or C, or a nonlinear weighting. Higher ranked projects are preferred over lower, and lower ranked projects are used to fill times where no highly ranked projects can be observed. For example, all remaining 'A' projects are high frequency, but the weather is not suitable for high frequency observing.

2.3.7 Observing Cadence

An observation can be filtered by the time since the last observation within the same project, in cases where periodic data is collected such as pulsar timing. This could be specified as a minimum interval (e.g. no more than once per week), a periodicity (e.g. at least 3 times per month), or both.

2.3.8 Stringency

Stringency is not a standalone measure, but rather a number which takes into account several factors in order to summarize how difficult an observation is to schedule. The specifics of how stringency is calculated is somewhat specific to each telescope, as weather plays a major role. An observation which needs "unicorn" conditions will be scheduled over others that have less demanding requirements.

Fixed-date observations such as co-observing with other facilities can be imagined as having an infinite stringency score, such that there is only one slot where they can be placed. Algorithmically, fixed sessions are usually placed first and then dynamic sessions are scheduled around them.

[3] Or in some cases like ALMA, the required sensitivity must be reached. Achieving the desired sensitivity can take more or less time than anticipated when creating sessions from a project, depending on the real world conditions when each is executed.

2.3.9 Hardware Availability

Generally when projects are split into SBs, some approximation of antenna resources is decided upon. A minimum number of antennas, receiver frequency bands, sub-band definitions, correlator resources, and so on are specified as part of the SB. If the required hardware resources are not available, it is a disqualifying factor for an SB to be scheduled.

2.4 Further Responsibilities

Currently, tagging antennas as requiring maintenance (as well as creating work orders) is a manual process. Passing predictive maintenance information between a maintenance management system and a scheduler would allow it to automate array management with respect to taking antennas out of service and bringing them back online for science observations. This would increase planning efficiency because the scheduler could have advance information about when antennas would leave or enter schedulable arrays.

While not nominally under the purview of a telescope scheduling system, there will be significant logistical challenges to deal with when it comes to dispatching maintenance teams to geographically distributed antennas. This is essentially a traveling salesman problem, a classic NP-hard problem in computer science. Scheduling maintenance crews has somewhat different requirements and constraints than scheduling observations, but could potentially benefit from algorithms or techniques used to schedule antennas. Whatever system is used to schedule maintenance crews, it could communicate with the observation scheduling system in the same way as the computerized maintenance management system (CMMS).

3 Existing Scheduling Systems

In this section, we will examine what other observatories use for scheduling observations.

3.1 Dynamic Scheduling System (GBO)

The DSS is Green Bank Observatory's solution for telescope scheduling.

GBO is somewhat unique because observers are responsible for running their own sessions on the Green Bank Telescope. Other facilities run scripts (typically standardized modes) on behalf of proposers. This makes the DSS largely a people scheduling system, though it must still deal with all of the standard parameters previously discussed.

The requirements for the DSS are summarized below. A complete enumeration can be found in [3].

The DSS requirements are somewhat unique, since observations are run directly by PIs (or a designee). This necessarily means that, while still taking into the usual scheduling inputs, the DSS is heavily weighting human availability in its algorithm due to the need to inform observers of their time allocation at least 24 hours in advance. As a result, observers are prioritized over absolute observing efficiency; if the weather turns out "too good" for the scheduled observation, it is still made as scheduled.

The DSS breaks sessions into three types: fixed, windowed, and unconstrained. Fixed sessions occur at a specific time and cannot be moved, for example coordinated observations with external radar signals or VLBI. Windowed sessions must be scheduled within a specific span of time, and are often used for things like pulsar timing observations that should be run periodically (e.g. every two weeks). Observers are allowed to limit the possible windows to a certain number so that they are not constantly on call for several days at a time. Unconstrained sessions can be scheduled at any point in time which is not a fixed session, or marked as unacceptable by the observer.

Stringency[4] is one of the primary ranking factors for the scheduling algorithm. Plotting stringency versus frequency shows the most difficult to schedule areas of the spectrum. Essentially, the algorithm favors the most difficult to schedule sessions first. "The smaller the number of remaining opportunities to schedule a session and the larger the session stringency, the more urgent it is to schedule that session, even if it is not currently the "best" session in the pool."[4]

Over-subscription in certain right ascensions is handled by weighting such that, at the end of an observing cycle, the remaining unobserved hours for each RA is roughly equal. Essentially, oversubscribed RAs are weighted heavier towards being scheduled until the over-subscription is reduced to be roughly equal to other RAs. The same weighting is applied to frequency.

Hard limits are in place to prevent observations from being scheduled or carried out in non-ideal or insufficiently ideal conditions. Additional factors that affect the score of an observation:

- Weighting factors that can increase the score of an observation:
  - whether an observer is on site
  - how much of a project is already observed
  - grade of the science
  - whether a project is part of a thesis
- Exclusion factors can prevent a project from being scheduled:
  - flag indicating that the observation is within 15 minutes of source transit
  - flag excluding daytime hours, usually for RFI reasons
  - LST range exclusion flag
  - observer availability flag

3.2 Dynamic Scheduling Algorithm (ALMA)

The DSA is the scheduling system for the Atacama Large Millimeter Array.

The requirements referenced here are from a truncated list discussed in [5]. The full requirements can be found in [8].

The DSA explicitly supports three primary modes, dynamic, interactive, and simulation. Manual mode operations are left to the telescope control system. (2.1-R2, 2.1-R3, 2.1-R4). Simulation mode is used for analyzing the algorithm's performance and predicting project completion. In interactive mode, the next scheduling block to run is selected by an astronomer. When dynamically scheduling, the scheduler chooses and executes the highest priority observation on the current array. Array management is done manually by telescope operators.

The requirements include support for multiple subarrays, but they take care to note that only a single "main" subarray will run at a given time. Secondary subarrays which run filler projects must be able to release their antennas on short notice in case the main subarray finishes and they are needed for a main research program (2.1-R11).

ALMA scheduling involves many of the factors discussed previously. However, scheduling differs somewhat because of the basic guarantee provided for PI data; rather than simply allocating

[4] "The reciprocal of the fraction of time that the following limits for a transit observation are all satisfied: efficiency, tracking error, and atmospheric stability."[4]

time on the telescope, astronomers are promised a certain sensitivity. For the scheduling system, this means tracking the status of data as it is processed to ensure the proper limits are reached (SNR, resolution) as well as predicting the feasibility of calibration in current conditions (4.0-R3). Another unique aspect of ALMA scheduling is enforcing the balance guidelines for dividing time between the supporting institutions.

Due to power and thermal constraints, only three front-end bands may be powered at once. While a requirement exists for the scheduler to be in control of band power management (4.0-R10), in practice power management is performed by operators (with the advisement of the DSA and pending observations). Future work aims to add this functionality.

The DSA is also used to calculate when antenna reconfigurations should take place, based on the availability of personnel, transporters, and the anticipated weather.

3.3 Observation Scheduling Tool (VLA)

The OST is the scheduling system for the Very Large Array. A full listing of initial requirements can be found in [1]. It is important to note that these requirements were later added to or modified in subsequent planning documents and JIRA tickets, but they can still provide some insight into the initial design.

The objective of the OST is to produce the highest scientific output per unit time. In order to accomplish this, it uses a number of criteria to evaluate SBs: science rank, LST, suitability, contributions to the project's sensitivity, current and future weather, RFI, stringency, and project status.

The OST is required to be able to run off-line with historic or simulated inputs, a feature which has been implemented in the other systems discussed as well. Simulation capabilities are important for verification, planning, and refining the algorithm or heuristic weights.

There is a requirement for support of subarrays, but no elaboration is made in regards to how many or how they will be managed with respect to one another. In practice, multiple science subarrays are seldom used on the VLA.

Like the other schedulers discussed so far, the OST also allows for manual operation and schedule modification. In practice, the scheduler is run shortly before a running SB ends in order to select the next SB to execute. Manual intervention is required for other aspects of the scheduler operation, including gathering feedback on proposed daily schedules and updating antenna positions when they are moved between array configurations.

3.4 CARMENES

CARMENES is an instrument installed at the Calar Alto Observatory. In order to optimize their use of telescope time, a genetic algorithm was used to schedule observations in short, medium, and long terms.[6]

The CARMENES scheduler uses a genetic algorithm, with each observation represented by a gene. Each bit of the gene indicates whether it is observed on a given day or not. The algorithm mutates the genes randomly, and grades the result by whether they represent a valid schedule (for instance, whether the number of hours in a given night are exceeded, or if the weather is suitable for those that have been scheduled).

4 Designing an ngVLA Scheduling System

4.1 Requirements

When designing a scheduling system for the ngVLA, all of the usual factors previously discussed should be accounted for. This is captured in two system level requirements:

SYS2227: The automatic observation scheduling system shall account for the system status, current and expected weather, project priority and percent complete, expected RFI, hour angle and frequency equity, source position limits, stringency of scientific observation requirements, and cadence (for recurring observations), when automatically scheduling observations.

SYS2228: The automatic scheduling system shall prioritize scheduling based on (1) scientific ranking priority, (2) band availability, (3) subarray extent, and (4) project completion percentage.[9]

Scheduling is expected to be fully automated at present, as dictated by SYS2302.

Some factors will need to be handled somewhat differently than they have been in the past, such as evaluating and predicting weather at geographically dispersed antenna sites to determine the feasibility of an array. In addition, the scheduler will be responsible for stopping observations when weather deteriorates, and rescheduling previously stopped observations as laid out in CAL0309.[7]

One major change in observatory operation that will affect scheduling is that calibration will be treated as an observatory function, rather than being tied to each observation. That is, to the extent possible, calibration information will be shared among observations. This should result in a dramatic reduction of calibration overhead, and is captured as a requirement in SYS1069.

Supporting subarray scheduling will be an important feature for the ngVLA, and is captured in SYS2217.

While not currently codified into a requirement, the system requirements note that atmospheric calibration may be done with a separate subarray in order to reduce calibration overhead when frequent source switching is normally required. This would require a number of antennas located close to those which are in the main science subarray, increasing the demand on array resources to schedule certain observations. In essence, this model would reduce time needed to perform an observation in exchange for increased antenna usage. Work is in progress to evaluate another option using water vapor radiometers (WVR) for the ngVLA, which could reduce the cadence at which source switching would need to be done, and possibly eliminate the plan to perform this calibration with a separate subarray.

4.2 Automation

Functions like band power management (if necessary), controlling the presence of antennas undergoing maintenance in the schedulable pool, and array creation could be handled automatically by the scheduling system (or delegated to a high-level executor in the monitor and control system). It is certainly possible to continue having humans performing these tasks, but in order to keep costs low and improve efficiency it is desirable to at least examine the feasibility of automation wherever possible.

4.3 User Interaction

It is imperative that scientists, operators, and other users have an understanding of the scheduling system in order to ensure trust in the resulting schedules. For existing systems that use straightforward weighting algorithms, this is a case of documentation and making sure it is visible to various audiences. In the event that ngVLA requires more sophisticated algorithms (for example machine learning, expert systems, or multi-variable optimization) it will be vital to build tools alongside the system that can provide accurate but succinct views into what the scheduler is doing.

4.4 External Systems

The scheduling system will require several data sources to be accessible in semi real-time: weather, M&C array and correlator statuses, SB availability/status, RFI, and data processing and storage capacity. Interfacing protocols are TBD but where possible we should use interoperable standards, as described in the ngVLA Computing Memo #2.

Despite this section's title, the ngVLA software components will need to be designed to work as a system rather than a collection of disparate parts. The scheduling system will rely on input from a variety of upstream tools to receive prepared observations, and affect or be affected by other systems such as the monitor and control system, predictive and reactive maintenance systems, and data collection and reduction. Taking a broader view of the system when considering such critical components is necessary to achieve an efficient ngVLA scheduler.

4.5 Evaluating Results

An important stage of scheduling is grading the output for quality, efficiency, cost-effectiveness, or other metrics as decided by the project. With current systems at NRAO this is done by visual inspection or manually tweaking the algorithmic weights used to score each observation. However, in the event that the ngVLA scheduler is afforded more degrees of freedom (e.g. the ability to change observing parameters), it could become an intractable problem to have people attempting to decide the value of two or more schedules in relation to each other.

In fact, this may be the case regardless with the addition of multiple operational subarrays. For example, given the choice between several projects utilizing subarrays versus one or a few projects utilizing larger subarrays, which is better? There are an infinite number of contrived examples which could break any naive model of what is "best".

At minimum, some effort should be put into producing an output scoring system for schedules so that they can be more easily summarized, with adjustable weightings for each metric to allow users or stakeholders to tweak what the telescope is being optimized to do. Such a system would also open the door to being able to automatically pick the optimal schedule.

5 Future Work

More investigation into several areas of scheduling is necessary. These areas include:

- The feasibility and usefulness of a "level 2" scheduler (see section 2).
- Coordinated current and future weather evaluation across a geographically distributed array.
- Determining the calibration strategy; i.e. whether WVR data can be used in lieu of either extra antennas or source switching.
- Avoiding RFI through scheduling, and reacting to sudden RFI during observations.

- Applicable techniques for generating schedules. Colome et al[2] provides a more in depth survey of algorithms, and analyzes how these algorithms performed at a number of astronomical facilities. However, they conclude that the best approach depends heavily on the requirements of the observatory.

6 Conclusion

In this memo we have briefly summarized the driving factors that will play a key role in defining and designing a scheduling system for the ngVLA. This information should assist in determining detailed requirements, and later the architecture of the scheduling software. There still remains research to be done into what techniques are best suited to solving ngVLA's scheduling constraints, but the team is seeking input from experts in the field of operational research.

References

[1] W. Brisken et al. EVLA e2e Science Software Requirements. Technical Report EVLA-SW-001, NRAO, 2005.

[2] Josep Colome et al. Research on schedulers for astronomical observatories. In Alison B. Peck, Robert L. Seaman, and Fernando Comeron, editors, *Observatory Operations: Strategies, Processes, and Systems IV*, volume 8448, pages 469 – 486. International Society for Optics and Photonics, SPIE, 2012.

[3] J. J. Condon. Scientific Requirements for Dynamic Scheduling. 2007.

[4] J. J. Condon and D. Balser. Dynamic Scheduling Algorithms, Metrics, and Simulations. 2015.

[5] A. Farris, S. Lucero, and D. Clarke. Scheduling Subsystem Design Document. Technical Report COMP-70.55.00.00-003-E-DSN, 2009.

[6] A. Garcia-Piquer et al. Efficient scheduling of astronomical observations - Application to the CARMENES radial-velocity survey. *A&A*, 604:A87, 2017.

[7] C. Hales. Calibration Requirements. Technical Report 020.22.00.00.00-0001-REQ, NRAO, 2020.

[8] R. Lucas et al. ALMA Software Science Requirements and Use Cases. Technical Report ALMA-70.10.00.00-002-I-SPE, 2003.

[9] R. Selina et al. System Requirements. Technical Report 020.10.15.10.00-0003-REQ, NRAO, 2020.

[10] I. Toledo et al. Astronomical Scheduling in the Era of Big Observatories. In *Astronomical Data Analysis Software and Systems XXVII*, volume 522 of *Astronomical Society of the Pacific Conference Series*, page 219, April 2020.

[11] M. C. H. Wright. ALMA Memo 282: Dynamic Scheduling. 1999.