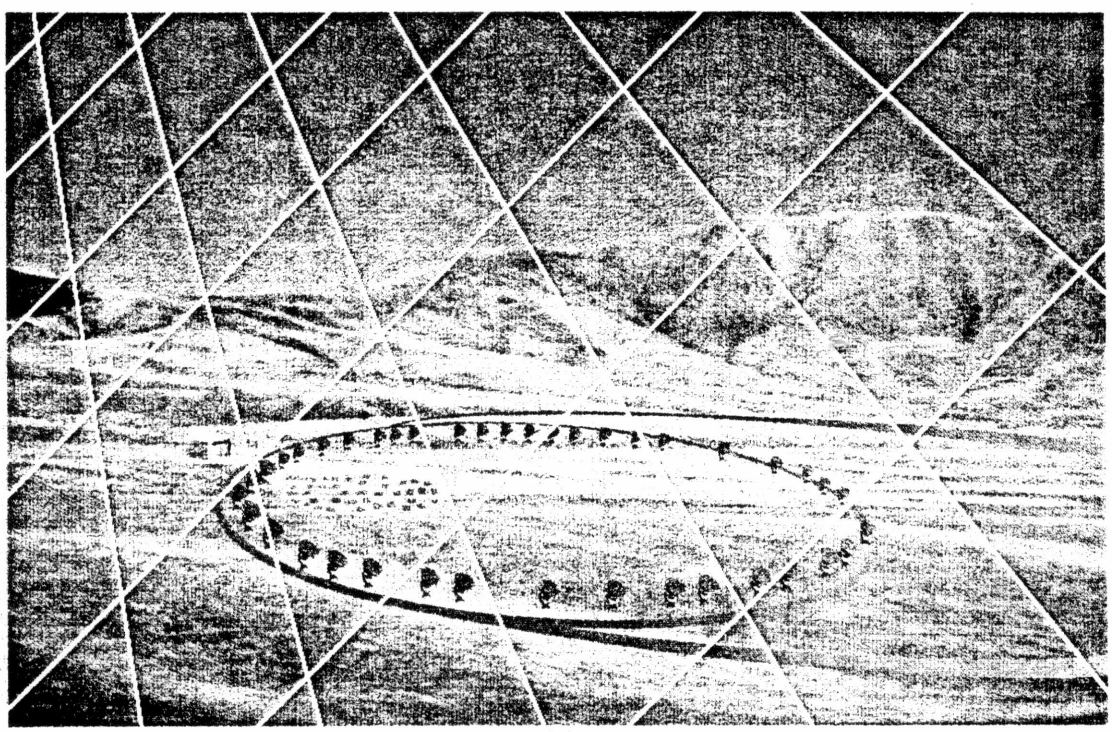
Millimeter Array Project Book

Version 1.0

July 1998



NATIONAL RADIO ASTRONOMY OBSERVATORY
A facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Millimeter Array

Project Book

Version 1.0 – July 1998

National Radio Astronomy Observatory
National Science Foundation

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

MMA PROJECT BOOK

Version 1.0, July 20 1998

Please note (July 1998) that this is the first draft of the MMA Project Book. Major revisions to most chapters can be expected after the first MMA PDR, at the end of July 1998, with regular revisions after that.

DTE

Table of Contents

	1. Algorithm development, Imaging Requirements Data Analysis, Data Flow and Data Storage issues are, for the time being, covered in Chapter 12.	-	7/13/98
14	**Site Characterization**	-	7/15/98
15	**Array Configuration**	-	7/15/98
16	**Site Development**	PDF	7/13/98
17	**Construction, Integration and Interim Operations**	PDF	7/13/98
18	**Post-construction Operations**	PDF	7/13/98
19	**Schedule and Timeline**	PDF	7/20/98

PDF files:

To print a chapter, click on the PDF box if it is available. This should invoke your Browser's PDF Reader. The PDF file will be of higher printed quality than the HTML file.

Please be aware that when printing HTML documents from Netscape using a PC running Windows 95, the printed format is affected by your screen resolution. If your screen format is set to less than 1024 horizontal pixels, some of the figures in the HTML documents will be truncated at the printer. Screen resolution in Windows 95 can be changed by choosing "Settings" from the "Start" menu, then "Control Panel" and "Display". Then choose the "Settings" tab and increase your screen resolution. 1024 by 768 or greater will be adequate; 640 by 480 is not. Other platforms and browsers may have similar problems.

Robert Brown
Last modified July 8 1998

MMA PROJECT BOOK: INTRODUCTION

Design of the Millimeter Array (MMA) is an ambitious undertaking. Uniquely, it is designed as a complete imaging instrument with the capability to image astronomical objects on angular scales from milliarcseconds to many degrees. It will measure fluctuations in the microwave background that are isotropically distributed all over the sky and it will image the kinematics of streams of gas smaller than 500 parsecs in length that are flowing onto the massive blackholes of quasars 3000 Megaparsecs away from us. MMA imaging on this tremendous range in angular scales, from degrees to less than 50 micro-arcseconds, brings with it both enormous capabilities for astrophysical research and enormous challenges for the technical design of the instrument. It is the purpose of the MMA Project Book to provide a comprehensive account of the scientific requirements of the MMA and the approaches being taken in design, fabrication and integration of the instrument to reach those goals.

The Millimeter Array will be constructed in two stages: a 3-year Design and Development (D&D) phase will be followed by six years of construction. The Design and Development effort has as its goal completion of designs for the array instrumentation including prototypes of the highest risk instruments or modules and decisions made among competing designs where that is necessary. The D&D work has a specified task, staff, timescale and budget. The D&D effort is very well defined. In contrast, the construction planning for the MMA involves open questions, answers for which are among the products or deliverables of the D&D work. The operations model for the MMA is even less well developed because it hinges not only on a clear understanding of what must be operated and maintained--the product of the D&D work--but also on a realistic assessment of what operational aspects of the MMA can and cannot be done in the vicinity of the MMA site, elsewhere in Chile, or which must be done at a U.S. base. The answers to these questions will become clear as experience is gained in Chile. Nevertheless, this MMA Project Book includes a description of the planning for the entire project, D&D, construction and operations. Where there are outstanding issues or several options still available these topics are included in the Project Book and noted as to be determined. The MMA Project Book is meant to be a living document growing both in breadth so as to encompass the entire project and in depth as design, construction and operations decisions are made.

For each task the MMA Project Book summarizes requirements and the technical approach being taken to meet those requirements. Where one task interacts with another either in design or integration, the interface requirements are specified and the integration assumptions being made are noted--or noted as missing and in need of definition.

The Project Book is meant to be the fundamental reference manual for what is and is not planned in the project. It is written by those people working in the project and its principal readership is meant to be other people in the project. As decisions are made or options are adopted the Project Book will evolve; that will be done using an audit trail of additions and changes appended to each chapter. The Project Book is of value only if it is kept up-to-date. For this reason the Project Book will be kept on line: the version that is printed from the web will always be the latest version.

MMA SCIENCE REQUIREMENTS

Frazer Owen,
Last revised July 8 1998

I.

Introduction

The concept of the MMA was first discussed in 1981 and written about in 1982 (MMA memo 1). From this beginning the ideas have evolved through numerous advisory committees and large scientific workshops. The goal has always been to produce a cutting edge instrument which will make great contributions in almost all areas of astrophysical research. As with the VLA, there are topics which now appear to be the most important and interesting areas for major progress with the array. These include cosmology, galaxy, star and planet formation, and chemistry/creation of the building blocks of life. However, with the VLA, other areas of astrophysics than the ones the instrument was sold on, have turned out to dominate the research; the same may be expected of the MMA. Thus the scientific requirements of the MMA have always been to build a very versatile instrument which, like the VLA, will be an important tool for decades to come regardless of the direction research takes us.

The most recent summary of the science goals of the US community are summarized in five reports from the November 1995 Tucson Science Workshop. These are divided into

1. Cosmology and Extragalactic
2. Star Formation and Stellar Evolution
3. Galactic Molecular Clouds and Astrochemistry
4. Solar System
5. Sun and Stellar

While these different areas emphasize different capabilities, they all require imaging over a range of resolutions and frequencies. This means that we must build an instrument which can image objects from milliarcsec (microarcsec for VLBI) to scales at least of 10's of minutes. This implies that an important feature of the array is that it must be reconfigurable in size , like the VLA, to match the problem and that it must be able to image objects larger than the antenna primary beam. Since the scales of many of the most interesting objects are much larger than the primary beam (which decreases in size linearly with wavelength), much of the observing may not use the interferometer mode at all but may use the array as a large collection of single dishes. This implies that sensitive, stable total power mode is an important requirement. However, current wisdom is that the interferometer mode will come closer to the theoretical sensitivity than will be possible for total power. Thus it appears to be attractive to do as many experiments as possible in interferometer mode. This, in turn, argues for smaller antennas. However, the desire for as simple an instrument as possible and as large a collecting area as possible for sensitive work on small sources conflicts with this desire at some level which leaves us a goal of obtaining the best compromise solution for a given amoount of money. This is the primary design conflict for the array.

In section II, I summarize the general instrumental requirements based on the science requirements as detailed in the five science reports listed above. In section III, I will outline some of the implications for the array to be able to reach these goals.

II.

General Requirements

1. Frequency Coverage

   The instrument needs to cover all the available frequency windows between about 30 and 900 GHz. These requirements are summarized in the MMA Frequency Bands whitepaper . This requires an outstanding site as discussed in the Site Selection Document.

2. Spectral Line and Continuum

   The MMA must operate as both a sensitive spectral line and continuum array. This implies using the widest continuum practical from the point of view of the IF and the correlator which appears now to be 8 GHz and a flexible correlator as described in the MMA Correlator Whitepaper .

3. Sensitivity

   The array must maximize both point source and surface brightness sensitivity. For antennas with the same overall properties, this requires maximizing the different quantities, nD^2 (for point source sensitivity), nD (for surface brightness sensitivity in a sparsely filled array) and $n^{(1/2)}$ (for a tightly packed array or total power mode). See MMA Memo #177.

4. High Resolution

   Given the expected brightness and size of sources considered in the science documents, this implies baselines of at least 3km and 10km as a design goal. This requirement demands the MMA be adequately phase stable both internally and in the presence of atmospheric phase flucuations. This will be discussed further in section 3.

5. Large Source Imaging

   On the other end of the size scale, this implies imaging objects both close to and bigger than the primary beam. This requirement has several implications. First, 3)+ 4) + 5) require that the array be reconfigurable into configurations optimized for the resolution and sensitivity required by each experiment. Second, the array must be able to make large mosaicked images (multiple pointings) to image regions on the order or larger than the primary beam. Third, the array must have a sensitive, stable total power system so that spatial frequencies smaller than are available in interferometer mode can be measured. Since the primary beams at the highest frequencies for antennas on the order of 10m are < 10 arcsec, modes 2 and 3 should be very common, perhaps the vast majority of all observations with the array.

6. High Fidelity Imaging

   Especially in the modes discussed in 5), a significant fraction of the experiments and some the most important require high fidelity imaging. That is the signal-to-noise is high enough and source complex enough that errors in pointing and calibration can degrade the scientific usefulness of the experiment (Cornwell memo/Rupen memo). Such problems require 1) excellent pointing, 2) high quality amplitude and 3) phase calibration which will be discussed in section 3.

7. Polarization

   Both linear and circular polarization of lines and continuum emission is a significant part of the MMA science program. At centimeter and longer wavelengths interferometers produce linear polarization by correlating the opposite circular polarizations from different antennas, that is R with L and L with R. However, it appears technically difficult to do this at millimeter wavelengths across the broad bands we want with the MMA. Thus it seems best to observe in the more natural linear polarization with the MMA. This means we crosscorrelate to calculate the V stokes parameter; we get I, Q and U from linear combinations of the two linear correlations. This requires both linears be present all the time and that either their relative gains remain very stable and/or we have the necessary internal calibration signals to measure their changes. (see Cotton, 1998).

8. Solar observations

   Requirements for observing the sun are discussed in the Sun and Stars science document and by Bastian et al, 1998.

9. VLBI

   The highest resolution with the MMA will be obtained from VLBI observations using the MMA as a single element. The requirements for this are discussed by Claussen and Ulvestad, 1998.

10. Pulsar/high speed

    Pulsar observations will require a gating mode with the correlator as well as a sum port like the VLA which can be attached to specialized external recording equipment. This latter capability would also be available for any high speed phenomena which may be discovered in the future.

III.

Implications

Several capabilities are implied by these requirements.

1. Phase Stability

As the frequency increases, the electrical path length through the atmosphere and the electronics must be more and more stable to produce high quality images.

a) Internal phase stability

The electronic systems must be stable enough so they they do not degrade the imaging relative to residual atmospheric path length fluctuations. This implies phase stability < 10 degrees at 900 GHz. This needs to be accomplished either by building a very phase stable system or by providing adequate internal calibration signals to calibrate out any remaining path length changes. If some or all of the MMA's frequencies operate in linear polarization, then the relative phase stability between the two receivers at each polarization needs to be high. A reasonable goal is 0.1% stability in the instrumental linearly polarized term. This requires about 0.1 degree stability between the two receivers.

b) Atmospheric phase stability

Atmospheric path length changes will be measured and corrected primarily using one of two techniques: fast switching or radiometric correction. In some special cases there may be other modes of phase calibration as well.

i) Fast Switching

In this mode the antennas are rapidly cycled between a nearby calibrator and the program source before the atmosphere can change significantly. This mode seems most certain to work but will trade calibration time for integration time on the source and thus reduces the effective instrumental sensitivity. It requires that the antennas be able to change pointing position very rapidly and stop quickly. Since most of the bright sources are available at 90 GHz or lower frequencies, it also requires that the array be capable of changing frequency bands during the slew between sources and that the internal phase differences between the observing and calibration bands be stable. See MMA Memo #139.

ii) Radiometric Phase Correction

Since water vapor in the atmosphere produces both opacity and changes in electrical path length, measurements of the sky brightness at millimeter wavelengths can be combined with model atmospheres and/or local calibration to estimate the phase changes caused by the atmosphere. However, these measurements require very stable total power radiometry. This technique sohould have much less observing overhead associated with it than the fast switching technique. However, it may be limited by the instrumental stability and the calibration/atmospheric modelling we are able to achieve. These problems are discussed in Carilli et al, 1998.

iii) Other techniques

In some cases there will be sources in the same primary beam which can be used as phase calibrators, either at the same frequency or at some different frequency than the observations. In the case where the frequency is different, for example in the case of a maser, the system must be cabable either of simultaneous observations at both frequencies or if different receivers are required of changing bands back and forth in a few seconds.

2. Amplitude stability

a) Atmospheric

Atmospheric opacity corrections need to be made often enough to prevent these errors from limiting each experiment. For a detection experiment or under good conditions, this may only require occasional opacity measurements for the entire array. For time variability or under marginal conditions, a more aggressive approach is probably necessary. Thus some form of total power monitoring off-source to allow such corrections to be made on an antennas-based approach is necessary.

b) Internal

i) Interferometer mode

Either the internal stability of the instrument needs to be good enough so that gain changes never limit an experiment or an internal amplitude calibration system (possibly a coherent signal) needs to be available to measure such gain changes.

ii) Total Power Mode

Perhaps the most demanding observations which will be made with the MMA will be the total power observations. For sources larger than the primary beam at any frequency the array will depend on this mode for its imaging information. The fluctuations in sky brightness and the instabilities in the receivers and other parts of the system will provide a great challenge to the instrument. The sky fluctuations will need to be dealt with by rapid on-the-fly mapping, that is scanning the antennas over the region of interest fast enough that the sky brightness has not had time to change very much or by rapid beam-switching. Either all the components of the array will need to be extremely stable or some sort of internal calibration will be necessary.

3. Integration times

The fastest integration times may be driven as much by total power and fast mosaicking as by high time resolution. See MMA Memo #192.

4. Contingency Scheduling

At millimeter and submillimeter wavelengths, the atmospheric conditions limit most experiments. That is why we are going to such as an extreme site as Llano de Chajnantor. To make optimum use of the site, we need to schedule experiments when atmospheric conditions allow. Experiments may be limited by atmospheric transparency, phase stability, wind, and possibly other conditions. To match the science to the changing conditions, we will need to schedule the array in real time, rather than blocking out time periods as is traditional on NRAO instruments now.

5. Dataflow

For such a complex instrument, the astronomer will need a complex, yet friendly, computer system to help the user to deal with the extremely, flexible MMA.

i) Since the array will be contingency scheduled, the process by which the astronomer submits and schedules a proposal will be different. This may only mean that somewhat flexible schedules, perhaps with more total time scheduled than the experiment will use, are submitted. It could also mean that the observatory computers and/or staff actually prepare the schedule. Also the telescope time allocation process will need to deal with assigning or approving the allowable weather conditions under which the experiment will be carried out.

ii) Experiments will probably more often be split up to allow them to be scheduled as close to transit as possible to minimize the air mass and opacity. Thus the astronomer will be able to get useful feedback on how an experiment is going during the time it is on the telescope. This needs to include diagnsotics and logs, as well as images. In some cases, these images might be the final output of the experiment. In other cases, the imaging may be too complex for quick turnaround without the astronomer's interaction. In these cases, the images may serve as quality control to allow the astronomer to adjust the parameters of the observation.

    This implies a more sophisticated communications/software system than has been use in astronomy to date.

iii) Almost real-time feedback to the staff on-site is also necessary to allow them to evaluate if an experiment has been successful. Weather and other data on the site as well as analysis of the quality of the data need to be part of this feedback to allow the scheduler (whether human, machine or some combination) to decide what to do next.

iv) A final data product which would be a combination of images and uv data also needs to be defined.

v) Off-line software is needed to allow the astronomer to take advantage of the necessariliy complicated imaging algorithms (often using large mosaics and combining interferometer and total power observations) as well as inspection and analysis of the results.

Calibration Issues for the MMA

Mark Holdaway
Last modified July 15 1998

1.0 Introduction

Calibration is the process by which the astronomer converts electronic signals from the telescope into meaningful astronomical data. Calibration is crucial for the MMA. As millimeter and submillimeter wavelength radiation will be adversely affected by the atmosphere and the electronic signal path in a variety of ways, and as the antennas will also be affected by the observing environment, we must understand how we will correct for or remove these various effects. The calibration strategy interacts heavily with the science requirements, the system design, the receivers' functioning, the antenna's physical behaviour, the site conditions (ie, site characterization), the scheduling of the telescope, the real time computing, and the post-processing software which will take most of the burden of implementing the required calibration schemes.

Never before has a radio astronomical instrument been built with such a detailed understanding of the site and its impact on the telescope. With this knowledge in hand, we can optimize the full calibration strategy to produce the maximum scientific output for the MMA.

At this point in time, we are convinced that all the calibrations which are required can be effectively achieved, but we have not yet made all the decisions as to how to achieve these calibrations. Furthermore, many of the options which we lay out for the various calibrations interact with each other, so we have a long way to go yet before we have a coherent picture of all calibration systems and their interdependencies. This chapter tries to express the astronomical requirements for the various calibrations which need to be performed, as well as the hardware and software requirements for the competing methods for performing these calibrations.

2.0 Pointing

2.1 Astronomical Requirement for Pointing

Cornwell, Holdaway, and Uson (1994) show a requirement of 1 arcsec for pointing accuracy, based on the requirements of good mosaicing image quality. Observations at the highest frequencies (850 GHz) will also require this pointing accuracy for even single pointing observations, and high frequency mosaics would often benefit from even better pointing. However, a large fraction of the MMA observations, such as single pointing observations at frequencies up to 500 GHz, mosaic observations at 115 GHz or lower frequency, or low SNR mosaics at millimeter wavelengths will not require this precision pointing spec.

With an increase in D to 10m, the pointing spec should tighten to 0.8 arcsec. Holdaway (1997; Memo 178) performed a more detailed analysis, showing the effects of pointing errors ranging from totally random to totally systematic. While we do not divide up the 0.8 arcsec pointing error specification into various systematic and random terms, we note that the effects of any pointing error budget with various systematic and random terms could be translated to an estimated image quality. We do note here that random errors have less of an effect than systematic errors, but we also assume that systematic errors can be calibrated out.

With only minor exceptions, pointing calibration must be performed prior to astronomical observations or the data are useless. This also means that we cannot generally interpolate pointing solutions backwards in time. This makes pointing calibration all the more crucial.

2.2 What Affects Pointing?

The antenna pointing will be affected by several slowly varying terms such as systematic imperfections of the antenna and the pad, gravitational forces, and thermal loading from the sun. Depending upon the strategy of the astronomical observations, much of these slowly varying effects can be removed by frequent offset pointing observations.

In addition to these slowly varying systematic pointing errors, there will also be highly random pointing errors caused by wind loading and anomalous refraction. Measurements of the wind indicate that a great deal of the power in the wind is often in a constant speed and direction, so it is only the gusts about this mean speed and direction which will result in differential pointing errors between the calibration and the target source (see Holdaway 1996, Memo 159). The refractive pointing will usually not be a severe problem, but will sometimes limit the pointing (Holdaway, 1997, Memo 186; Bulter, 1997, Memo 188). Since the refractive pointing is random on time scales of the antenna crossing time of the atmosphere (ie, 1 s), we mainly need to have statistically many different atmospheric instantiations for each of the five points of a pointing calibration to ensure that we are not applying an erroneous pointing position when we collect data on the target source.

Finally, at some level there will be a limit to the mechanical repeatability of the antenna pointing. At this level, we are left with completely random pointing errors which cannot be calibrated. If these purely random errors are too large, they will spoil the imaging characteristics of the MMA and will not be correctable. If they are small to moderate in magnitude (ie, < 0.5 arcsec), we can tolerate them quite well as these random errors are the least damaging of any pointing errors.

2.3 Pointing Calibration Strategy

Our basic goal is to have systematic pointing errors which can be calibrated and removed completely so we are left with purely random pointing errors which are small enough not to bother us.

2.3.1 Pointing Model

The first step in removing the systematic pointing errors is to determine the systematic imperfections of the antenna and pad and the effects of gravity. Most radio telescopes periodically undergo a pointing routine which samples the sky with pointing measurements on about a hundred astronomical sources taken at night to minimize thermal and wind pointing errors. The MMA will take about 60 minutes to perform 100 pointing calibrations across the sky. The results of these pointing measurements are then used to fit about 10 parameters in a pointing equation which accounts for various physical terms, such as misalignment of optical axes or four fold sag due to the antenna base being supported at four locations.

Some experimenting will go into determining the optimal form for the MMA pointing equation.

The MMA will probably rely heavily upon a lower frequency (e.g. 90 GHz or below) system for determining the pointing model. The wider beam at this low frequency and the high sensitivity and bright astronomical sources will facilitate pointing measurements, even after a reconfiguration. However, precise pointing offsets among the different frequencies will also need to be determined. It is our hope that the blind pointing after application of the pointing model is on the order of 2 or 3 arcsec rms. To achieve the precision pointing specification, frequent (i.e. every 30 minutes) offset pointing calibration will be required, to remove local deviations from the pointing model, or systematic slowly varying effects due to wind and/or thermal gradients.

2.3.2 Offset Interferometric Pointing on Quasars

Holdaway (1996; Memo 159) and Lucas (1997; Memo 189) have both demonstrated the feasibility of performing pointing calibration on quasars close to the target source; adequate SNR can be obtained with sufficient speed. The minimum calibrator flux, and hence the typical minimum distance to a pointing calibrator for pointing calibration, is a function of both the collecting area and the number of elements in the array.

A key question concerning the efficiency of these offset pointing calibrations: what are the differential pointing errors as a function of distance between cal and target sources? This depends upon the direction of wind, the sun angle, etc, and probably needs to be answered experimentally. Also, we must understand the stability of the pointing offsets among the different frequency bands, as the pointing calibration will usually be done at 30 or 90 GHz.

Both the pointing model solution and the offset interferometric pointing will require an extensive, up-to-date catalog of pointing calibrator sources, and the observing schedule program should allow for automation of the choice of a pointing calibrator and the pointing calibration strategy.

2.3.3 Infrared Pointing

It may be possible to perform infrared or optical offset pointing of the MMA antennas. Jack Welch is investigating the possibility of ensuring the infrared axis is aligned with the radio axis. At this point, this technique should be regarded as highly experimental. It doesn't need to work, as the offset interferometric pointing will work well enough. However, if infrared pointing does work, it will increase the overall efficiency of the instrument, and may improve the antennas' pointing. The infrared pointing will be largely immune to refractive pointing. Since our main strategy concerning refractive pointing is to minimize its random effect on the pointing measurements, the infrared pointing is not at a disadvantage because it is not affected by anomalous pointing.

2.3.4 Scheduling and Editing

A particular experiment's demands for precision pointing need to be combined with the current site conditions (phase stability for refractive pointing, wind and wind stability, and solar loading and variability) and qualitative rules of thumb for the success of pointing calibration and pointing stability in various conditions to determine when that experiment should be scheduled. Records of the pointing solutions, phase stability, wind fluctuations, and solar loading can be used to locate times when the pointing solutions are suspect, and the astronomical data during these times can be editted accordingly.

If the pointing solutions show large drifts with time, but all antennas are behaving similarly, the mean array pointing position as a function of time can be corrected in post-processing.

2.3.5 Pointing Self-Calibration

Pointing self-calibration, an unimplemented algorithm, may not work well at all unless there is at least one bright point-like source within the target area. Once the antenna pointing offsets have been determined, it is simple and not too cpu-expensive to apply the mean array offset as a function of time to mosaic or single pointing data and use that in the imaging. However, if there is significant scatter among the antennas' pointing positions at each time, imaging wide field sources considering the correct pointing data may be prohibitively expensive (Holdaway, 1993; Memo 95).

3.0 Primary Beam Calibration

A detailed understanding of the primary beam will be required to image wide field astronomical objects. At low frequencies, pointing errors will tend to limit mosaic image quality, but at high frequencies, surface errors resulting in primary beam distortions will dominate the errors in mosaic image reconstruction (Cornway, Holdawell, and Uson, 1994). Early simulations (Holdaway, 1990; Memo 61) indicated that an understanding of the primary beam down to the 2-3% level is desirable. If this cannot be achieved, mosaic dynamic range will be limited by primary beam uncertainty more than pointing errors.

The desired primary beam information will result naturally from the low frequency holography campaigns. While the high frequency primary beams will not simply scale like the frequency, we can estimate them from a surface error model and the feed placement. However, we will probably want an independent measurement of the beam at several frequencies. At the highest frequencies where we expect the most problems with the primary beams, the primary beams will be hardest to measure due to decreased sensitivity and a lack of appropriate sources.

We may require primary beam models with different levels of complexity to achieve different scientific goals. The simplest primary beam model, which will suffice for low to intermediate dynamic range observations, will be a mean rotationally symmetric primary beam, measured out to the first or second sidelobe. The next level of complexity may be a mean 2-d (ie, non rotationally symmetric) beam. It is concievable that we would someday require the use of different 2-d primary beams or voltage patterns for each antenna, for several different elevation angles, or even for both.

4.0 Baseline Determination

The relative positions of the antennas must be determined accurately so the geometrical delay can be correctly applied to the antenna voltages prior to correlation. Residual delays will result in phase errors which change across the observing band and differential phase errors between two different sources on the sky. For the MMA with a 10 GHz bandwidth, a reasonable limit of 1/3 radian phase difference across the band requires a baseline accuracy of about 1 mm. Requiring the differential phase error between two sources 5 degrees apart on the sky to be on the order of 5 degrees results in a baseline accuracy of about 0.1 mm. Atmospheric phase errors of more than 5 degrees would not severely impact the imaging, as these errors are random in time and among antennas. However, baseline errors will be partially systematic as they will be slowly varying in time, so we need to be more conservative with them than with the atmosphere.

4.1 Baseline Measurement Strategy

The baselines, or delays, may be measured by determining the delay on each baseline for on order of a hundred observations of point sources sampling the entire sky. Individual delays can be fit across the spectrum, as in VLBI. The complete set of delays is used to solve for the three dimensional locations of all antennas relative to a reference antenna. The observing strategy is similar to the pointing model determination, and should take about an hour to complete. Signal to noise is not an issue for 0.1 mm accuracy, and the 1 hour time scale is set more by the minimum time to sample many sources around the sky. Atmospheric phase fluctuations may affect the baseline delays, so ideally the observing conditions should be excellent. In poor conditions, the delays can probably still be determined based on the statistics of many differential measurements, as the atmosphere should tend towards a zero mean in differential measurements.

Of some concern is the time scale over which we can expect the antenna positions to remain fixed to within 0.1 mm. We do not expect any permafrost on the MMA site, which rules out an entire class of soil movements. However, we can probably expect some amount of soil creep, especially after earthquakes. We will gain experience concerning the frequency of baseline calibration once the MMA begins to be operational at Chajnantor.

5.0 Flux Calibration

5.1 Astronomical Requirements for Flux Calibration

Yun et al. have written a white paper on flux calibration which addresses most of the issues mentioned here. The primary scientific requirement for accurate absolute flux calibration is the comparison of astronomical images made at different frequencies. While the current 10% estimated absolute accuracies permit many qualitative conclusions, 1% absolute flux accuracy will really open up new quantitative scientific possibilities. This flux calibration accuracy must apply to both total power and interferometric modes.

5.2 What Affects Flux Calibration?

Changes in the receiver temperature, electronic gain drift, variable atmospheric opacity and emission, variable ground pickup, decorrelation (atmospheric and electronic) and gravitational deformation at high frequencies are some of the parameters affecting accurate flux calibration.

5.3 Flux Calibration Strategy

5.3.1 Instrumental Amplitude Calibration

The currently used ambient load chopper wheel method is accurate to about 10%. Bock et al. (1998) are investigating a two load system located behind the secondary and viewed through a hole in the secondary. This system could theoretically achieve 1% amplitude calibration of a well understood antenna/receiver system, but would rely upon accurate ancillary measurements of the decorrelation and the opacity. BIMA is prototyping this system, so we should find out in the coming years how well it works. An accurate estimate of the atmospheric opacity will be required for accurate flux calibration. At frequencies at which the atmosphere is opaque at the low elevation angles and partially transparent at high elevation angles, it will be possible to solve for the sky temperature and the opacity with a sky tip. However, at frequencies at which the atmospheric transmission is excellent, a sky tip will only give the product of the opacity and the sky temperature, and the temperature must be assumed to calculate the opacity. Currently, atmospheric models are not sufficiently accurate to measure the opacity and sky temperature at a partially opaque frequency and accurately estimate the opacity at another frequency. However, an actual measurement of the sky temperature and water vapor profiles via radiosonde would be beneficial to the accurate determination of the opacity at the more transparent frequencies. Since the sky temperature and water vapor profiles change during the day and with weather systems affecting the site, several radiosonde launches per day would be required. A more cost effective solution would be to float a teathered balloon over the site several times a day. The temperature, pressure, and water vapor information would also be useful for the radiometric phase correction schemes.

5.3.2 A Gain-Based Instrumental Amplitude Calibration

Larry D'Addario points out that tracking Tsys fluctuations may not be the easiest way to get good flux calibration. He points out that with a multi-bit correlator, we can measure the correlated power instead of the correlation coefficient; hence, we need to track the electronic gain from RX to correlator. The electronic gain does not vary with atmospheric opacity, changing ground pickup, etc, and is therefore much easier to track, perhaps by injecting a broad-band signal through the system. Even so, we still need an accurate opacity measurement.

5.3.3 Astronomical Flux Calibration

While an instrumental amplitude calibration accurate to 1% is a desirable goal, it is unclear that we will be able to understand the antennas and the atmosphere well enough to achieve this ambitious goal, so it is important to find astronomical sources which could serve as flux standards. Currently, planets are used as astronomical flux standards for millimeter wavelength observations, but the estimated accuracy of the planets' fluxes is only about 10%, so new flux standards need to be found.

The MMA will have the sensitivity to use much fainter sources as flux standards. At the highest frequencies (650 and 850 GHz windows), some stars will be bright enough to achieve the 1% flux calibration goal in a few minutes. A knowledge of their temperatures from optical data will determine the expected submillimeter flux, but this may be complicated by confusing emission from dust or even time variable gyrosynchrotron emission from sunspots or flares.

As the blackbody spectrum falls off very fast at lower frequencies, millimeter observations will not be able to use stars as astronomical flux standards. Currently, asteroids appear promising. Many are unresolved to many MMA baselines, the bright ones are in the range of 50-1200 mJy, permitting fast detection at 1% accuracy, and they are fairly simple systems. One drawback is their non-uniform emission as they rotate and as they move with respect to the sun, which could be change their flux by several percent. Observational tests are required on these prospective astronomical flux standards as the MMA comes on line, and we can always hope for a less problematic class of sources for an accurate astronomical flux standard. See Yun et al. (1998) for a more detailed discussion of stars and asteroids for flux standards.

5.3.4 Phase Decorrelation

An uncorrected antenna based phase error of 10 degrees rms will result in a 3% decrease in the visibility amplitude due to decorrelation. As the characteristics of the phase noise change, the amount of decorrelation will also change. The primary defense against decorrelation is to try to correct the phase as much as possible. However, when the phase cannot be fully corrected, we can estimate the magnitude of the decorrelation and correct the visibilities. Decorrelation could be estimated from:

- phase calibrator data (fast switching)
- independent phase monitor (atmospheric) PLUS injected LO signal (antenna mechanical & electronic)
- radiometric data (atmospheric) PLUS injected LO signal (antenna mechanical & electronic)

For atmospheric coherence times, see Holdaway 1997 (Memo 169).

5.3.5 Changes in Tau

At millimeter wavelengths, the changes in atmospheric opacity will be very modest, under 1% over 10 minutes about 80% of the time. Since the same amount of water vapor results in much larger opacities in the submillimeter, the opacity fluctuations in the submillimeter over characteristic calibration time scales will be much larger, typically several percent during median stability conditions. Due to the lack of submillimeter calibration sources available for fast switching and the current uncertainty in the transmission models, we will probably need to perform frequent tipping measurements to solve for the opacity at the observed frequency.

5.3.6 Polarization Complications

As mentioned below under polarization calibration, if a linearly polarized calibration source is used to track changes in the amplitude gain or opacity, a telescope with linear feeds will produce parallel hand visibilities which are modulated by the linear polarization. The extra signal varies as a sinusoid of the parallactic angle, so the errors are systematic. This will not be a problem for the astronomical flux calibrators mentioned here, but would be a problem for quasars.

6.0 Phase Calibration

Phase errors limit resolution, limit the dynamic range of images, introduce artifacts, and reduce sensitivity by decorrelation. Without effective phase calibration, the maximum usable MMA baseline would generally be about 300 m. Amplitude errors would limit image dynamic range and skew the flux scale.

6.1 Astronomical Requirements for Phase Calibration

The phase calibration working group report (Woody 1995; Memo 144) considered three cases at 230 GHz: high quality imaging with 8 deg phase errors, median conditions with 19 deg phase errors, and poor imaging with 48 deg phase errors. The phase errors have a budget which includes the atmosphere, the antenna, and the electronics.

6.2 What Affects the Phase?

At millimeter wavelengths, the main atmospheric constituent which causes phase errors is inhomogeneously distributed water vapor. Up to about 300 GHz, atmospheric water vapor is very nearly non-dispersive. Above 300, water vapor can be quite dispersive, especially near the water vapor lines in the atmosphere. Submillimeter wavelength observations will need to account for this dispersion if the phase is being calibrated indirectly (ie, scaled from a lower frequency or determined by scaling the differential water vapor column as determined by water vapor radiometry).

The dry air results in a major contribution to the absolute phase. If there are appreciable temporal or spatial fluctuations in temperature or pressure in the dry air above the array, phase fluctuations will result. Furthermore, the absolute dry air phase depends upon the observing elevation angle and the topographical elevation, which will change from one source to another. It is believed that the dry phase is non-dispersive at millimeter wavelengths.

Any change in the distance between the subreflector and the feed will cause phase errors.

The stability of the LO and other electronics will also influence the phase.

6.3 Phase Calibration Strategies

6.3.1 Fast Switching

If a calibrator is sufficiently close and the telescope is sufficiently fast, fast switching between a calibrator source and a target source can effectively stop the atmospheric, electronic, and antenna phase fluctuations. If fast switching is used as the phase calibration method, it makes minimum requirements on the system sensitivity, the slew speed and settle down time of the antennas, and the online and data taking systems. Fast switching has been studied extensively (MMA Memos 84, 123, 126, 139, 173, 174), and we are fairly confident that it will work for the MMA.

6.3.1.1 Sensitivity Requirements

The basic criteria for fast switching to work is that the phase calibration source needs to be detected with sufficient SNR and the target source be observed for some amount of time within the coherence time and distance of the atmosphere. This translates into a requirement that there be sufficiently many calibrator sources which are sufficiently bright (Memo 123), and a requirement on the sensitivity of the array. In practice, this means that the calibrator source will typically be within a degree of the target source, the calibrator will usually be detected in less than a second, and the entire cycle time will be about 10 s, though the details vary with observing frequency. Spectral line observations will need to use wide bandwidth continuum observations of the calibrator.

With the current sensitivity of the MMA and our understanding of the quasar source counts and their dependence on frequency, we will not always be able to perform fast switching calibration at the target frequency, but often we will get a higher SNR phase solution by observing the calibrator at a low frequency (like 30 or 90 GHz) and scaling the solution up to the target frequency.

6.3.1.2 Scaling the Phase to High Frequency Observations

The falling source counts and sensitivity at high frequency will often require fast switching to observe calibrators at low frequencies and scale the phases up to the observing frequency of the target source.

This requires a much more accurate phase solution at the lower frequency. Since the dry atmosphere and the electronics terms are non-dispersive, this extrapolation basically relies upon the wet differential delay to be non-dispersive as well. In the submillimeter, the wet differential delay is dispersive, which will either limit the effectiveness of fast switching or require more complications in the fast switching observing strategy, such as less frequent multi-frequency calibrator observations to help separate out the non-water vapor phase contributions.

6.3.1.2 Requirements on Antenna Movements

The antenna movement requirement is currently a slew of 1.5 degrees and settle down to 3 arcsec pointing in 1.5 seconds.

6.3.1.3 Requirements on Antenna and Electronics Stability

At the very least, the antenna needs to be mechanically stable to within a small fraction of a wavelength (ie, 5-10 degrees at the target frequency) over a calibration cycle time, even when the antenna is moved by a few degrees on the sky. Similarly, the electronics need to be equally stable over the calibration cycle time. However, if we are to succeed in the submillimeter, the antenna and electronics need to be stable over much longer times, such as the cycle time between the multifrequency observations required to separate the wet and dry phase errors.

6.3.1.4 Requirements on Computing

The on-line system needs to control the antennas gracefully enough to move them quickly without exciting the lowest resonant frequency. Also, the quanta of integration time and scan length need to be sufficiently small so as not torestrict the integration time spent on the target source and calibrator or the time spent between sources. Flexibility at the 100-200 ms level is desirable. Fast switching data can be calibrated with existing software, but some extensions in spatial-temporal interpolation will be useful.

6.3.1.5 Sensitivity Loss from Fast Switching

Fast switching will reduce the sensitivity of observations due to time lost observing the calibrator and moving the antennas, and due to decorrelation from residual phase errors. Both effects can be reduced by observing in the best conditions, which often result in very low residual phase errors at a minimum expense in time lost to the calibration process. However, not all projects can be observed during the best phase conditions. MMA Memo 174 concludes that fast switching will generally result in less than a 20% decrease in sensitivity for the phase conditions at the Chajnantor site.

6.3.1.6 Interaction with Scheduling

During poor phase stability conditions, fast switching won't work at the high frequencies. Also, a given target field may have a dearth of calibrator sources, requiring that the field be observed during better phase conditions than the average field. For reasons like these, dynamical scheduling is absolutely required to optimize the utility of the MMA. We envision one or more phase stability monitors providing real time information to the array control center, and contributing to observing decisions - e.g.:

- what project should run on the telescope?

- do the present conditions permit the current project to continue?
- what is the optimal calibrator for the current project in the current atmospheric conditions and hour angle?

6.3.1.7 Calibrator Survey and Maintenance of a Calibrator Database

The quasars which will form the bulk of the fast switching calibrators will be highly variable at millimeter wavelengths, and a quick survey of a few square degree region about the target source will sometimes be required. The MMA has the sensitivity to perform a blind search for calibration sources in a few minutes. Surveys directed with lower frequency source catalogs will be even faster. Whenever a potential calibrator is observed, the source information will need to go into a comprehensive calibrator database, which can also be used for choosing an appropriate calibrator.

6.3.2 Radiometric Phase Correction

The most promising alternative to fast switching is radiometric phase correction (MMA Memo 209, Radiometric Correction white paper). Radiometric phase correction utilizes the variable emission caused by inhomogeneously distributed atmospheric water vapor to determine the phase fluctuations caused by water vapor. While water vapor is not the only source of phase errors, it is the dominant source of short time scale phase fluctuations. This method has had several early successes, but the correlation between the radiometric fluctuations and the interferometrically measured phase fluctuations changes with time, and there are some times when the method does not work well at all.

The current thinking for radiometric phase correction is that the 183 GHz water vapor line should be exploited. The partial saturation of this line, even in the driest conditions on Chajnantor, initially seemed problematic, but Lay (Memo 209) indicates the unique line shape helps to discriminate between water vapor and errors like spillover, water droplets, temperature fluctuations, height fluctuations, and gain fluctuations. A total of 16 channels each of 500 MHz bandwidth would permit good discrimination between the water vapor and these errors. A cooled system is desirable. When the PWV column is under 4 mm, residual antenna based rms path errors of under 50 microns can be achieved. Larger water vapor columns preclude high frequency observations, so the larger phase errors associated with high opacity conditions will not be critical.

We should build a pair of prototype cooled 183 GHz spectrometer systems to verify Lay's predictions. We also need to decide how to implement the 183 GHz water vapor spectrometer on the MMA: do we use a standalone cooled or uncooled system, a dedicated radiometer in the receiver dewar, or do we simply use the 183 GHz astronomical receiver as a water vapor spectrometer?

6.3.4 Calibration of the Electronic and Antenna Phase with an Injected Signal

Radiometric phase correction will only correct for those phase fluctuations which are caused by water vapor, and will not correct for any phase errors caused by variations in the dry atmospheric delay, mechanical instabilities in the antenna, or instabilities in the electronics. Therefore, radiometric phase correction requires some supporting observations or calibration technique to remove phase errors caused by these other sources.

It should be possible to periodically inject a stable signal, perhaps derived from the LO, into the feed to calibrate the electronic contributions to the phase errors. If the calibration signal is injected from the subreflector, then this calibration system will also track the most important mechanical phase drifts of the antenna. If the calibration signal is derived from the LO, and the LO itself has phase instabilities, they will either cancel or be doubled, depending upon the relative phase of the LO and the injected signal. In fact, by alternating the relative parity of the injected signal and the LO, we can solve for both phase errors in the LO and in the rest of the electronics and the antenna up to the subreflector. So, between a reference signal injected at the subreflector and radiometric phase correction, only fluctuations in the dry atmosphere will be unaccounted for.

The on-line system would need to control the details of the injected signal. Information about the injected signal would need to be recorded with the data, and an option for determining and correcting for the electronic phase errors in real time should exist.

The injected signal calibration scheme is an area of research for the design and development phase of the project and will be developed in coordination with the LO system.

6.3.5 Paired Array Phase Correction

It is possible to use some of the antennas to observe a calibrator and the rest of the antennas to observe the target source. At this time, no special plans are being made for this "paired array" phase calibration technique. Specifically, the array is not being designed in a way that closely pairs antennas to optimize paired array calibration. In the smaller arrays, the configurations will naturally permit paired array calibration.

7.0 Bandpass Calibration

In rough terms, the dynamic range of a single spectral channel which is limited by errors in continuum subtraction caused by bandpass errors will be

$$DR = \frac{S_{line}\sqrt{N}}{S_{cont}\sigma_{BP}} \quad (1)$$

where S_{line} is the strength of the line, N is the number of antennas, S_{cont} is the continuum strength of the target source, and $sigma_{BP}$ is the rms error in the bandpass. For the spectral line observations to be limited by thermal noise and not by bandpass errors, and assuming the bandpass errors are themselves due to thermal noise in the observations of the bandpass calibrator, we have the condition that

$$N\sqrt{t_{cal}}/\sqrt{t_{line}} > S_{cont}/S_{cal} \quad (2)$$

where S_{cal} is the flux of the bandpass calibrator.

7.1 Scientific Requirements

A majority of the spectral line observations made with the MMA will probably have no problem meeting the condition of being limited by thermal noise before they are limited by bandpass calibration.

Observations of weak spectral lines in the host galaxy of a bright flat spectrum quasar will be about as demanding on the bandpass calibration as these experiments currently are for the VLA or the AT. However, observations of spectral lines on planets in the solar system will be extremely demanding.

For example, the continuum brightness temperature of a planet might be 200 K, and the thermal noise for a bandwidth of 20 m/s and a 1 hour observation at 200 GHz with a T_{sys} of 40 K would be about 0.006 K, requiring a lot of time on a very bright source free of spectral emission to provide a good bandpass calibration. Furthermore, a large fraction of the MMA's targets will be galactic, and it will be increasingly difficult to find bandpass calibrators which are not affected by galactic emission or absorption.

7.3 Bandpass Calibration Strategy

We will always have the capability of calibrating the bandpass in the current manner, by observing a strong source for a sufficiently long time. However, injection of a strong noise signal which is flat over the observed frequency range would be an ideal solution to both the planet problem and to the galactic confusing line problem. The injection can be made directly into the feed, or can be part of the ancillary LO system which has been proposed to be injected from the subreflector. The injected noise will need to be much stronger than the 400 K signal at a few percent coupling mentioned in the flux calibration section.

At millimeter and submillimeter wavelengths, the atmosphere will also contribute to the bandpass for wide bandwidth observations, so we must either perform an independent determination of the bandpass astronomically to solve for the atmospheric bandpass component, or we would measure the precipitable water vapor from opacity measurements made at a fiducial frequency and determine the atmospheric contribution to the bandpass through the use of an atmospheric transmission model. Currently, the atmospheric transmission models are probably not good enough for this sort of work, but the MMA would provide enough data for an ad hoc model or to improve the theoretical models. We will be concerned with changes in the atmospheric component of the bandpass on reasonably short time scales and among the different antennas.

There is an implicit specification placed on the system design that the electronic bandpass be either stable or that it vary linearly with time to something like 10000:1 to 100000:1. If the bandpass changes are mainly linear, we can remove them through interpolation if we calibrate often enough.

8.0 Polarization Calibration

Linearly polarized feeds have a wider usable bandwidth than circularly polarized feeds, and the MMA will most likely use linear polarization in order to get complete coverage of all millimeter wavelength atmospheric windows with a reasonable number of receivers. Sault et al. (1991) and Cotton (1998, Memo 208) have both treated the problem of polarization calibration with linear feeds in detail. The main details that we must be concerned with here are that linear polarization leaks into the "parallel hand" visibilities, and that it is not easy to distinguish circular polarization from the instrumental polarization terms.

Because the linear polarization is entangled with the total intensity in the parallel hand visibilities, there are times when all four cross correlations per baseline will need to be performed, which will probably result in halving the bandwidth and cutting the sensitivity by $\sqrt{2}$. We consider several cases which could come up with the MMA to demonstrate when we may need to consider all four cross correlations and when we may use approximations to make use of just the two parallel hand cross correlations:

- Amplitude calibration is performed instrumentally and phase calibration is performed on a quasar (or a combination of radiometric plus a quasar). The quasars will generally be a few percent linearly polarized, but may be as much as 10-20% polarized, and hence Stokes Q and U will influence the parallel hand visibilities. These sources have almost no circular polarization. For a point source, the calibrator's linear polarization will not affect the phase, only the amplitude. If the amplitude calibration is performed instrumentally, as in the scheme of Bock et al., there is no problem with a polarized calibrator and linear feeds. We further consider two subcases:
  - Total intensity imaging with no polarization in the target source. Many millimeter spectral line sources will have little or no linear polarization. Nothing special needs to take place, as the parallel hands will basically contain Stokes I.
  - Total intensity imaging with appreciable linear polarization in the target source. The linear polarization in the target source will corrupt the parallel hand visibilities in a systematic way. However, when the XX and YY visibilities are added together, the linear polarization corruptions cancel out. This is acceptable for low to moderate dynamic range total intensity observations, but may not be sufficient for high dynamic range total intensity observations, as residual gain errors will limit the cancellation of the linear polarization and adding the XX and YY correlations results in a condition in which gain errors no longer close, limiting the use of self-calibration. High dynamic range total intensity imaging of a source with appreciable linear polarization may require full polarization calibration and imaging.
  - Polarization imaging. A bright calibration source must be observed to determine the instrumental polarization leakage or "D" terms. If the calibrator has known (or zero) linear polarization and no circular polarization, the D terms can be determined in a single snapshot. If the calibrator has unknown linear polarization, the calibrator must be observed through sufficient parallactic angle coverage to permit separation of the calibrator and the D terms. Application of the D terms will permit the polarization imaging.
- Amplitude calibration is performed astronomically. If the amplitude calibrator is not polarized, there is no problem. If it is linearly polarized, then the parallel hand visibilities will vary systematically with parallactic angle, the XX and YY visibilities varying in opposite senses. There are several options:
  - For total intensity observations of a target source at low to moderate SNR, the array-wide XX and YY gain ratios can be determined and corrected for.
  - High SNR total intensity observations will require accounting for the different parallactic angles of each antenna, which will result in imperfect cancellation when using the array-wide gain ratios. In this case, the full polarization calibration will need to be performed, even if there is no interest in polarization.

In all cases in which the cross hand visibilities are explicitly used, the X-Y phase offset must be monitored for each antenna. As there is no simple way to determine the X-Y phase offset astronomically, the MMA could inject a tone into the feeds, as the AT does. Cotton (1998) points out that it is difficult to generate a millimeter RF tone, and that injecting an IF tone further downstream in the electronics is simpler, though not as good instrumentally. On the other hand, we could derive an RF signal from the LO and inject it into the feeds for the X-Y phase calibration.

The choice of a flux calibrator may also interact with the polarization calibration. Unresolved asteroids which are not azimuthally symmetric will have some time dependent linear polarization, which will complicate the flux calibration. If stars are used for a flux standard, they may display some circular polarization, which would require that another source be used for the D term calibration.

As stated above, the full polarization calibration requires good coverage in parallactic angle to separate the constant instrumental polarization (D term) signal from the sinusoidally varying astronomical polarization signal. This causes some concern since the MMA is envisioned to be predominantly a near-transit instrument with real time imaging capability. If instrumental polarization calibration is required for many observations, it may be prudent to keep a database of the instrumental polarization solutions at the various frequencies and bandwidths and rely upon that whenever possible. Unlike the VLA, the ATNF compact array shows essentially no time variability in the instrumental polarization (less than 1:10000 over 12 hours, with variations of 0.1% over months). Given the constraints of the MMA, time constant instrumental polarization may be a good design goal for the feeds, but not a strict requirement.

One way around the complication of good parallactic angle coverage is to use sources of known polarization (ie, unpolarized sources). Holdaway, Carilli, and Owen (1992, VLA Scientific Memo 163) have demonstrated that it is possible to solve for the instrumental polarization for a single snapshot, (ie, a single parallactic angle) if the source polarization is known in advance. So, it would be beneficial to MMA observing to identify bright, compact sources with known polarization or no polarization for use as polarization calibrators.

9.0 Special Single Dish Calibration Issues

The MMA differs from any other aperture synthesis array in that, from the outset, the instrument will support no-compromise single-dish observing modes in addition to the more usual interferometric modes. Some of the issues are discussed in **MMA Memo 108** (*"Single Dish Observing and Calibration Modes"*, D.T. Emerson, P.R. Jewell).

Because single-dish observing is in total power, albeit it switched against, for example, blank sky, there are extraordinary demands on instrumental gain stability. In addition, the extra, variable emission from the sky comes in directly, and tends to mask the much weaker (by perhaps 4 orders of magnitude) astronomical emission. This is in contrast to interferometry, which of course by the use of cross-correlation rather than self-correlation, is relatively immune to these factors.

Astronomical sensitivity calibration in single-dish mode has to be on a dish-by-dish basis; calibration sources need to be detectable with adequate signal-to-noise ratio by one single dish of the array. This is again in contrast to interferometric astronomical calibration measurements, in which the large collecting area of the entire array can contribute to the signal-to-noise ratio achievable in calibrating individual dishes of the array.

Polarization calibration of single dish observations has its own problems. At mm-waves, polarization measurements are conventionally made with a "widget" in front of the receiver feed. This "widget" introduces changes in the polarization response of the receiver - for example a rotating grid and screen combination can continuously rotate the incident plane of linear polarization. The astronomical polarization is then detected by synchronous changes in total power intensity through the receiver as the sense of polarization changes.

The MMA may indeed have to provide such "widgets" for each of the antennas. However, the complexity and potential unreliability of such a device could be avoided if it were shown possible to measure polarization reliably, in single-dish mode, by cross-correlation of the signal from orthogonally polarized feeds. Tests of the feasibility of this techniques are planned by early 1999.

9.1 Atmospheric Emission Cancellation

The emission from the atmosphere is much stronger than the emission from most astronomical sources, and, even worse, the atmospheric emission is variable as well. The variable part of the emission is mainly due to inhomogeneously distributed water vapor, which also causes the phase fluctuations. Since we have excellent statistics of the phase stability on the Chajnantor site, we can infer the severity of the variable atmospheric emission at any desired frequency by using a transmission model or FTS measurements.

For an interferometer, the atmospheric emission above two different antennas is not correlated, so it does not affect the visibilities. In total power continuum observations, the variable atmospheric emission is a major problem which requires some sort of switching on the sky. The total power spectral line case is much less demanding, as large atmospheric fluctuations can be tolerated, considering the much smaller channel widths and much higher thermal noise and the possibility of fitting an average baseline to each spectrum. The spectral line data will have secondary effects, such as the bandpass changing in response to the changing atmospheric load. However, the spectral line observations are much easier than the continuum case, so if we can beat the atmosphere for continuum observations, the spectral line observations will be no problem. The detailed treatment of this problem is presented in an upcoming MMA Memo (Holdaway, Lugten, and Freund, 1998).

9.1.1 Beam Switching

Traditionally, beam switching by nutating subreflector has been used to remove the variable atmospheric emission. Our study indicates that most beam switching is non-optimal. For any given observation, we would like to be roughly equally limited by thermal SNR and by the residual variable atmospheric emission. If the noise is dominated by the variable atmospheric emission, we need to switch faster. The faster we switch, the better the atmospheric cancelation, but the lower the duty cycle, so the thermal noise will increase. Furthermore, the distance of the throw also needs to be considered. In general, it is optimal to have the smallest throw which gets completely off source. However, in an unstable atmosphere, multiple short throws are better. Hence, the detailed use of a nutating subreflector needs to be fine tuned to match the atmospheric conditions and the observing frequency. As with fast switching, we hope that the observer does not have to perform the calculations to find the optimal switching strategy; the observer should provide high level guidelines, and the program which performs the micro-scheduling should calculate the optimal switching strategy for the current atmospheric conditions.

John Lugten's nutator design allows for a maximum angular acceleration of 13200 deg/s/s, or 2000 deg/s/s on the sky, and a maximum throw of about 2 * 8.6 arcmin for symmetric beam throwing. With such a nutator, we would like possible beam throws ranging from 0.3 arcmin to 17.2 arcmin. Maximum nutating frequencies of about 40 Hz could yield acceptable duty cycles for very small throws, and larger throws would require lower frequencies such as 5 Hz. The key aspect of this nutator is that it must be flexible, permitting any combination of throws and frequencies which is physically possible. If it is affordable, nutators with higher peak acceleration and larger maximum throws would be desirable, but the stated acceleration and maximum throw are acceptable. The two beams should be as similar as possible to reduce the level of systematic errors in beam switching.

The analysis of the On-The-Fly technique for total power continuum observations indicates that it will be as good or better than beam switching in all situations. However, there is considerable risk involved in the On-The-Fly method. For this reason, it is generally agreed that the prototype antenna needs to have a nutating subreflector.

9.1.2 On-The-Fly

In On-The-Fly (OTF) observing, the antennas scan quickly across a source at constant elevation angle, using the off-source regions on other side of the source region to define the sky emission. Very large sources will need to be pieced together at some SNR expense. The OTF technique promises to be quite effective at removing the atmospheric emission for three reasons:

- each Nyquist sample on the sky is observed for a very short time, so the system noise is large and a larger amount of sky fluctuation noise is tolerable. (The large number of Nyquist samples observed in each scan compensates for this large noise per Nyquist sample.)
- since more time is spent observing the OFF than an individual ON Nyquist sample, the atmosphere is well determined, unlike beam switching where we are differencing two noisy numbers.
- since the OFF's are observed over a range of time, we can remove a second order polynomial trend in the atmospheric emission time series, which greatly reduces the residual sky emission fluctuations.

For sources which are about one beam across, the OTF observing strategy works about as well as beam switching. For larger sources, OTF wins because of the relative increase in the SNR of the atmospheric determination and because multiple throws begin to degrade the beam switching SNR.

Because the entire antenna is moving, many systematic errors which plague beam switching (such as differences in the shapes and gains of the ON and OFF beams) are eliminated. However, it takes much more energy to move the entire antenna, and there is more risk in general with an observing strategy that attempts to move the entire antenna.

9.1.2.1 Controlling Antenna Movements

OTF will work only if we can slew and reaccelerate the antenna quickly without exciting the lowest resonant frequency of the antennas. An initial analysis of this problem has been performed by Holdaway, Lugten, and Freund (1998). Using a Gaussian acceleration profile and an error function velocity profile, they predict the antennas will be able to turn around from one scan direction to the other in about 0.2 s without appreciably exciting the lowest resonant frequency. This acceleration profile is a good one, but probably not an optimal one, so further work could help optimize the profiles for both OTF antenna motion and fast switching antenna motion.

In order not to excite the antenna motions, the acceleration must be very smoothly varying. This will put strong constraints on both the control system and on the servo system.

9.1.2.2 Maximum Velocity and Acceleration

OTF simulations of sources of various sizes indicate that the optimal slew velocity varies linearly with source size. For a maximum interesting source size of 1 deg, a maximum slew rate of about 0.5 deg/s is required. This requires a maximum antenna angular acceleration of about 12 deg/s/s. Since the profile is Gaussian, we do not require this maximum acceleration for very long. These maximum velocities and accelerations are for an antenna with lowest resonant frequency of 6 Hz. An antenna which was less stiff could not utilize such large accelerations and velocities in OTF observing. A stiffer antenna would permit faster turnarounds, requiring larger accelerations and velocities. However, the 6 Hz antenna is effectively beating the atmosphere already, so not much is gained from a stiffer antenna.

9.1.2.3 Reading Out Encoders, Dump Time

OTF requires that we know where the antenna is for each Nyquist beam. At the 0.5 deg/s maximum slew rate, observing at 850 GHz with a half beam size of 0.001 deg will require that we dump the data and know where the antenna is every 2 ms. We don't need to make the antenna go to any precise place at any precise time, we just need to know where the antenna was at a precise time. We may not need to read the encoders every 2 ms; if the antenna position changes smoothly over time scales of 10 ms, we can read out the encoders more coarsely and interpolate. We do not require that the encoders be accurate to within the pointing specification of 0.8 arcsec.

9.1.2.4 Antenna Motions Don't Need to be Synchronized

Since we are only talking about total power OTF here, we need not synchronize all the antennas in their dance across the sky. The antennas could be staggered to permit a more constant utilization of electrical power.

9.1.2.5 1/f Noise

In addition to atmospheric brightness fluctuations, beam switching and OTF will remove a portion of the receivers' 1/f noise. From the optimizations we have performed, we can set specifications on the 1/f noise for each observing frequency. Even though the beam switching is performing the switching faster than OTF, the integration time spent on each ON is often larger than the integration time spent per Nyquist sample of an OTF observation, so OTF and beam switching are similar in their ability to switch out 1/f noise. If these specifications cannot be met, we must reoptimize the OTF observing strategy, which would result in moving more quickly to accomplish faster switching and less time or more white noise per Nyquist sample on the source. This would favor both higher maximum accelerations and a stiffer antenna.

Freq	Beam Size Source		0.5 deg Source	
[GHz]	noise	break ν	noise	break ν
	[Jy]	[Hz]	[Jy]	[Hz]
90	0.047	1.2	0.081	0.34
230	0.088	1.2	0.25	0.29
345	0.14	1.2	0.47	0.29
650	0.33	1.3	1.6	0.34

Table 1: For continuum (8 GHz

bandwidth per polarization) OTF observations, what noise level must the 1/f noise be below, and at what frequency, for 1/f noise to have essentially no effect on OTF observations' sensitivity?

10.0 Solar Observing

Tim Bastian will write a section on special solar observing calibration requirements.

11.0 Editing

Both the on-line and post-processing software should provide for carrying various monitor data through the system and allowing easy editing of astronomical data based on the monitor data, on a per time or per antenna basis.

References

Bock, D., J. Welch, M. Flemming, and D. Thornton, 1998, "Radiometer Calibration at the Cassagrain Secondary Mirror".

Cornwell, Holdaway, and Uson, 1994

Cotton, 1998, MMA Memo 208.

Holdaway 1997 (Memo 169).

Holdaway, Carilli, and Owen (1992, VLA Scientific Memo 163)

Lay, O. 1998. MMA Memo 209.

R.J. Sault, N.E.B. Killeen, M.J. Kesteven, 1991, ''AT polarisation calibration'', ATNF Technical Document Series 39.3015.

Yun et al., 1998, White Paper.

More, more, more

mma@nrao.edu

MMA Project Book, Chapter 3 Section 2.

Calibration: hardware schemes

Darrel Emerson
Last modified July 16 1998

1. Introduction

The previous section describes a number of schemes for calibrating the MMA amplitude and phase. This section outlines two specific hardware calibration schemes which can help to calibrate the instrumental phase and amplitude of the MMA. Other instrumental calibration schemes, such as round-trip phase calibration for the local oscillator, and AGC/total power monitors in the I.F. chain, are described elsewhere.

The two calibration schemes outlined here are:

- **Absolute temperature sensitivity calibration,** single dish mode
- **Relative amplitude and phase calibration,** with an artificial coherent calibration signal suitable for interferometric calibration

2. Absolute Temperature Calibration

This technique is only relevant to single dish total-power observations; it has been suggested by Jack Welch and others, and is currently (July 1998) the subject of a joint MDC development between BIMA and NRAO. It gives an absolutely calibrated signal of a few K at the receiver, over the complete frequency range covered by the MMA. It has been described in an (as yet unpublished) memo by Bock, Welch, Flemming and Thornton.

The essence of this technique is that a black body radiator is placed at the center of the subreflector of each antenna, within the unused area of subreflector matching the central blockage of the antenna. In this way there is no effect on antenna sensitivity.

Within this central part of the subreflector, a plane mirror switches between two (or more) hot loads of different temperatures. The two loads have very precisely controlled and calibrated temperatures. The total power output of the receiver is sampled synchronously with the mirror switching between the two calibrated loads.

The added switched receiver noise is, to a first approximation, equal to the difference in temperature of the two hot loads, multiplied by the beam solid angle of the absorbers at the subreflector seen from the receiver feed, divided by the beam solid angle of the subreflector - assuming the receiver feed itself is matched to the angle subtended by the receiver. This ratio will be reasonably constant with frequency, but at a given frequency can be calibrated precisely by measurements of the feed antenna pattern.

For more details see the memo by Bock, Welch, Flemming and Thornton. The joint development with BIMA will show, on a timescale of a few months, how well the technique can be expected to work in practice.

Reference:

D. Bock, J. Welch, M. Flemming and D. Thornton, 1998 unpublished memo: *"Radiometer Calibration at the Cassegrain Secondary Mirror."* (See the Appendix at the end of this chapter.)

3. Interferometric relative phase and amplitude calibration

In the debugging stage of the MMA, there will be a need for a generic test signal that can be used to debug the entire electronic system of a given antenna or antenna pair, from front-end to correlator. When the antenna surface and pointing are sufficiently reliable, astronomical sources can be used for this purpose, but having an independent, artificially generated signal that is not dependent on antenna performance will be invaluable in checking out and maintaining the system.

If the calibration signal can be made coherent at all individual antennas, it opens up the possibility of calibrating the entire receiver system, front-ends, back-ends and correlator, amplitude and phase as a function of frequency, in a way independent of antenna tracking, pointing, or efficiency performance. The calibration system should be sufficiently stable that it can be used as a secondary calibration system, with only occasional cross-calibration with astronomical sources.

The Photonic Calibration System

The photonic calibration system has a broad-band, radiating antenna situated at the center of the subreflector, where no extra antenna blockage is introduced. At the feed of the broadband antenna, there is an uncooled photomixer device. A single optical fiber, carrying laser signals generated at a central laboratory or control room, feeds this photomixer. In the simplest form, the optical signal would come from two lasers, whose difference frequency corresponds to the telescope observing frequency, and which is phase-locked to the telescope frequency standard. The equipment required to do this would be nearly identical to that being developed for the photonic laser local oscillator system. Only one pair of lasers would be required for the entire array; the combined laser output would be split optically N ways (where N is the number of antennas) and routed via N independent fibers to each antenna.

In slight variants of this scheme, either a single laser signal, or the dual laser system tuned to the required mm-wave difference frequency, could be modulated. The modulation might take the form of a regular comb spectrum, simulating broadband noise. This becomes quite analogous to the pulse cal system developed for the VLBA, and could be used for checking the relative amplitude and phase response over the entire interferometric IF passband. The modulation might also be a truly random, or a pseudo-random digitally generated sequence, which would also provide a broad-band coherent test signal. This random or pseudo-random noise needs to be coherent at each antenna, so timing considerations, within a fraction of the reciprocal bandwidth, are important.

Naturally this injected signal needs to be stable, both in amplitude and phase. It may require round-trip delay compensation of some type, and perhaps an AGC system to keep the signal amplitude constant. However, attention to the stability of this calibration signal may relax the technical requirements elsewhere in the system.

Most of the development for this coherent photonic calibration scheme is already being undertaken in the context of the photonic local oscillator development. The calibration scheme should in principle be much simpler, because several orders of magnitude lower radiated mm-wave power is required. The main additional development needed is that of the broad-band radiating antenna, to be sited at the subreflector, fed by the signal from the photomixer. The broadband antenna is the subject of a joint

NRAO-MDC (OVRO) development project.

4. Combined Calibration System

The incoherent calibration scheme described earlier switches between blackbody radiators of different temperatures using a mirror. In principle, by allowing an extra position on this mirror, the radiated calibration signal can be switched between the incoherent blackbody loads and the coherent radiator. At a later stage in the development, when the feasibility of both coherent and incoherent calibration schemes has been demonstrated, the combination of the different calibration radiators into one package will receive attention, as will studies of how to achieve the necessary amplitude and phase stability.

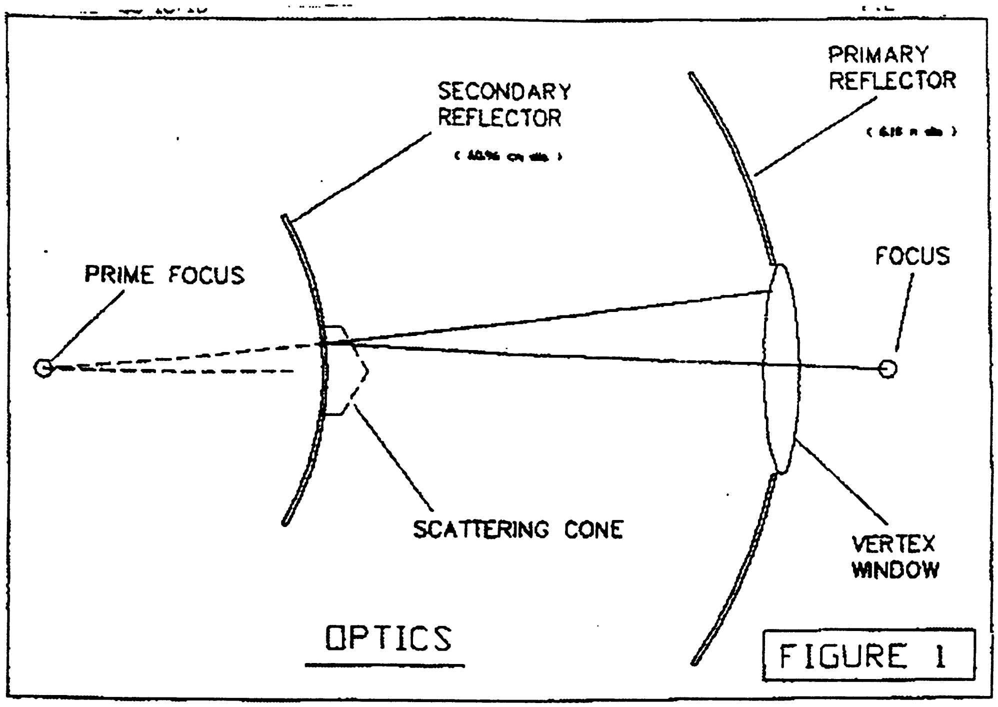
5. Work to be done

Much of the work is being carried out at BIMA (incoherent calibrator) and OVRO (coherent calibrator). The photonic calibration builds on work already in progress in the context of the photonic local oscillator scheme, with the exception of the broadband antenna. Some simple design work is required now (e.g. exactly how much coherent power needs to be radiated from the subreflector, with what requirements on amplitude and phase stability?) but the bulk of the effort can be expected fairly late in the MMA development phase.

APPENDIX

The following figures are from the unpublished memo by Bock, Welch, Fleming and Thornton, *"Radiometer Calibration at the Cassegrain Secondary Mirror."*

Figure 1 shows the general Cassegrain optics, which normally has a scattering cone covering the central part of the subreflector to direct unwanted rays on to cold sky. **Figure 2** shows a scattering mirror behind the subreflector, giving much the same effect.



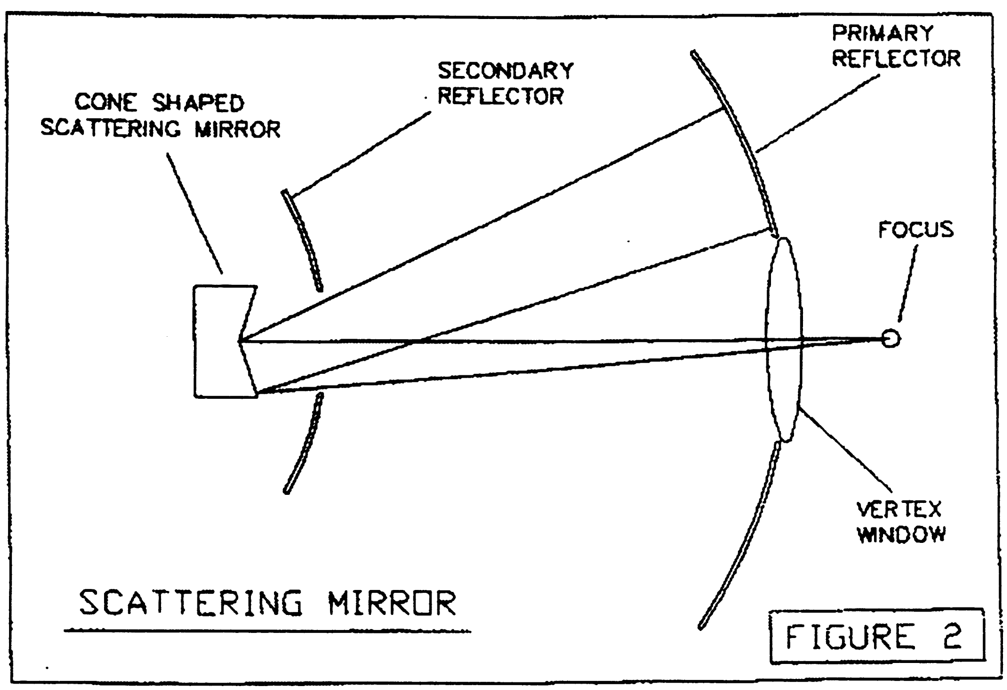


Figure 3 shows the absorbing black body load that would be placed behind the central hole of the subreflector

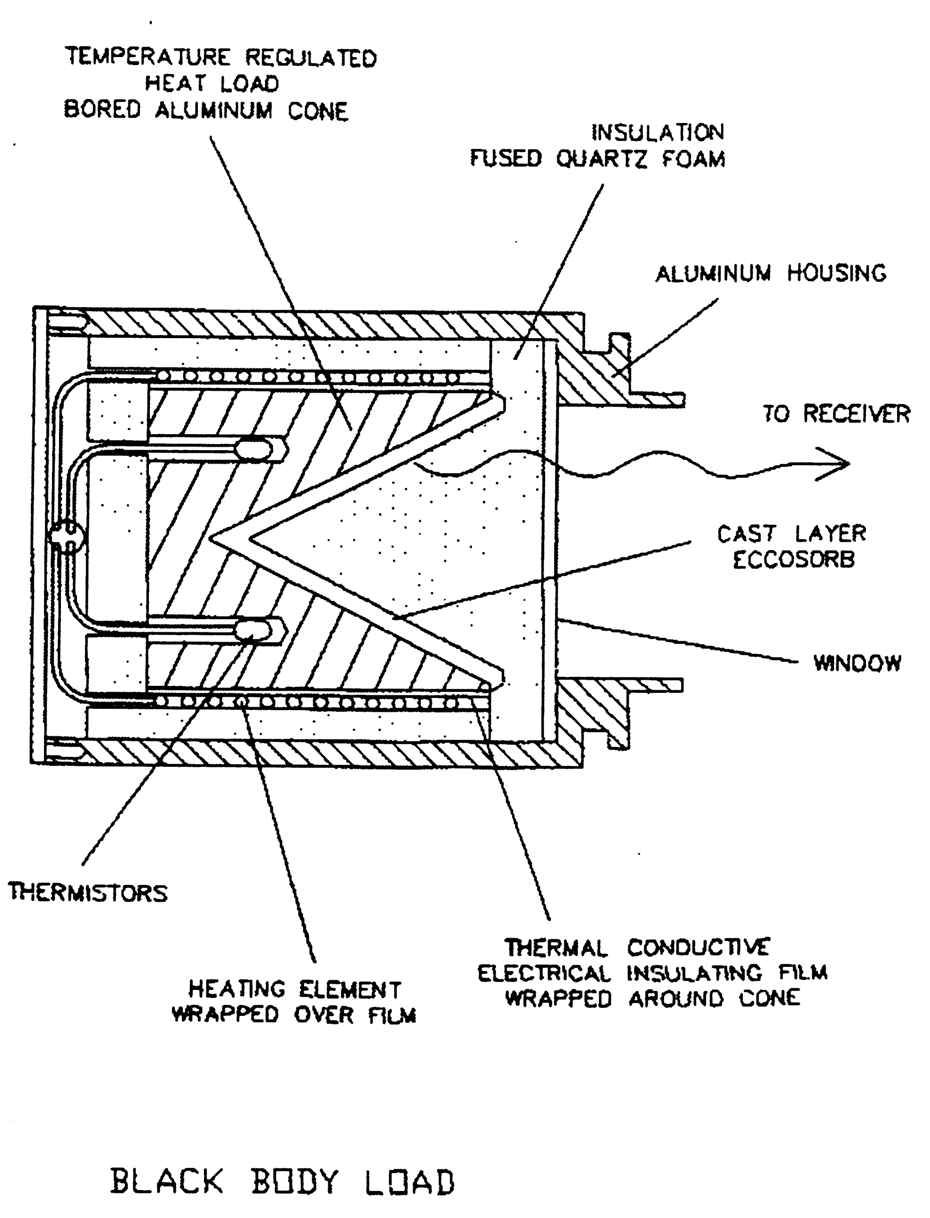


FIGURE 3

Figure 4 shows the arrangement of a rotating 45-degree mirror which will choose between one of two hot loads, whose temperatures differ by ~100 K, and the scattering cone from which rays which eventually reach cold sky.

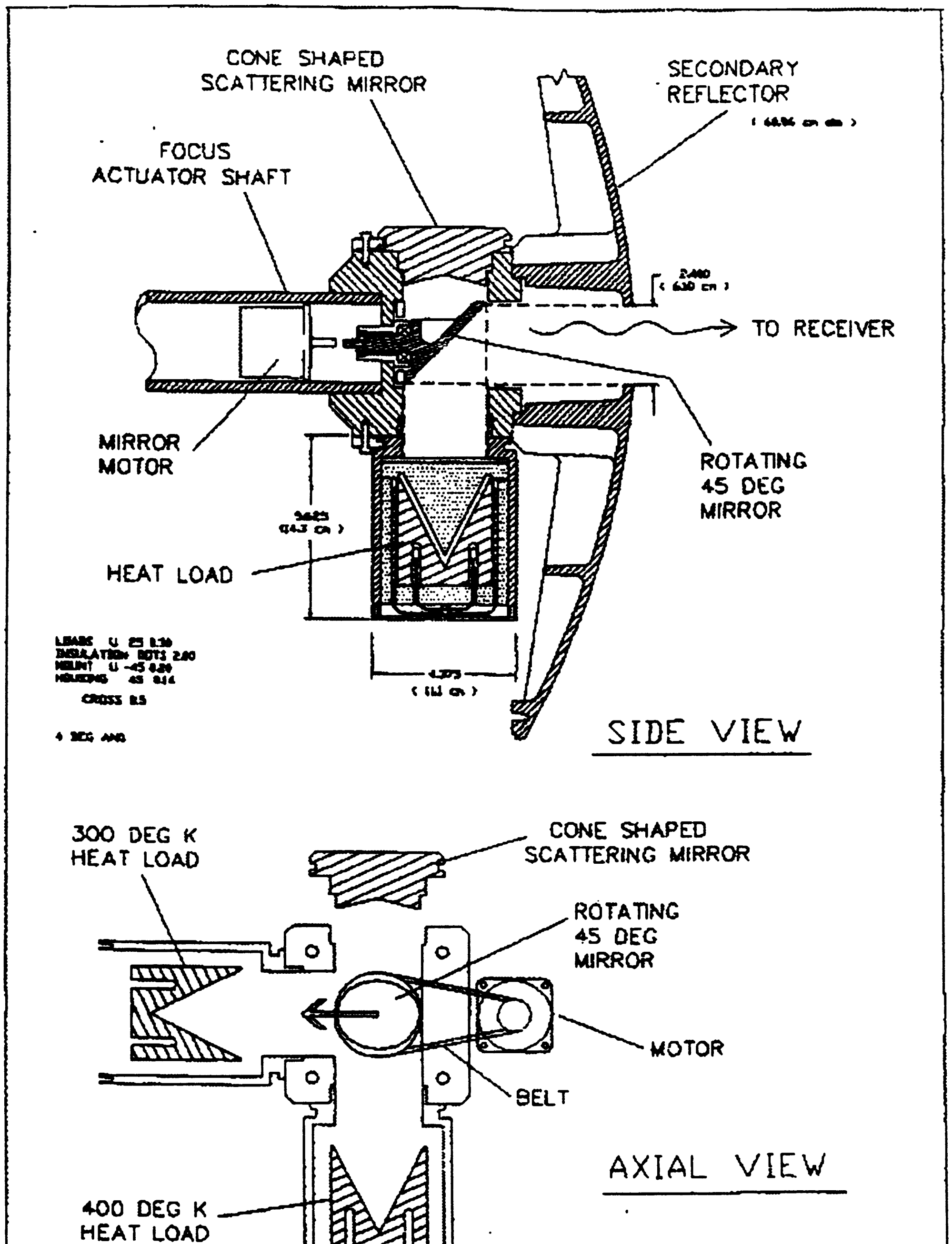

CONE SHAPED SCATTERING MIRROR
SECONDARY REFLECTOR
FOCUS ACTUATOR SHAFT
TO RECEIVER
MIRROR MOTOR
ROTATING 45 DEG MIRROR
HEAT LOAD
SIDE VIEW
300 DEG K HEAT LOAD
CONE SHAPED SCATTERING MIRROR
ROTATING 45 DEG MIRROR
MOTOR
BELT
AXIAL VIEW
400 DEG K HEAT LOAD

FIGURE 4

Peter Napier
Last modified July 16 1998

ANTENNAS

1. Introduction

This chapter presents the current state of understanding of the performance requirements for the MMA antenna elements and describes the current strawman design concepts which will satisfy these requirements. Also described are the current plans for the procurement and testing of the antennas.

The "antenna" subsystem of the MMA is here defined to include the following equipment:

10 m diameter primary reflector including tripod or quadripod subreflector support legs.

Secondary reflector and its servo-controlled positioning platform, including nutation.

A receiver cabin and its HVAC system.

Alt/az mount, the drive systems on the mount and the servo-system controller for the drives.

Metrology instrumentation such as temperature probes, tiltmeters, laser metrology systems, etc.

Power distribution cabling on the antenna and the cable wraps for these cables.

Platforms for mounting auxiliary equipment such as cryogenic compressors.

Antenna foundation design but not fabrication.

Antenna transporter vehicle.

The following equipment is not included as part of the "antenna" subsystem and must be supplied as part of another subsystem:

Cables and special cable wraps required for IF and LO signal distribution. The antenna will provide suitable mounting interfaces for this equipment.

2. Specifications and Requirements

2.1 Operating Environment

The following operating environment defines the environment on the MMA site in Chile. The first few antennas will be tested at a location in the Continental US and it is possible that all antennas will be assembled and undergo preliminary testing at San Pedro de Atacama (altitude 2440 m). Before ordering the first antennas, these environmental specifications must be reviewed to ensure that they are adequate to ensure safety of the antenna at these alternate test location sites.

2.1.1 Location: Northern Chile, latitude -23d01m S, longitude 67d45m W.

2.1.2 Altitude: 5000 m (16400 ft) The barometric pressure at this altitude is 55% of its sea-level value.

2.1.3 Maximum Wind Velocity: The antenna must survive 65 m/sec (130 mph) without damage when positioned in its stow position.

2.1.4 Temperature: The antenna must operate correctly in the temperature range -20 C to 40 C. The annual average temperature on the site is -4 C.

2.1.5 Precipitation: Annual precipitation on the site is in the range 10 cm to 30 cm per year. Most of this falls as snow but thunderstorms do occur and so brief periods of heavy rain and hail are possible. The antenna must be designed to survive, without damage, the following conditions: maximum rate of rainfall 5 cm/hr, hailstones 2 cm diameter, snow load 100 kg/sq.m on reflector surface, radial ice on all exposed surfaces 1 cm. Surface heating to prevent snow and ice buildup not required.

2.1.6 Humidity: The monthly average humidity in the summer (January) is 53% and in the winter (June) it is 31%. The annual average is 39%. The monthly average water vapor pressure in the summer (January) is 4.0 hPa (4 gm/sq.cm) and in the winter (July) it is 1.2 hPa. The annual average is 2.3 hPa.

2.1.7 Insolation: The site location on the southern tropic, the high altitude and low water vapor result in insolation rates amongst the highest in the world. The median midday solar flux in the wavelength range 0.3-60 micrometers for the months of December and June are 1290 w/sq.m and 840 w/sq.m respectively. Ultraviolet radiation will be approximately 70% higher than at sea-level.

2.1.8 Lightning. Thunderstorms occur on the site so the antenna must be equipped with a lightning protection system.

2.1.9 Dust and Grit. The site ground surface is volcanic soil and gravel with no vegetation of any kind to stabilize the surface. It is likely that wind-blown dust and grit will be a factor for machinery operating on the site but this problem has not yet been quantified.

2.2 General Configuration

The antenna will be a symmetric paraboloidal reflector, of diameter 10m, mounted on an elevation over azimuth mount. Subreflector support legs will be either tripod or quadripod configuration. The reflector will consist of machined aluminum panels mounted on a carbon fiber reinforced plastic (CFRP) reflector backup structure (BUS). The BUS could be built completely of CFRP or could consist of CFRP struts connected by metal nodes. The tripod or quadripod will be made of CFRP. The reflector surfaces of the antenna will not be painted.

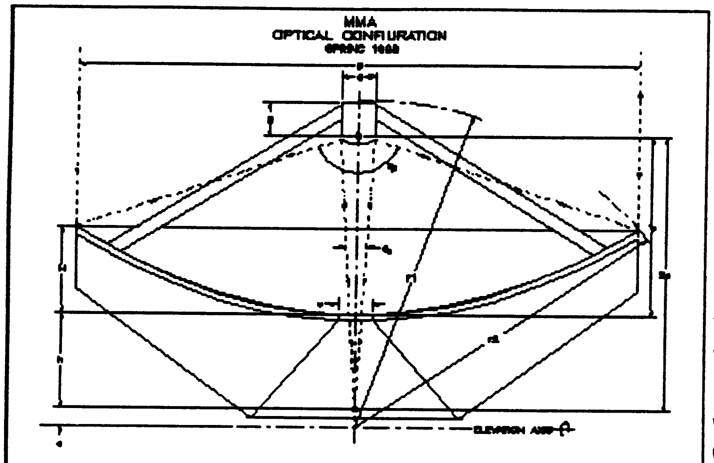
The antenna will be designed so as to conform to all relevant Occupational Safety and Health Administration (OSHA) safety codes.

2.3 Reflector Geometry

The receivers will be located at the secondary focus of a Cassegrain geometry. A strawman design for the Cassegrain geometry is shown in Figure 1 (taken from Lugten (1998)) but there is still flexibility to make changes to this layout.

2.4 Range of Motion

Antenna foundations will be constructed so that the azimuth axis of an antenna is parallel to local gravity at the pad. For observations close to the zenith this will result in a difference in parallactic angle between antennas.



	OPTICAL CONFIGURATION			
D	Primary diameter		10.01 m	394.0 inches
f_p	Focal length of primary		3.803 m	149.72 inches
	f_p/D of primary		0.38	0.38
d	Secondary diameter		0.762 m	30.00 inches
	Final f/D		8.78	8.78
	Magnification factor		17.85	17.85
ϕ_p	Primary angle of illumination		133.38 deg	133.38 deg
ϕ_s	Secondary angle of illumination		8.44 deg	8.44 deg
2c	Distance between primary & sec. foci		5.327 m	209.72 inches
H	Depth of primary		1.646 m	64.803 inches
r1	Tipping structure apex radius (1" clearance)	[illegible] 1.320	6.624 m	260.8 inches
r2	Tipping structure dish radius (1" clearance)	[illegible] 1.280	6.388 m	251.5 inches
h	Distance from vertex to secondary focus		1.524 m	60.00 inches
e	Distance from elevation axis to focus		0.762 m	30.00 inches
q	Distance from primary focus to top of quadrapod		0.508 m	20.00 inches
v	Clear aperture through primary		0.660 m	26.00 inches

Figure 1 Strawman MMA antenna Cassegrain Geometry

Minimum elevation angle for observations: 0 deg

Maximum elevation angle for observations: 125 deg. Cone of avoidance at the zenith: 0.2 deg in radius for normal sidereal tracking. Because of the high velocities and accelerations required for fast switching or on-the-fly mapping (see section 2.8 below) there will be a region around the zenith, probably about 30 deg in radius, where azimuth switching times are degraded .

Stow position for wind survival: elevation 0 deg (this position was chosen so that, during a winter storm, the reflector can be oriented with its back into the wind to prevent build up of snow and ice in the dish. In the event that the close packed array configuration requires a higher limit than 0 deg. (see Section 2.11 below), the high wind stow position will be this higher limit.

Stow position for maintenance: zenith (this position was chosen to prevent an antenna undergoing maintenance from mechanically interfering with an adjacent antenna in the most compact array).

Range of azimuth motion: 270 degrees either side of due north.

2.5 Reflector Surface Accuracy

The surface accuracy will be no worse than 25 micrometers rms, including the subreflector contributions. This will provide an antenna surface efficiency of 91% at 300 GHz and 41% at 900 Ghz. At night this accuracy is to be achieved in a wind of 9.5 m/s which is the 90th percentile wind for nighttime (2000 hrs to 0800 hrs) observing. During the day this accuracy is to be achieved for any orientation of solar illumination in a wind of 6 m/s. During the day the focus can be calibrated astronomically every 30 min.

The final, precision measurement of the surface will be done using holography. Reflector panels will be adjusted manually from the front side of the reflector with the antenna pointed at the horizon and the mechanic supported on an external platform or cherry-picker. The panel adjusters will be calibrated so that an adjustment point can be moved with a resolution of 5 micrometers without needing a dial-gauge or other motion indicator. A full surface adjustment must require no more than 16 person-hours of work.

2.6 Pointing Accuracy

An absolute pointing accuracy is required of 1/30th primary beamwidth rms at 300 GHz, which corresponds to 0.8 arcsec. At night this accuracy is to be achieved in a wind of 9.5 m/s which is the 90th percentile wind for nighttime (2000 hrs to 0800 hrs) observing. During the day this accuracy is to be achieved for any orientation of solar illumination in a wind of 6 m/s. During the day the pointing can be calibrated astronomically every 30 min if required to meet the accuracy requirement.

2.7 Metrology

Provision will be made in the antenna design for inclusion of metrology equipment which will allow antenna pointing to be corrected for structural deformation caused by wind or thermal loading. This equipment will be included on the prototype antenna both to evaluate the performance of the antenna and to determine if the metrology equipment is required on the production antennas. Metrology systems to be considered for including in the antenna include: a laser/quadrant-detector system to measure quadripod movement, tiltmeters, temperature probes, laser/retroreflector systems and an IR camera for offset-pointing on stars.

2.8 Fast Motion Capability

Two observing modes require the MMA antenna to have special fast motion capabilities: fast switching phase calibration and on-the-fly total power mapping.

Fast switching: The goal is to have the antenna move 1.5 degrees on the sky and settle to within 3 arcsec pointing error, all in 1 second of time. A possibly acceptable upper limit for this switching time is 2 seconds. It is expected that the switching acceleration profile will be carefully designed so as to avoid exciting the lowest structural resonant frequency of the antenna, in which case the lowest resonant frequency should not be lower than 6 Hz. The maximum velocity and acceleration required for fast switching are 3 deg/sec and 12 deg/sec/sec on the sky respectively, with both axes able to move at this rate simultaneously. It is expected that this velocity and acceleration will be achievable in azimuth only for zenith angles greater than 30 deg (this implies maximum azimuth velocity and acceleration of 6 deg/sec and 24 deg/sec/sec respectively).

Analysis of the expected use of this fast switching mode (Holdaway, 1997) indicates that the antenna should be designed to survive 30-50 million cycles of fast switching during an assumed 30 year life.

On-the-fly mapping: In this mode the antenna will scan at a rate of up to 0.5 deg/sec across a large object, several or many beamwidths in size, and then turn around as rapidly as possible and scan back across the source in the opposite direction. A maximum acceleration of 12 deg/sec/sec is required for the turn around. While the antenna is scanning across the source the antenna position must be recorded at a rate sufficient to provide an angular sampling interval on the sky of wavelength/(2D) radians. For 0.5 deg/sec motion and 900 GHz observations this requires antenna position readout every 2 msec. The antenna positions should be accurate to 1 arcsec. As the antenna tracks across the source it is not necessary for the position at any time to be precisely a precommanded position; it is sufficient to simply know where the antenna is actually pointing and all antennas need not point precisely at the same position.

A third observing mode requiring motion of the antenna faster than sidereal tracking rate is on-the-fly interferometric mosaicing, in which interferometry data is taken while the antenna is continuously scanning across the source. It is expected that the antenna velocity will be only one-tenth of its mapping-on-the fly value (see previous paragraph), but in this case all antennas must point to the same position at the same time to within 1 arcsec rms.

Fast slew to a new source position: 1.5 deg/sec elevation, 3 deg/sec azimuth. These velocities are less than the maximum velocities specified above for the antenna in fast-switching mode. This is because fast motion for fast-switching is required with low duty cycle whilst slewing could last for many seconds possibly resulting in motor overheating.

2.9 Subreflector Position Control

The subreflector will be supported on a platform which allows movement in all 3 linear directions. The precision of the mechanism will be adequate to allow the subreflector to be positioned, under computer command, with sufficient accuracy to prevent gain loss of more than 1% at 900 Ghz due to focus, comatic or astigmatic aberration. Position will be correctable on timescales of tens of seconds.

In addition to the above listed linear motions the first antenna will also be equipped with a subreflector nutator which will allow beam throws of three beamwidths at 86 GHz (4.3 arcmin) at rates up to 5 Hz in the cross-elevation direction only. The decision as to whether all antennas will be equipped with nutators will be made after testing the first antenna.

2.10 Phase Stability

Phase errors caused by variations in the propagation pathlength through the antenna can be rapidly or slowly varying. Fast phase changes are primarily caused by the wind and the peak pathlength variation in a 9.5 m/s wind must be no more than 7 microns. Slow phase changes are primarily due to variations in the temperature of the antenna and the goal is to keep these phase errors small enough so that the residual errors after an astronomical phase calibration every 3 min are small enough to allow observations at 900 Ghz.

2.11 Close Packing

In the smallest array the antennas must be placed close together. The goal is to be able to place the antennas within 12.5 m (1.25 D) of each other without any possibility of the antennas hitting each other, no matter what the relative orientation of the two antennas. An acceptable fallback on this requirement would be to have no possibility of interference for elevations above 20 deg. In the event that it proves necessary to have a higher elevation limit of this type when antennas are in the close packed configuration, an electronic interlock on the antenna pad will ensure that the higher limit is activated.

2.12 Solar Observing

Direct observations of the sun will be allowed. All surface accuracy and pointing requirements must be met while observing the sun and a suitable surface treatment of the primary reflector surface must be provided to prevent solar heating damage of the subreflector or its support legs. When observing the sun the solar heating of the secondary focal plane must be less than 100 watts.

2.13 Low Antenna Noise

Contributions to system noise from the antenna, due to such mechanisms as scattering of ground noise into the feed and resistive loss of reflector surfaces, will be minimized as much as possible without compromising the surface accuracy and pointing requirements. Design features to be considered to achieve this goal include supporting the subreflector support legs close to the edge of the reflector, shaping the underside of the support legs to reduce ground pickup and locating the feeds at the Cassegrain focus to avoid the need for tertiary reflectors.

2.14 Transportability

To move the antennas from one array configuration to another the antennas will be picked up and carried on a transporter vehicle which runs on a gravel road on rubber tires. The transporter with an antenna on board will be able to negotiate a 15 % grade, turn a corner with a minimum turning radius of 10 m and travel at 10 km/hr on the flat and 5 km/hr up a 10% grade. An unloaded transporter must be able to travel at 20 km/hr on the flat. The transporter must be able to safely move an antenna in winds up to 16 m/s (this is approximately the 95th percentile for the winds on the site at 1600 hrs local time, the time at which the winds are maximum each day). A stationary transporter with an antenna on board will survive winds up to 65 m/s; if necessary, structure carried on the transporter can be deployed to stabilize the transporter on the ground in this survival mode.

The transporter will carry an auxiliary generator to keep all electrical systems on the antenna operational during a move. The transporter will pick up the antenna above its azimuth bearing so that the azimuth bearing and drive can be used to rotate the base of the antenna to simplify bolt hole alignment when an antenna is placed on a pad. It may be desirable to oxygenate the air in the transporter operator's cabin so the cabin must not have large air leaks.

When an antenna is picked up a time goal of 20 min is required from the time of arrival of the transporter to the time of departure with an antenna on board. When an antenna is placed down on a pad a time goal of 30 min is required from the time of arrival of the transporter until the transporter has departed and the antenna is ready to be pointed.

2.15 Receiver Cabin

A receiver cabin with dimensions approximately as shown in Figure 1 of Napier et. al., 1996,
Napier et.al. 1996
will be provided at the Cassegrain focus. Temperature in the cabin will be maintained by an antenna mounted HVAC system at 16 C to an accuracy of +- 1C. A built-in mechanism will be provided so that a receiver can be lifted from the ground, through the cabin door and into its observing location, all without significant man-handling of the receiver. Part of the installation of a receiver may involve the use of a separate special purpose vehicle, such as a high fork-lift, which lifts the receiver through the cabin door.

The cabin will be watertight and a thin RF-transparent membrane will cover the aperture through which the RF beam enters the cabin. A computer actuated shutter will be deployable to protect the membrane when necessary.

It may be desirable to oxygenate the cabin air when workers are inside so the cabin must not have large air leaks.

2.16 Monitor and Control

The following functions on the antenna will be controllable under remote computer control:

Antenna position and scan rate

Subreflector position in x,y,z and nutation

Power distribution switching from normal to critical power and complete power down

Receiver cabin HVAC temperature set point

The following functions on the antenna will be monitored by a remote computer:

Antenna position and rate (velocity and direction)

Motor currents and all power supply voltages in the servo system

Subreflector position

Readout from any metrology devices

Readout from any temperature probes

HVAC system performance

Limit switch status

The following fault conditions will be automatically sensed and acted on at the antenna:

Power down after smoke detection

Power down after loss of a phase of the power supply

Loss of a drive motor

Drive shutdown if antenna oscillation detected

Antenna stow if command link from control building lost

2.17 Interfaces to Other Subsystems

The following interfaces must be defined:

Monitor and control digital interface

Interface between antenna and transporter

Interface between subreflector support legs and the subreflector support mechanism

Interface between antenna and its concrete observing pad

Interface to the electrical power system

Interface to any special cable wrap required for RF signals

Interface to the receiver package in the receiver cabin

Interface to any equipment racks in the receiver cabin

Installation procedure for receiver package

2.18 Maintenance and Reliability

Because of the remote site and large number of antennas the reliability and maintainability of the antennas are important. The antennas will be designed so that, with normal preventive maintenance, they should operate for 30 years without requiring elevation or azimuth bearing or reflector surface replacement. Although they should not be required, straightforward elevation and azimuth bearing replacement procedures must be included in the antenna design. All normal repair and maintenance actions should be able to be completed by a two- person crew in 4 hours. To the maximum extent possible all equipment on the antenna should be "modularized" so that a failure can be cured by simply swapping out the failed component without the need for any repair in place. Examples of equipment which should be designed for easy replacement includes gear boxes, drive motors, HVAC equipment, servo-system electronic components and the subreflector position control mechanism.

2.19 Manufacture and Assembly

The antenna will be designed for economic production costs.

It is expected that the first two antennas will be tested initially at a site in the US and later shipped to the MMA site so the ability to disassemble the antenna into pieces for overseas shipping is required.

The high altitude and remoteness of the MMA site make it desirable to minimize the amount of work required on the high site. It is expected that the antennas will be assembled, outfitted and tested at an Operations Support Facility 50 km from the MMA site at an altitude of 2400 m. They will be carried to the MMA site on the transporter vehicle or, in the event that this proves not to be feasible, they will be disassembled into just two pieces, the mount and the reflector, for transportation to the site on trucks. Thus the antenna should be designed for easy disassembly at the elevation axis and both the reflector and mount must have pickup points for handling as single units.

3. Current Design Concepts

At present, two antenna concepts are being investigated, both of which are expected to meet MMA requirements.

3.1 Concept A

This concept has been developed principally within NRAO and is described in Section 4.2.3 of Lugten et.al., 1998b.
Lugten et.al., 1998b

3.2 Concept B

This concept has been developed principally within OVRO and is described in Woody and Lamb, 1998.
Woody and Lamb, 1998

4. Procurement and Construction Plans

The plan for procuring the MMA antennas is currently as follows:

(1) The MMA Antenna Working Group is developing two design concepts in order to be sure that we fully understand all of the requirements of the antenna and that these requirements are affordable within the MMA budget. Full information on MMA antenna requirements and design concepts will be given to interested companies at a workshop in September 1998 to allow the companies to begin preparing for the bid.

(2) An RFP (Request for Proposal) for the design and fabrication of the first MMA antenna will be issued at the end of 1998. The procurement will be a fixed-price contract to a performance specification. All information developed by NRAO in its studies of the concepts described in Section 3 above will be provided to prospective bidders, and bidders can propose one of the NRAO concepts or an alternative design, but in either case it will be the responsibility of the contractor to meet the performance specifications. In the RFP response the bidders will be required to describe their proposed design in some detail, to provide the cost of the design and fabrication of the first antenna and a not-to-exceed estimate for the cost of the antenna in production quantities. An option for a second copy of the antenna, to be exercised in October 2000, will also be requested.

(3) A possibility, still under consideration, is that two parallel design contracts will be issued to two different companies, with the first antenna fabrication contract to be awarded competitively, based on predicted performance and production cost, after completion of the design phase.

(4) The first antenna will be delivered in June, 2001, and then tested extensively for 1.5 years so that there is high confidence that the design is adequate to meet the astronomical requirements.

(5) With the antenna design validated by adequate testing, the production run of the antennas can be procured on a "build-to-print" basis rather than to a performance specification. That is, NRAO can take the risk that the production antennas will meet performance requirements, thereby reducing cost and increasing the number of possible vendors who could produce antenna components. The RFP for the production run will be issued in January 2003, with a required antenna delivery of 8 antennas per year.

The antennas will be fabricated in the vendor's factory and pieces of the antenna will be accepted on the basis of mechanical acceptance tests in the factory before the pieces are shipped to the Operations Support Facility for assembly. It is not yet decided whether antenna assembly will be performed by a contractor or by NRAO. The main reason to consider having NRAO take responsibility for assembly is that in the long term NRAO must have full capability to mainain, repair and improve the antennas on the remote site and an effective way to develop this capability would be to also assemble the antennas.

5. Test Plans and Results

Testing of the first two antennas will take place at the VLA site. The first antenna will be tested initially in single-dish mode. Single dish tests will include pointing accuracy, single-dish holography, surface stability, elevation dependent gain curve, fast movement ability and dynamic performance. When the first interferometer is available testing will be extended to phase stability and interferometric holography.

6. Acknowledgements

This chapter represents the work of the MMA Antenna Working Group: J. Bieging, J. Cheng, D. Emerson, M. Fleming, M. Holdaway, J. Kingsley, J. Lamb, J. Lugten, J. Mangum, J. Payne, J. Welch, D. Woody

7. References

M. Holdaway, "How many switching cycles will the MMA make in its lifetime", MMA Memo 174, 1997

J. Lugten, 1998a, New Strawman Optics Memo, in preparation.

J. Lugten, J. Cheng, M. Fleming and J. Kingsley, "Antenna design for the Millimeter Array", SPIE paper 1998b,.

P.J. Napier et. al., "A strawman optics layout for the MMA antenna", MMA Memo163, 1996.

J. Woody and J. Lamb, "Yet another design for MMA antennas", MMA antenna note, 1998.

MMA RECEIVERS: -- Prototype Receivers

Harry Fagg
Jeff Clarke
Last revised July 20 1998

1. Prototype Receivers

1. Introduction

It will be necessary to construct a first-generation receiver system as a test instrument for the prototype MMA antennae. This prototype receiver will also serve as a test bed for many state of the art components and procedures to be used in the MMA operational receiver systems. This section will present electrical configurations for two observational test bands and one atmospheric phase correction band.

2. The Bands and Configurations

The observational test bands are chosen to provide thorough testing of the MMA prototype antennae at a good test site. Desire to cover certain spectral lines for antenna and receiver tests drive the choice of test bands. These test bands may, or may not be identical to observational bands desired for the operational MMA receiver systems (see the Whitepaper "Frequency Band Considerations and Recommendations"). Other limitations of test band choice include test site observing conditions. For these reasons, the chosen prototype receiver bands are: 210-270 GHz, 86-115 GHz and the atmospheric test band at 183 GHz. A final IF band of 4-12 GHz has been tentatively chosen to provide the desired 8 GHz IF bandwidth; however, for the prototype the IF band will be restricted to about 1 GHz near the lower part of this range. The 210-270 GHz configuration is shown in Figure 1.

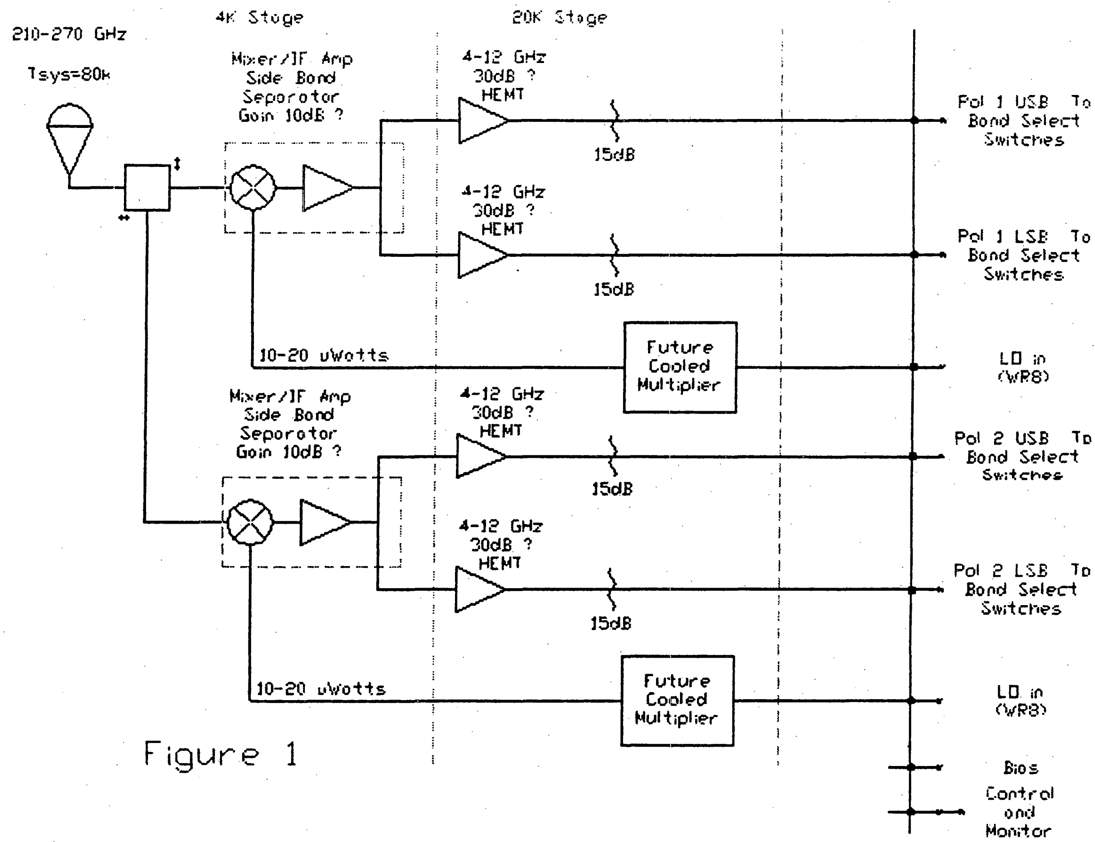


Figure 1: The 210-270 GHz configuration.
Note that the IF band shown in Figures 1 to 3 is 4 - 12 GHz;
the prototype system will be restricted to about 4 - 5 GHz

It is planned that this receiver system will implement the first NRAO use of an ortho-mode-transducer (OMT) at this frequency. An orthomode transducer for the 3-mm band has already been developed, and although this has yet to be scaled to the 1-mm band, we are optimistic that this will be successful. If that is not available, a conventional cross-grid polarization diplexer could be used instead. Also, a balanced, sideband separating SIS mixer is planned; a prototype mixer of this kind has recently been developed at the CDL with very promising results. A fallback position would be to use the existing single-ended mixers that have been in use at the 12 M Telescope for many years. The precise LO configuration is uncertain as of writing, and depends on the cooled multiplier development now in progress at the CDL. A fallback position is to copy the LO system presently in use at the 12 M Telescope. Four IF bands result from the two polarizations, each with two sidebands. Cooled HEMT amplifiers follow the mixer and drive room temperature amplifiers. This band, as well as the other observational test bands, is selected by a band-select switch which feeds a common IF filter and amplifier system. This common section is shown in Figure 2.

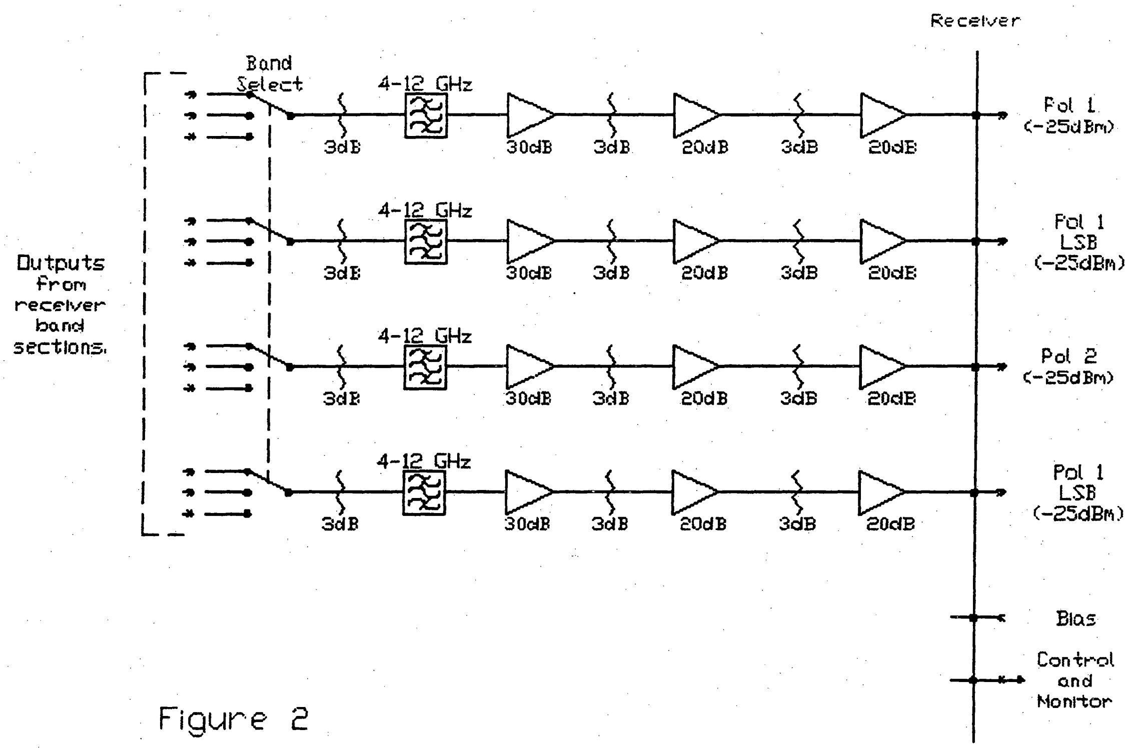


Figure 2: HFET downconversion and IF

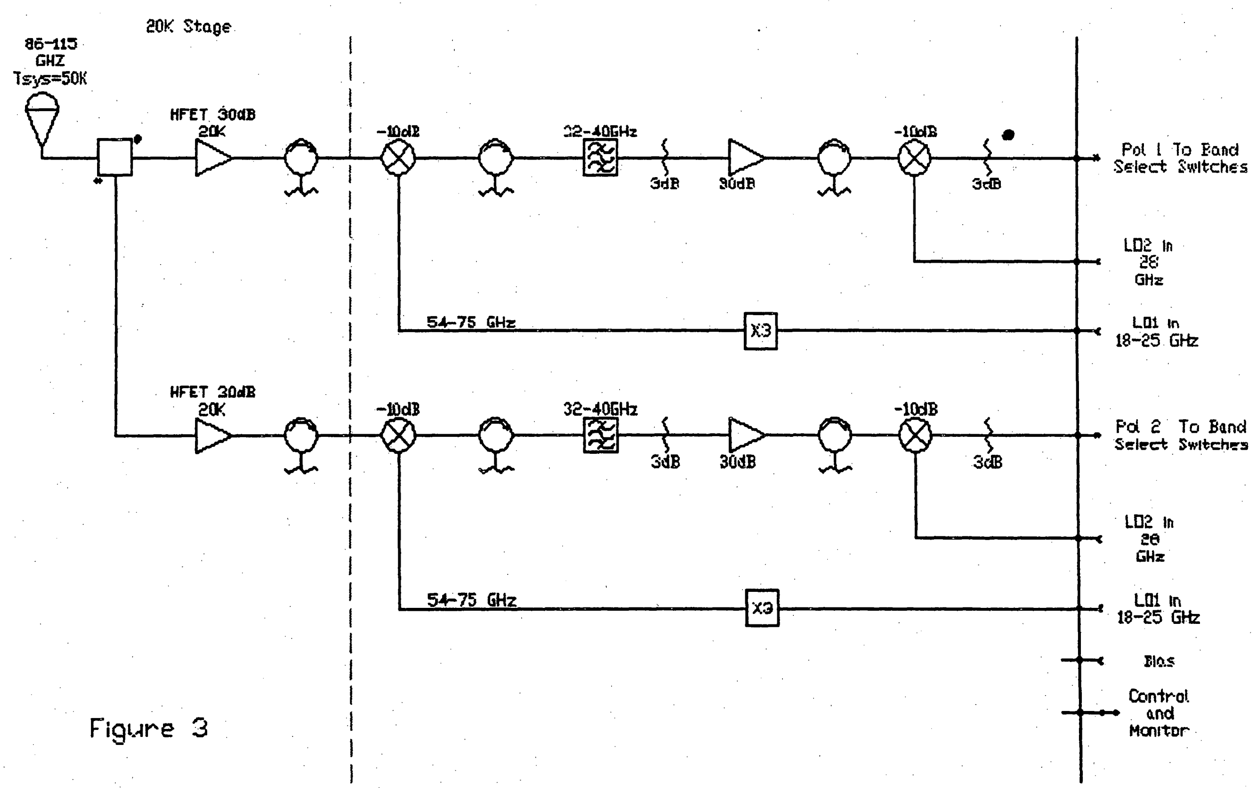


Figure 3: 86-115 GHz configuration

A schematic of the 86-115 GHz band configuration is shown in Figure 3. An OMT is used to separate the two polarizations in this configuration also. Two IF bands result. This band will afford the opportunity to test a HEMT amplifier front-end element at 100 GHz; the choice between HEMT and SIS for this band is still open. Down conversion to the final IF band is a dual conversion process. The first LO and first IF are chosen to fall into conventional waveguide bands. A second conversion results in the final 4-12 GHz IF band. The conversions, added gain and isolation take place in a room temperature environment. Again, the band-select switch, bandpass filtering and final gain stage are common to the three observational test bands.

3. Atmospheric Monitor Receiver

An atmospheric phase monitoring receiver is a desired component of this test receiver system. This receiver does not require wide band tuning and will use a single conversion with a fixed LO frequency at 183 GHz. This band configuration is shown in Figure 4. A decision on the type of mixer to be used has not been made. This decision will likely hinge upon the system temperature required for these measurements.

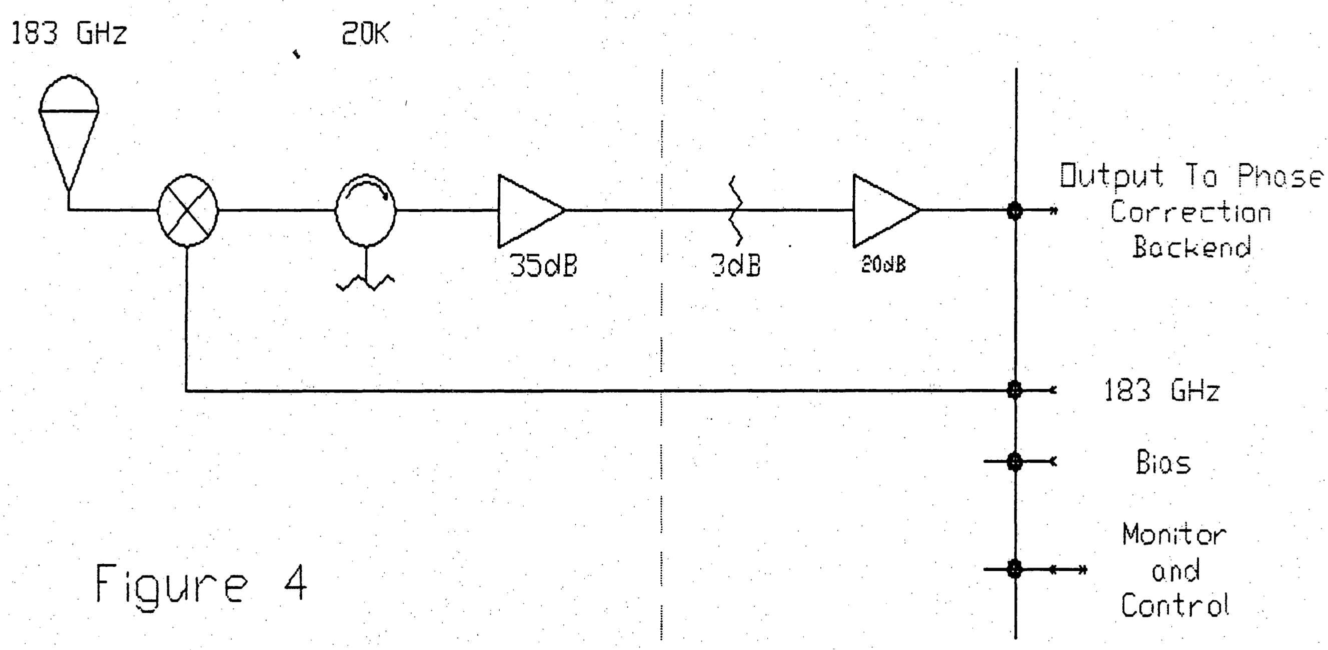


Figure 4: 183 GHz atmospheric monitor receiver configuration

4. Ongoing Discussions

Realization of this prototype MMA receiver system is dependent upon development of high frequency HEMT amplifiers and ortho-mode-transducers. Proven technology exists in SIS mixers and quasi-optical polarization splitters. Should it be necessary for the timely completion of this receiver system, these and other technologies could be considered. Discussions continue on approaches to the completion of this important test and prototype receiver system.

Back to Chapter 5 TOC | Back to Chapter 5 Section 1 top | On to Chapter 5 Section 2

MMA RECEIVERS: Receiver Optics

Most of this section is from
MMA Memo #163 and revisions,
from the MMA antenna group
(Lugten, Napier et al.)

Last revised July 10 1998

MMA RECEIVER OPTICS

Most of the material in this section is taken from MMA Memo #163, "A Strawman Optics Layout for the MMA Antenna" and later revisions of that memo.

The current concept is shown in Figure 1 below, and representative dimensions are given in Table 1. Some of the considerations which have led to this concept are discussed below. It is likely that some of the details will change as the antenna design proceeds. However, the various options currently being considered for the antenna design will have less impact on the optimum optical parameters than will the choice of feed arrangement.

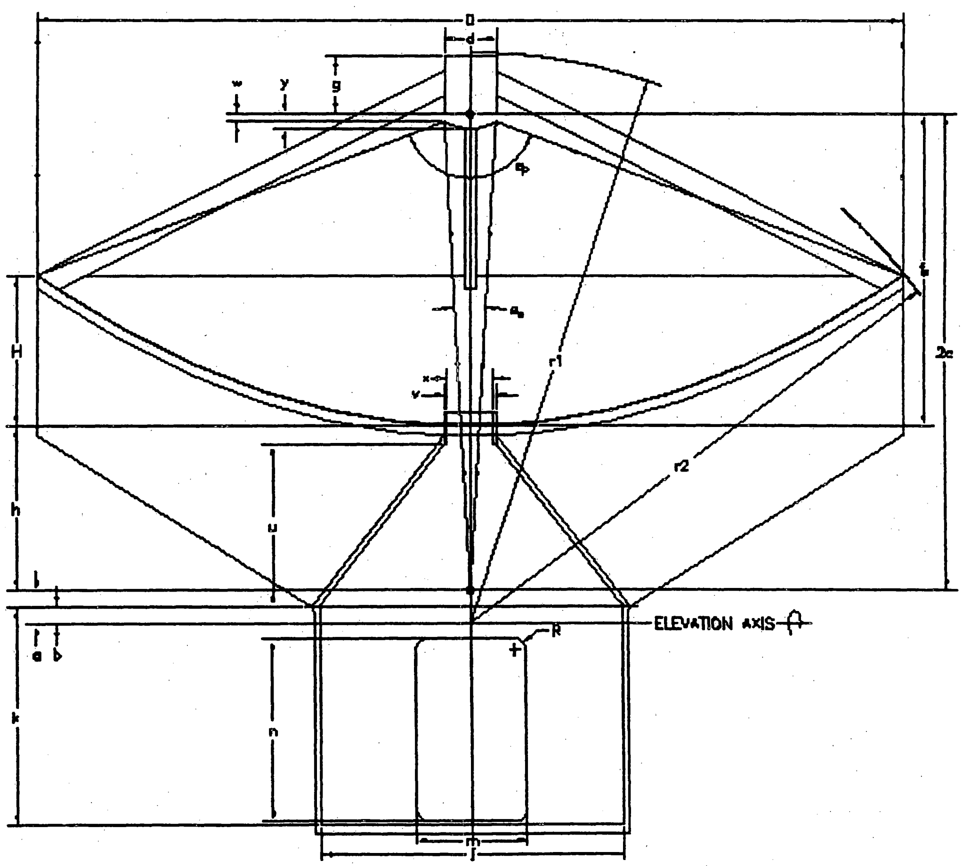


Figure 1: MMA Optical Configuration, Version 1.

Choice of Cassegrain Geometry

The proposed geometry is a minimum-blockage Cassegrain geometry. This choice is driven primarily by the desire for low antenna noise (to match the low atmospheric opacities available on the 5000m site in Chile) and for high aperture efficiency. The minimum blockage aspects of the design include supporting the quadripod legs near the edge of the reflector and making the diameter of the subreflector equal to the diameter of the hole in the center of the primary reflector. The use of the Cassegrain focus avoids additional reflectors and their unavoidable contributions to the system temperature. At the Cassegrain focus it is possible to have more than one feed at a time looking at the subreflector. This is necessary, for example, if one wishes to monitor the atmospheric total power fluctuations at one frequency in order to correct for atmospheric phase fluctuations at another frequency.

Choice Of Primary Mirror Focal Ratio

The choice of 0.38 for the primary F/D is driven by the trade off between close packing (short interferometer baseline without the possiblility of antenna collisions) and sensitivity to mis-alignment. For the geometry described, the antennas can be placed on a baseline of length 1.28 D without possibility of collision at elevation angles greater than 18.5 degrees. At zero elevation angle the baseline without possibility of collision is 1.32 D. Other geometries being studied would allow even closer packing.

Possible Feed Arrangements and Secondary Mirror Magnification

The location of the cassegrain focus is chosen to lie 1.5 m below the vertex of the primary to allow plenty of space in front of the receiver for the various selectable quasi-optical devices that have been proposed. Examples of these devices include calibration devices, circular polarizers, solar observing devices, dual frequency reflectors and possibly a beam directing reflector for the 30 GHz receiver which could be mounted off to the side of the mm/submm receiver. The receivers are located near the elevation axis so that antenna balance is not substantially affected by their removal. Designs with better antenna close packing performance either have less space available in front of the receivers or locate the receivers aft of the elevation axis.

Several feed systems can be considered to provide the ten receiver bands proposed over the 30 GHz to 950 GHz range proposed for the MMA (Wootten et al, 1998). Four possible systems and some of their advantages and disadvantages were discussed in memo 163. They are a) off-axis feeds with a symmetrical subreflector, b) off-axis feeds with a rotating asymmetric subreflector, c) movable dewar(s) which place the selected feed on axis, or d) a rotating, cooled on-axis beam director.

For the MMA receiver and antenna designs we have tentatively chosen option (a), off-axis feeds with a symmetrical subreflector. The following discussion summarizes the optical aberrations with such a configuration.

Aberrations

The optical configuration described here is the classical cassegrain type. The main parameters are summarized in Figure 1. The classical cassegrain configuration has zero spherical aberration but suffers from both coma and astigmatism off axis. The relationship between the angular aberrations given by these formulas and the wavefront rms deviation are derived in Lugten (1998).

For a feed located 0.152 m (6 inches) off axis, the angular tangential coma (ATC) is 1.89" and the angular astigmatism (AAS) is 0.98" . This results in wavefront errors of 1.80 microns RMS and 2.42 microns RMS, respectively, assuming uniform illumination. These wavefront errors result in a loss of on axis gain of 0.10% and 0.19%, respectively at 850 GHz (353 microns wavelength). With tapered illumination, the RMS wavefront errors are slightly smaller. The radius of curvature of the median image surface is 0.290 m (11.4 inches), so that for a feed located 0.152 m off axis it is approximately 0.043 m (1.7 inches) closer to the secondary mirror than for a feed located on axis. If receiver feeds were to be configured for a flat median image surface and the subreflector moved to bring the selected feed into focus, a subreflector motion of 136 microns (0.0054 inches) is required. This motion would introduce 8.4" of angular spherical aberration (ASA), which corresponds to 3.79 microns RMS wavefront error and would result in a 0.45% loss of gain at 850 GHz.

Misalignment of the telescope also produces aberrations. Decenter of the subreflector by 39 microns produces a wavefront tilt (pointing error) of 2" and ATC of 2.75" , and negligible AAS. Likewise, tilt of the subreflector by 13.5" produces a wavefront tilt of 2" and ATC of 1.30" and negligible AAS. The aberrations resulting from tilt and decenter of the subreflector are strongly coupled - for example, a combination of tilt and decenter equivalent to a rotation of the subreflector about the prime focus location produces nearly perfect cancellation of aberrations. Finally, as noted above, despace of the mirrors by 136 microns produces 8.4" of angular spherical aberration (ASA) resulting in a loss of forward gain of 0.45% at 850 GHz.

For the classical cassegrain configuration, ATC varies as the inverse square of the final focal ratio. AAS varies approximately inversely with the primary mirror focal ratio as does the curvature of the image surface. Thus, choosing a faster primary mirror increases AAS and increases the image surface curvature. Likewise, ATC due to secondary decenter or tilt is worse for a faster primary mirror, and ASA is more sensitive to despace errors for a faster primary mirror.

A Possible Arrangement of Receiver Feeds

Many decisions on the receiver layout have not been made; in particular, the question of whether the receivers themselves will be built into one, or two or more, dewars is still unresolved. The choice of dewar configuration is a function of cooling capacity, refrigerator running costs and reliability, and general reliability questions. It is now likely that any eventual 30-GHz system will be built as an add-on, probably with additional mirrors, rather than being placed in the main dewar or dewars.

MMA Memo #183 discusses one possible configuration of receivers. The following figure is taken from that memo, and illustrates one arrangement to be considered if the receivers are built into 2 dewars. Although it is probably unlikely that exactly this arrangement will be followed, it gives some idea of possible options. The choice of receiver is made operationally simply by changing the pointing of the telescope, so that the correct receiver window at the top of the dewar is illuminated.

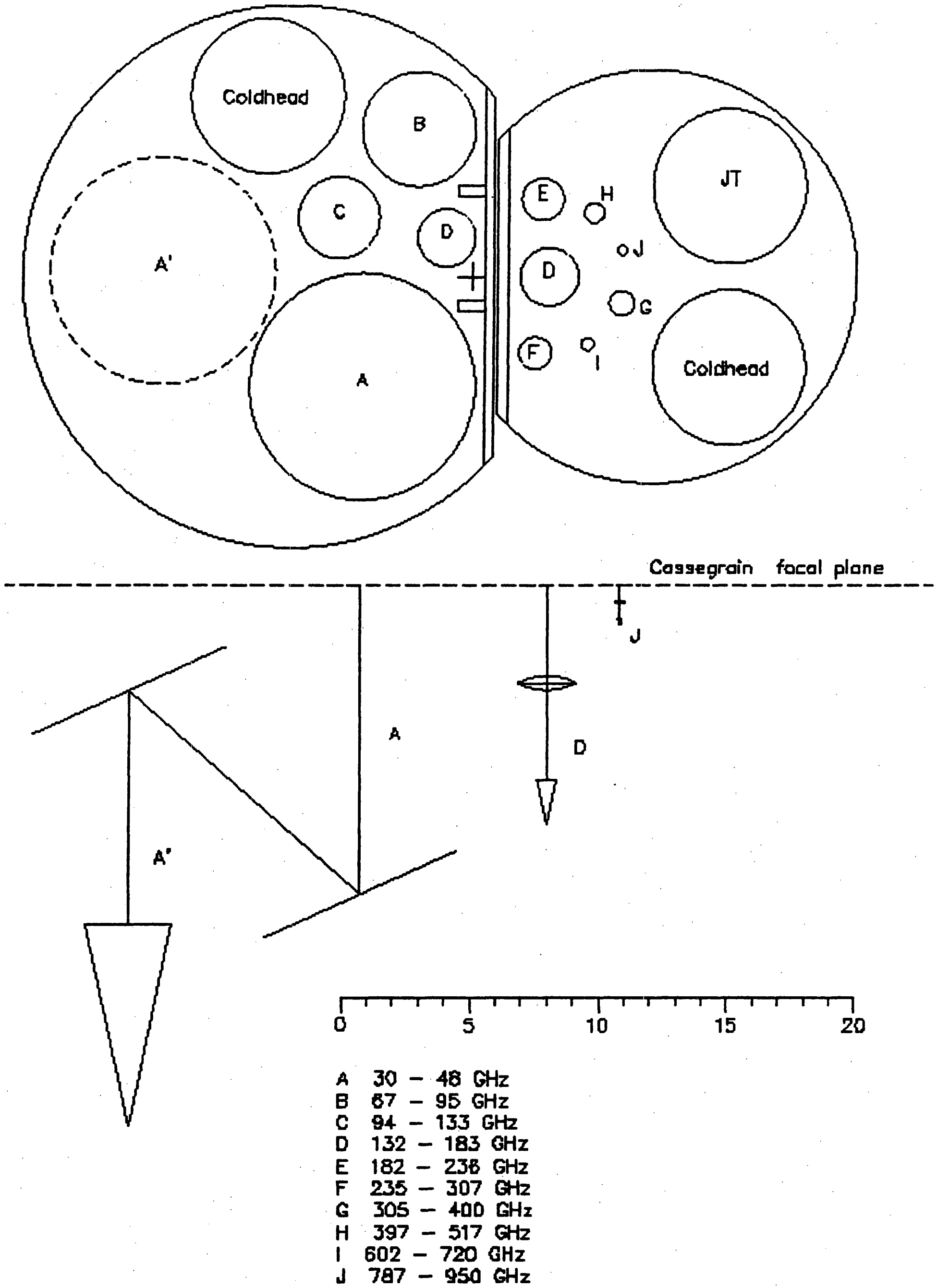


Figure 2: A possible receiver feed configuration, for the 2-dewar case.
See MMA Memo #183

D	Primary Aperture	10.01 m	394.00 inches
	Focal length of primary	3.803 m	149.72 inches
	f_p of primary	0.38	0.38
d	Secondary aperture	0.762 m	30.00 inches
	Final f/D	6.78	6.78
	Magnification factor	17.85	17.85
f_p/D	Primary angle of illumination	133.36 deg	133.36 deg
$theta_p$	Secondary angle of illumination	8.44 deg	8.44 deg
2c	Distance between primary and secondary foci	5.327 m	209.72 inches
H	Depth of primary	1.646 m	64.803 inches
r1	Tipping structure apex radius (1" clearance)	6.624 m	260.8 inches
	[close packing 1.32 D]		
r2	Tipping structure dish radius (1" clearance)	6.388 m	251.5 inches
	[close packing 1.28 D]		
h	Distance from vertex to secondary focus	1.524 m	60.00 inches
a	Distance from elevation axis to focus	0.762 m	30.00 inches
g	Distance from primary focus to top of quadrapod	0.508 m	20.0 inches
x	Clear aperture at receiver cabin window	0.660 m	26.0 inches

Table 1: Dimensions of MMA Optical Configuration, Version 1.

References

Faber, Sandra M., ''Formulae Relevant to the Optical Performance of Telescopes'', U.C. TMT Report No. 55 (1981)

Lugten, John B., ''Relationships between Angular Aberratiions and the Wavefront Deviation from Flatness'', MMA memo in prep. (1998)

Napier, P. et al., ''A Strawman Optics Layout for the MMA Antenna'', MMA memo 163 (1996)

Schroeder, Daniel J., ''Astronomical Optics'', Academic Press, San Diego (1987)

Wootten, A., Snyder, L., van Dishoeck, E., and Owen, F., ''Frequency Band Considerations and Recommendations'', MMA White paper (1998)

Back to Chapter 5 TOC | Back to Chapter 5 Section 1 | On to Chapter 5 Section 3

MMA RECEIVERS: SIS mixers

Tony Kerr
S.-K. Pan
John Webber
Last revised July 10 1998

3. SIS Mixers

1. Specifications

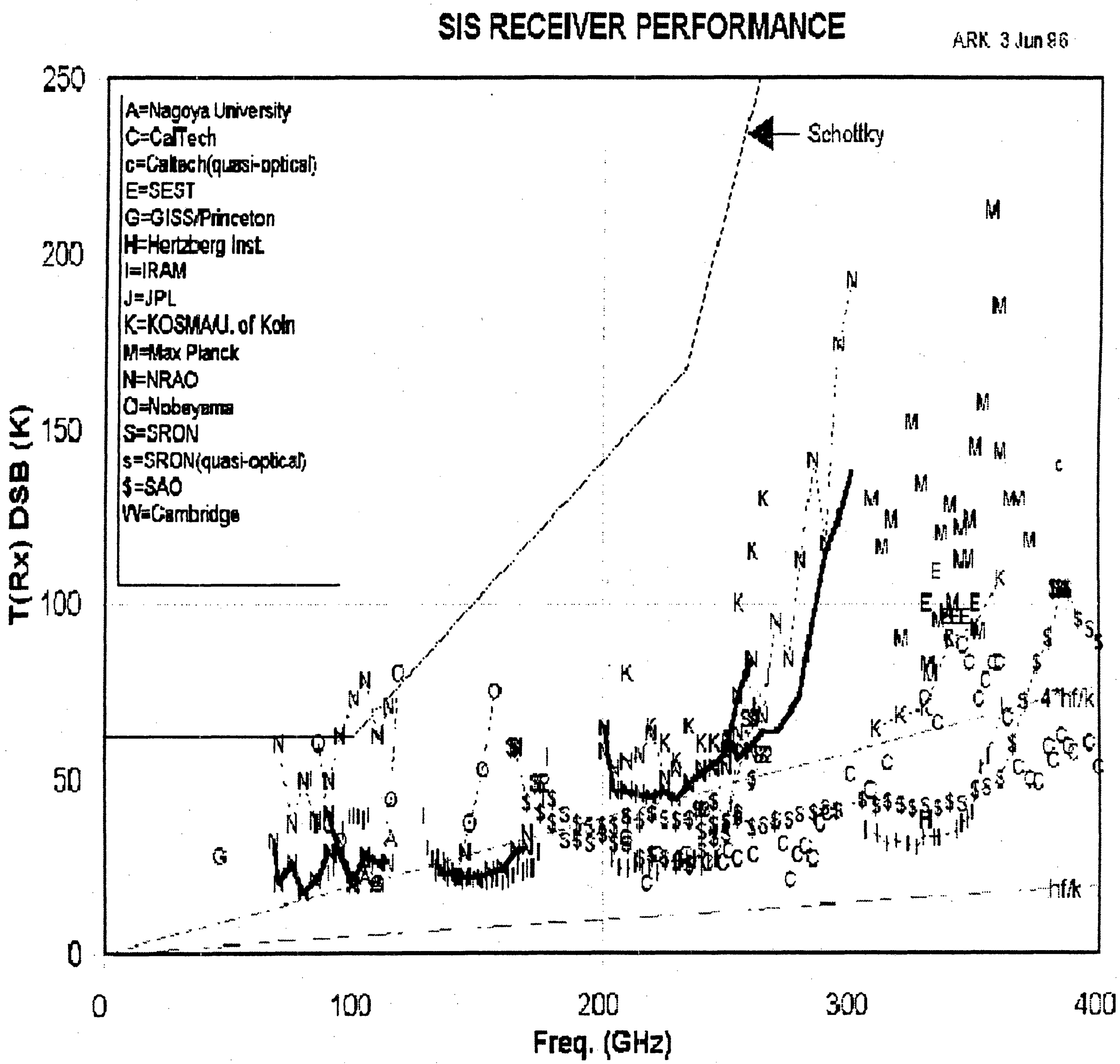
SIS mixer receivers are expected to be used for all frequencies above 116 GHz. It is undecided whether SIS receivers will be used below this frequency (perhaps down to 90 GHz), or whether HFET amplifiers will be preferable for their greater immunity to interference and possible lower cryogenics cost. All MMA receivers must meet the following general requirements:

i. for spectral line observations, the SSB system noise must be as low as possible;
ii. for sensitive continuum observations, the IF bandwidth of each receiver should be 8 GHz, and
iii. for the sake of reliability there should be no mechanical tuners.

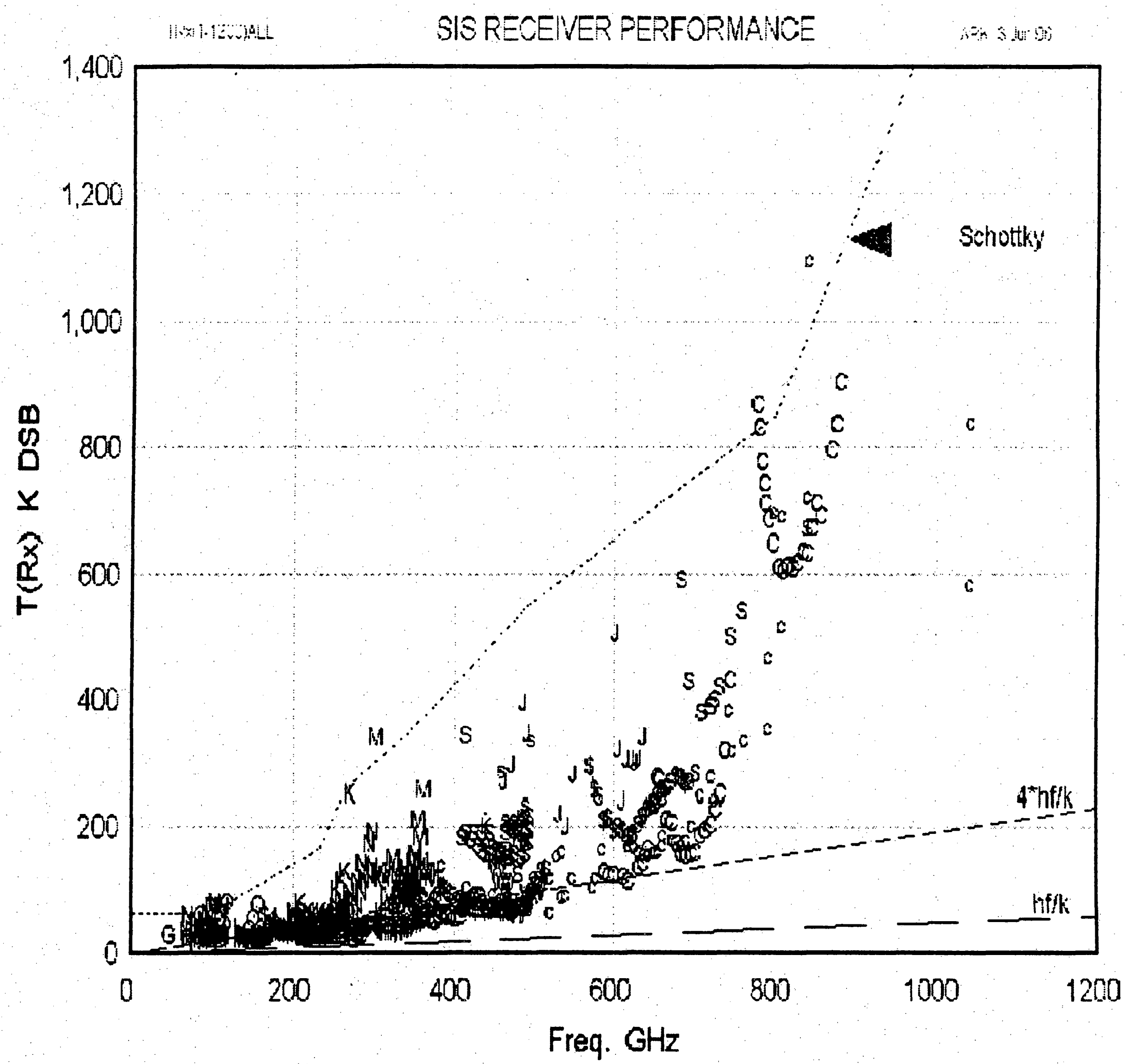
2. Performance

Figure 5.3.1 shows the DSB noise temperatures of SIS receivers reported in the last few years. The best fixed-tuned receivers have DSB noise temperatures in the range 2-4 hf/k up to ~600 GHz. Above ~600 GHz, receiver noise temperatures rise rapidly because of RF loss i the Nb conductors. Work on new materials is likely to improve high frequency results in the next few years (e.g., NbTiN for 600-1200 GHz).
(Note that in calculating SSB system noise temperatures from DSB receiver noise temperatures, care must be taken to include the appropriate image input noise. The appropriate value of SSB receiver noise temperature is given by:

$$\mathbf{TR_{SSB} = 2TR_{DSB} + T_{image}}.)$$



Click on image to zoom



Click on image to zoom

Figure 5.3.1 Reported SIS mixer receiver temperatures

Most of these receivers use a ~1.5 GHz IF, an exception being the SAO receivers which use 4-6 GHz. The IF for the MMA is tentatively chosen as 4- 12 GHz to give the desired 8 GHz IF bandwidth. The final choice of IF will depend largely on the results of work now under way to develop an internal IF stage for SIS mixers which will allow isolators to be eliminated from the IF system. The best (individual) tunerless SIS receivers reported to date in the 150-400 GHz range have frequency ranges 1.37:1, 1.42:1, and 1.54:1. Their noise temperatures degrade quite precipitously beyond the band edges. In making the 80 receivers required for each band on the MMA, we cannot expect to achieve identical Tr vs. freq. characteristics, and the maximum bandwidth common to all 80 receivers will be somewhat less than that of the individual receivers. (Nb process control is something we are starting to work on with our SIS fabricators, but hitherto there has been little consideration given to such matters in SIS mixer production). It is hoped that by the time we start building the MMA receivers we will be able to achieve a 1.5:1 common bandwidth, but until this is actually demonstrated we should be conservative to ensure we do not end up with unexpected gaps in the frequency coverage.

3. Development

1. Capacitively coupled coplanar waveguide

To achieve wide RF bands (an upper to lower frequency ratio of 1.3 or greater) without mechanical tuning, a fully integrated (MMIC) mixer design is required. The resulting ''drop in'' mixer chips are relatively easy to mount in blocks in which they are coupled to RF and LO waveguides. Conventional microstrip MMIC technology is difficult to use above ~100 GHz because of the very thin substrates necessary to prevent coupling to unwanted substrate modes. The use of coplanar waveguide (CPW) circuits allows a thick substrate, but is prone to odd-mode resonances excited by bends or near-by obstacles, and has poor isolation between adjacent lines. CPW also requires inconveniently narrow gaps when a substrate of low dielectric constant is used. To overcome these difficulties, we have developed capacitively loaded coplanar waveguide (CLCPW), a CPW with periodic capacitive bridges. The bridges are grounded at the ends, thus suppressing the odd mode, but they also add a substantial capacitance per unit length to the CPW, which allows desirable characteristic impedance levels to be obtained with convenient dimensions. Figure 5.3.2 shows a quadrature hybrid composed of CLCPW with periodic capacitive bridges.

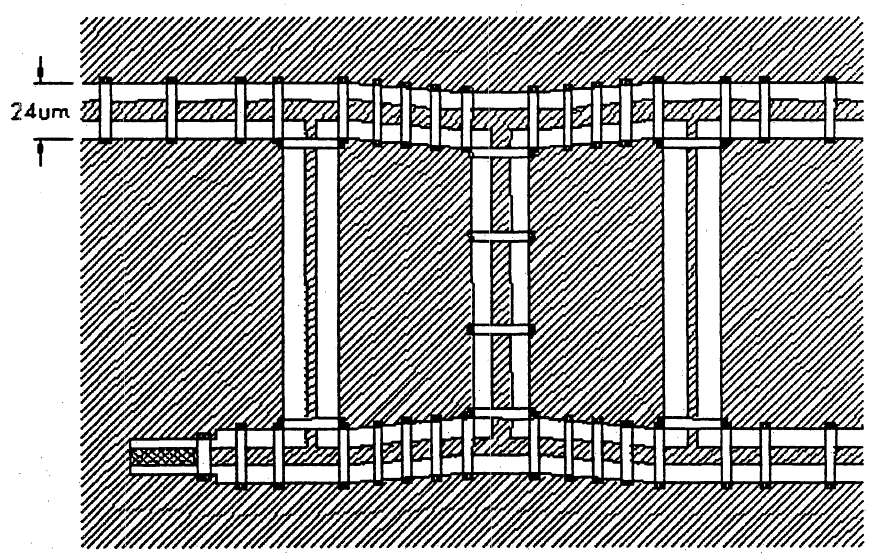


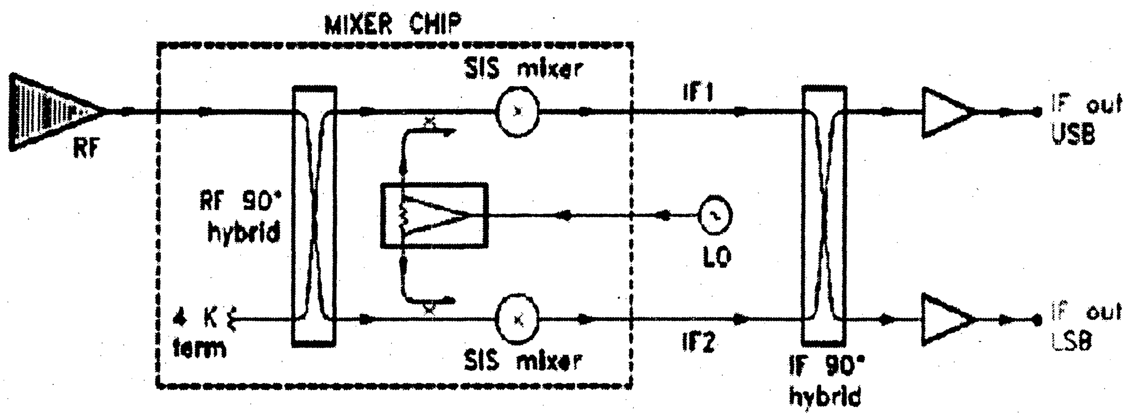
Figure 5.3.2. A 200-300 GHz quadrature hybrid using capacitively loaded coplanar waveguide (CLCPW).

The bridges are 4 microns wide, and are connected to the ground plane at their ends.

The fourth port (lower left) has a built-in matched termination. The substrate is 0.0035" fused quartz.

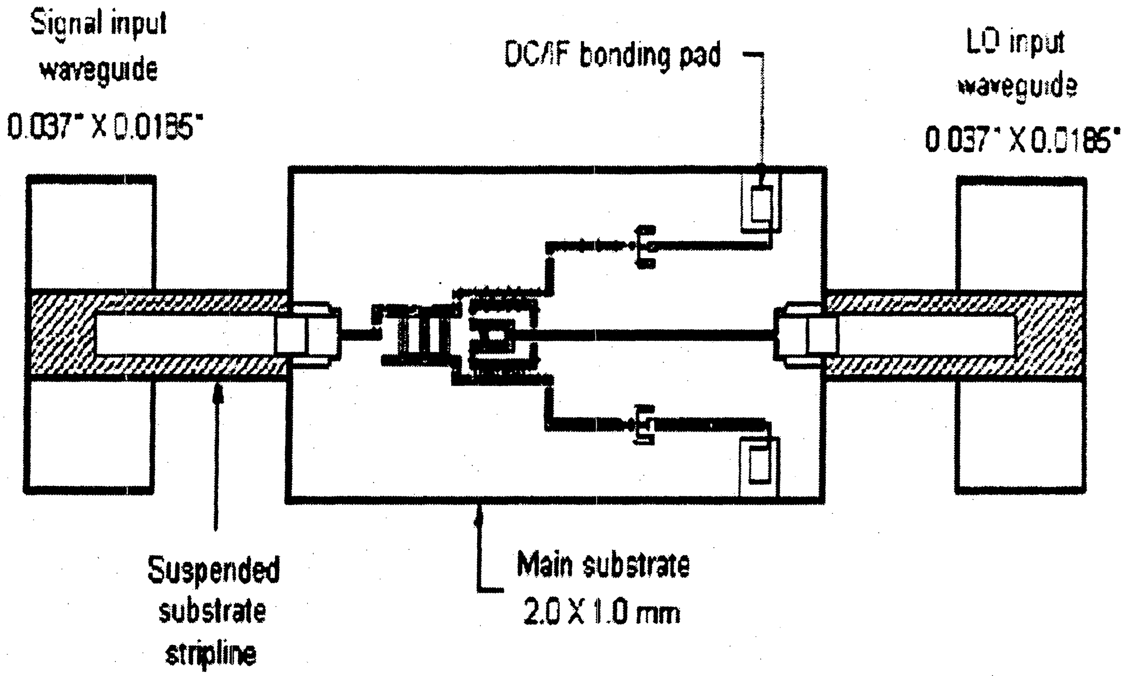
2. **Sideband separating mixer**

Even at the proposed site in Chile with its low atmospheric water vapor, atmospheric noise in the image band of an SIS receiver will add substantially to the system noise. The advantages of sideband separating mixers with their image terminated in a 4 K cold load have been discussed (see MMA Memos 168 and 170), and we expect to use sideband separating mixers in at least the lower frequency SIS receivers. A developmental MMIC 230 GHz sideband separating mixer is shown in Figure 5.3.3. The IF outputs from the mixer are combined in an external quadrature hybrid which phases the down-converted signals from the upper and lower sidebands so they appear separately at the output ports of the hybrid. A useful property of the sideband separating SIS mixer is that the sidebands can be swapped between the two outputs simply by reversing the polarity of the bias on one of the component mixers.



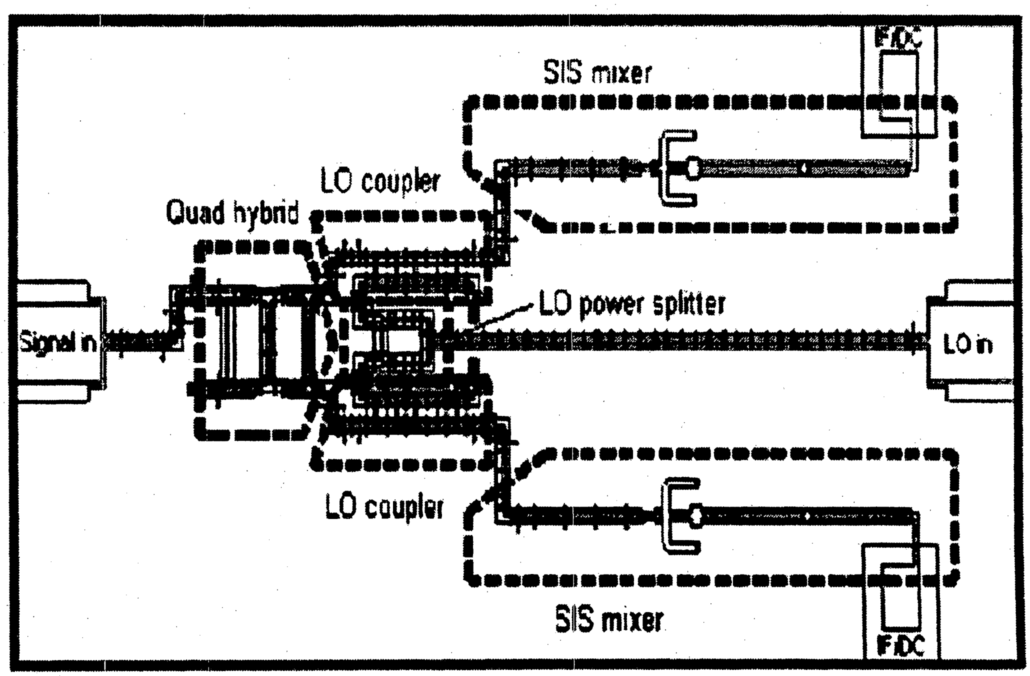
Click to zoom

Figure 5.3.3(a). Block diagram of an SIS sideband separating mixer.



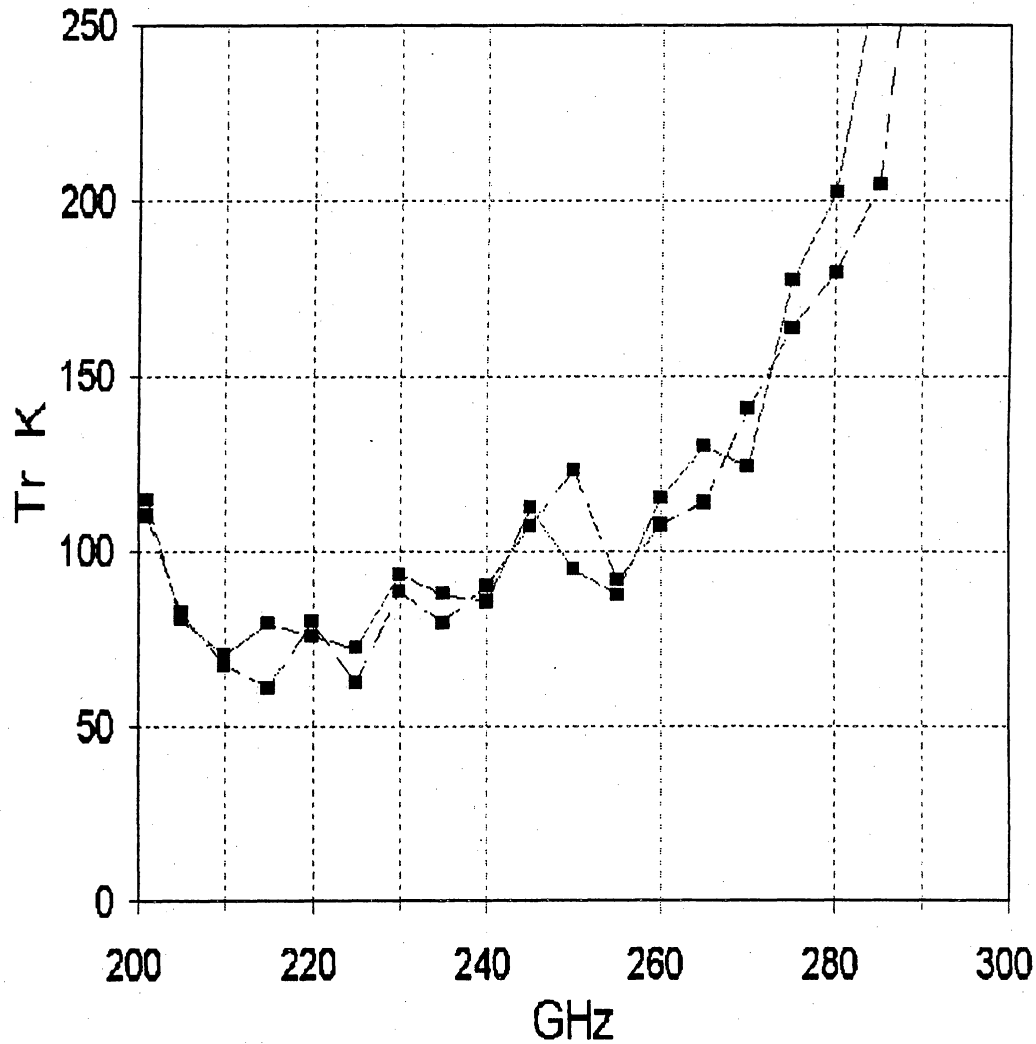
Click to zoom

Figure 5.3.3(b). 230-GHz sideband separating mixer, showing the signal and LO waveguides, suspended stripline coupling probes, and the main substrate.



Click to zoom

Figure 5.3.3(c). Substrate of the 230-GHz sideband separating mixer, showing the main components.



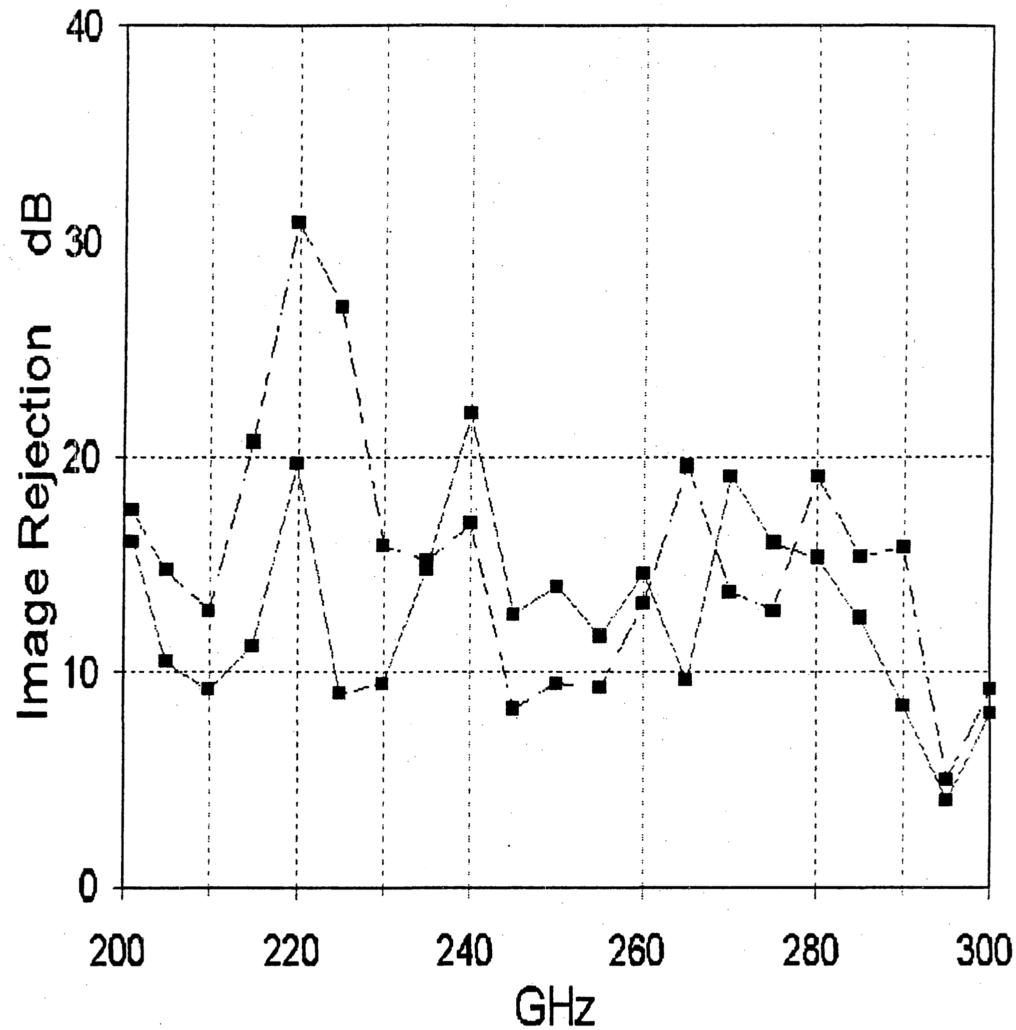
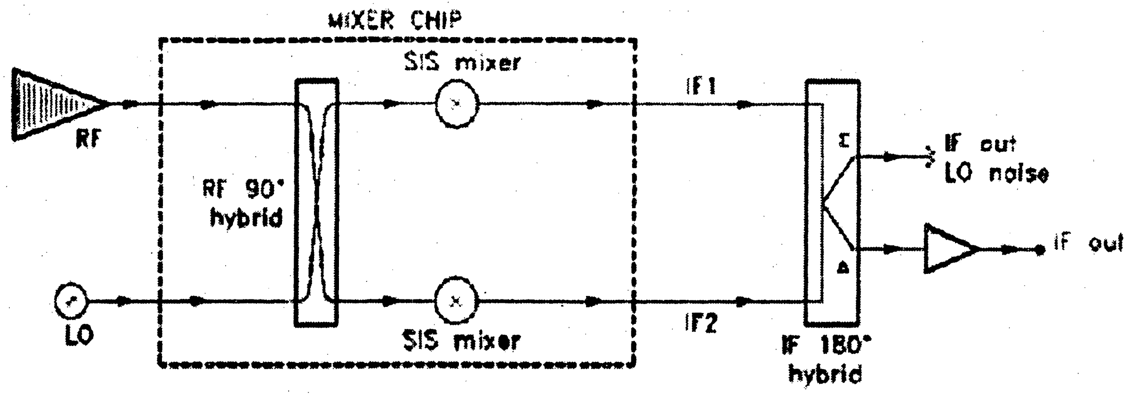


Figure 5.3.3(d) Receiver temperature and sideband separation for the experimental mixer.

3. **Balanced mixer**

The use of balanced SIS mixers has two potential advantages for the MMA. Compared with the usual 20 dB LO coupler or beam splitter in front of the mixer, a balanced mixer requires ~17 dB less LO power. This greatly eases the task of developing wideband tunerless LOs. The other benefit of a balanced mixer is its inherent rejection of AM sideband noise accompanying the LO. A MMIC balanced mixer design is shown in Figure 5.3.4. Some of these mixers have been fabricated, and tests to verify the design are expected to continue through 1998.



Click to zoom

Figure 5.3.4(a). Block diagram of a balanced SIS mixer.

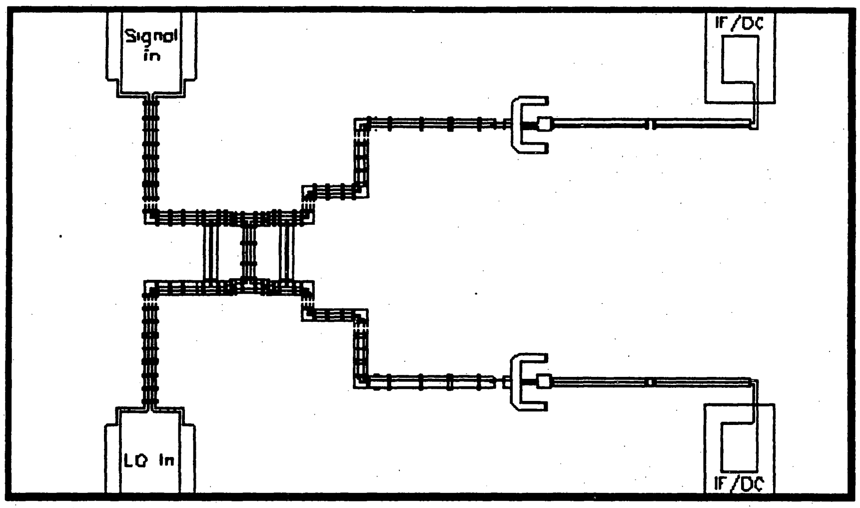
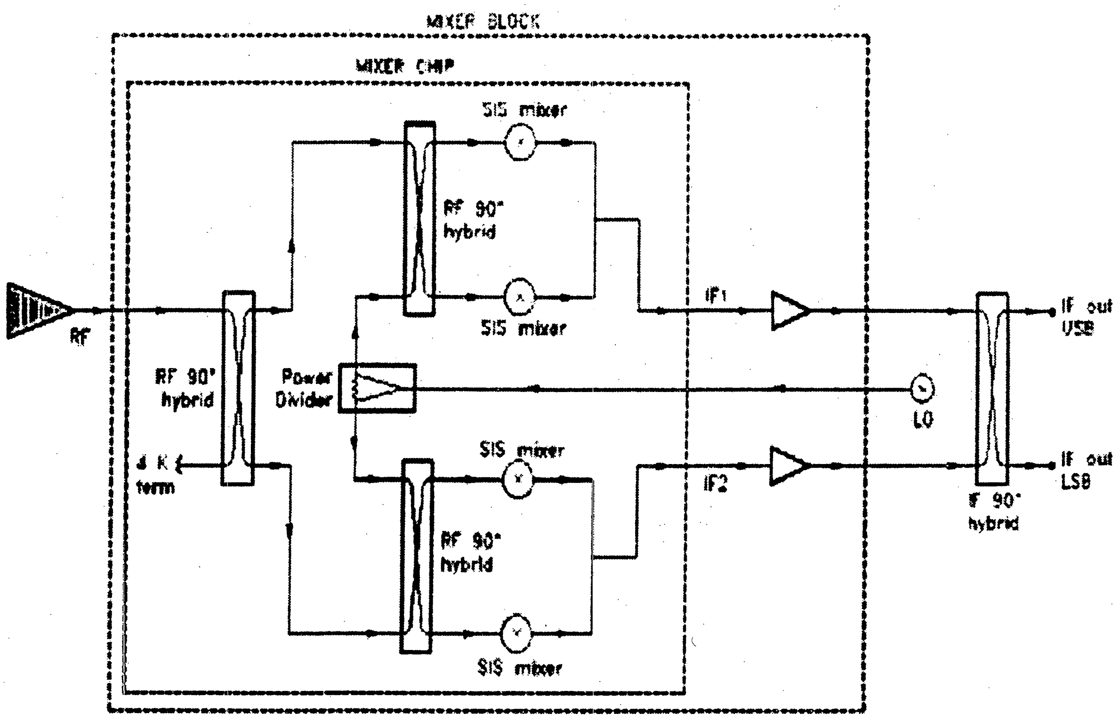


Figure 5.3.4(b). Substrate of a 230 GHz balanced mixer, showing the quadrature hybrid and two SIS mixers.

4. **Sideband-separating, balanced mixers**

When the designs of the sideband-separating and balanced mixers have been thoroughly tested, we will design and build a mixer which incorporates both these features: a balanced, sideband-separating mixer. This will incorporate the circuit elements whose design has already been proven individually. This will produce for the MMA a mixer which requires a minimum of LO power, provides good immunity to LO noise, and substantially reduces the contribution to system noise of atmospheric noise in the unwanted sideband. The schematic is shown in Figure 5.3.5. We expect that the mixer chip will be about 2 X 2 mm in size.



Click to zoom

Figure 5.3.5. Block diagram of a balanced, sideband-separating SIS mixer.

5. **Integrated IF amplifier**

Two options are being considered for the 8-GHz-wide IF in the SIS receivers for the MMA. The conventional approach uses an IF isolator between the mixer and IF amplifier, while a new scheme, developed at OVRO, uses an IF amplifier stage inside the SIS mixer block and no isolator. The latter scheme allows an IF covering more than an octave, tentatively 4-12 GHz. The need for an isolator in the conventional scheme forces the IF center frequency to at least 12 GHz (IF = 8-16 GHz) to achieve an 8 GHz bandwidth, probably with a significant noise penalty. The penalty is not simply a result of the increase in amplifier noise temperature at the higher frequency, but includes the noise from the cold termination of the isolator which is reflected from the mixer output.

The use of a high IF, as required by both the above schemes, imposes a constraint on the output capacitance of the SIS mixer. In most SIS mixers, the RF tuning circuit adds substantial IF capacitance in parallel with the SIS junction. We have developed an SIS mixer with low IF capacitance, and this design was used as a building block in the sideband separating and balanced mixers described above.

In collaboration with S. Weinreb at the University of Massachusetts, we have begun the design of an integrated IF amplifier which will permit the MMA goal of 8 GHz instantaneous bandwidth per sideband to be realized. The MMIC amplifier chip employs a grounded gate first stage and is expected to give good performance over the

4-12 GHz target band.

6. **Further plans**

It is planned to continue the development of the 200-300 GHz sideband- separating, balanced SIS mixer with integrated IF amplifier until the goals of 5.3.1 are met. This will then be used in the prototype receiver for the test antenna. Once a number of these mixers have been evaluated, the frequency bands for the MMA will be frozen and new designs will be developed to cover those bands.

Back to Chapter 5 TOC | Back to Chapter 5 Section 2 | On to Chapter 5 Section 4

MMA RECEIVERS: HFET AMPLIFIERS

Marian Pospieszalski
Ed Wollack
John Webber
Last revised July 14 1998

4. HFET AMPLIFIERS

1. Specifications

HFET low-noise amplifiers [1] may be used for MMA front ends between 26.5 and 116 GHz, depending on final decisions regarding the lowest frequency at which the MMA must operate and the highest frequency at which amplifiers are preferable to SIS mixers. In any event, the HFET amplifiers will be designed to produce the lowest noise compatible with the receiver band (up to a whole waveguide band), with sufficient gain to overcome second stage noise. It should be noted that HFET amplifiers have greater immunity to in-band interference than SIS mixers, a feature which may be important in deciding between the two below 116 GHz due to the possibility of interference from satellite-borne transmitters in this region of the spectrum. In addition, HFET amplifiers require cooling only to 15 K instead of 4 K; this may be relevant or not, depending on the dewar configuration (see section 6).

Possible bands to be covered, along with characteristic performance data for InP amplifiers, are listed in Table 5.4.1. Note that the first two bands are required if the MMA must operate below 33 GHz, thus requiring 2 receivers to cover 26.5-50 GHz; but if the lowest frequency is chosen to be 33 GHz, then the entire band 33-50 GHz can be covered with just one receiver. If standard waveguide is used, then in order to provide complete coverage above 68 GHz (the frequency at which the atmospheric oxygen line begins to permit observations) we must stop at 90 GHz for this receiver. If non-standard waveguide is used, then the region 68-102 GHz could be covered. Thus, it is possible to cover all atmospheric frequencies from 33 to 102 GHz with just two HFET receivers; an examination of Table 5.4.1 will also reveal that the receiver temperatures at some frequencies would be slightly higher in this case than if narrower-band amplifiers are used.

Table 5.4.1		
Band (GHz)	**Waveguide Designation**	**Trcvr (K) at Center of Band**
26.5-40	WR-28	16
40-50	WR-22 or 19	25
33-50	WR-22	25
68-102	WR-10 or non-standard	50
68-90	WR-12	40
90-116	WR-08	60

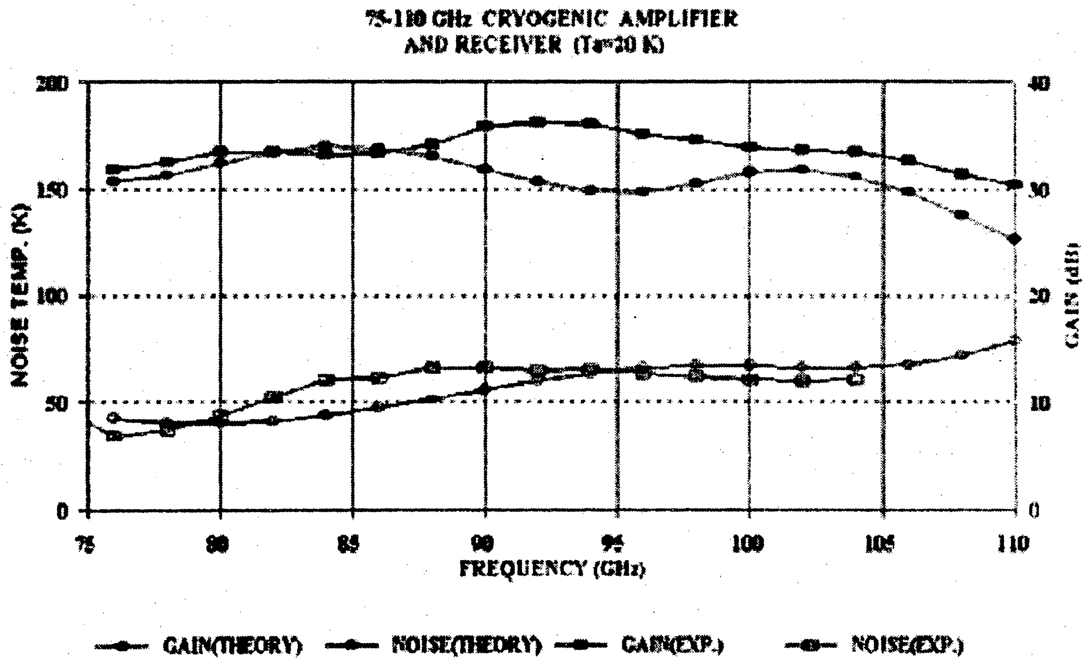
Low-noise HFET amplifiers have a higher threshold for problems caused by interference than do SIS mixers, which will be important if a proposed 94 GHz satellite-borne cloud radar is implemented.

2. **Development work**

The InP HFETs used in current NRAO amplifiers consist of a 250-nm undoped AlInAs buffer, a 40-nm GaInAs channel, a 1.5-nm undoped spacer, an 8-nm AlInAs donor layer doped to approximately 7e18 cm-3 and, finally, a 7-nm GaInAs doped cap, all grown lattice-matched to an InP semi-insulating substrate. These devices typically exhibit an electron sheet density of 2.5 to 2.8 $.10^{12}$ cm^{-2} and a room-temperature mobility of 10,000 to 11,000 cm^2/Volt-sec.

Successful design of cryogenically coolable amplifiers requires knowledge of both signal and noise models of HFETs at cryogenic temperatures. These models have been developed with sufficient accuracy to allow for computer-aided designs of cryogenic amplifiers with optimal noise bandwidth performance. An example of a cryogenic amplifier covering WR10 waveguide bandwidth (75-110 GHz) is shown in Figure 5.4.1, with calculated and measured performance data shown in Figure 5.4.2. This amplifier, built for the Microwave Anisotropy Probe (MAP) project, was realized in hybrid technology using pure polytetrafluoroethylene (PTFE), 0.003" thick substrates. The choice of "chip and wire" technology was dictated not only by the objective of achieving the lowest possible noise performance, but also by much lower cost than full-blown development of MMICs, in relatively low volume radio astronomy applications. The amplifier uses full waveguide bandwidth, E-plane probe, waveguide-to-microstrip transitions. It employs six InP HFET chips, each having gate dimensions 0.1 x 50 microns. The input and first two interstage coupling networks were designed to achieve the lowest average noise temperature across the band while the last three interstage coupling networks were designed to achieve a flat gain across the band. The bias networks also use "chip and wire" technology with bond wires being treated as transmission lines in the design process. All the bias network elements having influence on millimeter-wave characteristics and coupling capacitors are manufactured using 0.003" thick quartz substrates.

Figure 5.4.1 Six-stage 75-110 GHz amplifier



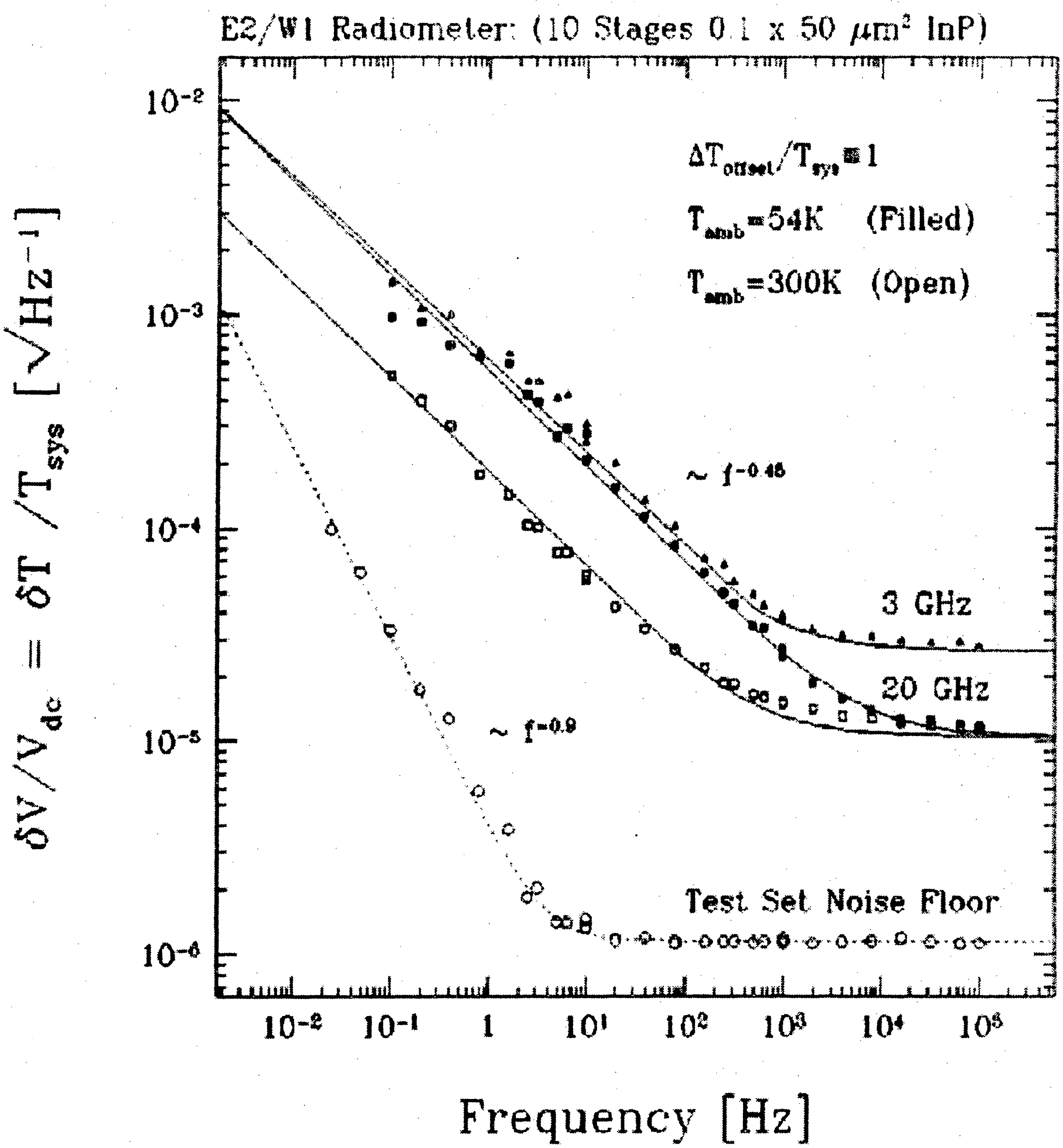
Click to zoom

Figure 5.4.2 Theoretical and experimental gain and noise of a laboratory receiver employing the amplifier shown in figure 5.4.1

An amplifier can be designed to provide good performance in any of the bands listed in Table 5.4.1. The only region of the spectrum in which an NRAO amplifier has not already been designed and tested is 110-116 GHz. In this highest frequency range appropriate for use of an HFET amplifier, performance becomes critically dependent on device geometry. The dimensions of the presently available devices from the NRAO wafer designated "518" are 10 by 12 mils, which is marginally too large for a robust design reaching 116 GHz. A wafer procured for the MAP project has somewhat smaller devices, 6 by 8 mils, but the surface passivating layer on this wafer increases capacitance and precludes its use above about 100 GHz. In order to make a design reaching 116 GHz, and to have enough devices to build all the HFET amplifiers for the MMA with the best possible noise temperature, the plan calls for fabrication of a new wafer with devices similar to the MAP wafer, but without the surface passivating layer.

3. Gain Fluctuations

A feature of the transistors is the gain fluctuation commonly called "1/f noise" which can be significant in single-dish astronomical observations with very large detection bandwidth X integration time products. Since the MMA will be used in this mode to make large-scale maps, probably on-the-fly, the effect of gain fluctuations must be considered. Theoretical calculations and laboratory experiments have been reported by Wollack [2] and by Wollack and Pospieszalski [3]. Using direct total power detection in a receiver at 80 GHz using 10 stages of InP amplification at ambient temperatures of 54 K and 300 K yielded the results shown in Figure 5.4.3.



Click to zoom

Figure 5.4.3 Fluctuations in power spectral density under varying temperature and bandwidth conditions for a test total power radiometer

For the narrow bandwidths used in spectral line observations, these gain fluctuations are not an important factor, since the bandwidth per spectral channel will be small and the noise will be dominated by pure noise, Tsys/sqrt(time*bandwidth). For wideband continuum observations, however, gain fluctuations will be the dominant factor. As a concrete example, consider a time scale of 1 second, which might be typical of a single unidirectional scan employed for observing a continuum point source or making a continuum map. The square of the variance in gain at 20 K ambient temperature for InP devices is about $3.6 \cdot 10^{-8}$ [1/Hz] per stage; conservatively, 6 stages of gain will be needed to obtain at least 30 dB of gain before a mixer. With 6 stages and a bandwidth of 8 GHz, a total power radiometer will have $\mathrm{sqrt}(3.6 \cdot 10^{-8} * 6 * 8 \cdot 10^{9} * 0.5) = 29$ times the variation in measured (delta T)/T than predicted from pure noise.

Various techniques can be used to alleviate this problem and provide continuum sensitivity closer to the situation in which pure noise, rather than gain fluctuations, dominates. The situation is analogous to the early days of radio astronomy, when bandwidths were much smaller but amplifier gain fluctuations much larger; the solution was a switching radiometer. In order to achieve full theoretical sensitivity with an 8 GHz bandwidth, a switch rate of about 10 kHz would be required; the problem is, how to switch and what to switch against. In the MAP project, a phase-switching scheme is used in effect to switch between two beams at a rate of 2.5 kHz; this results at W-band in only about a 20% loss in sensitivity for the effective 20 GHz bandwidth. A sky-switching option for the MMA would significantly increase expense, since a 2-feed receiver with all its electronics would cost at least twice as much as a single feed receiver; there might be severe difficulty going as low as 33 GHz with such a plan, due to the size of the feeds. Experiments in measuring and alleviating gain fluctuations using an out-of-band "pilot tone" have been carried out in the laboratory, with only modest success (e.g. Weinreb [4]). It is possible that a noise-adding radiometer could be employed, at least up to 50 GHz; GaAs noise diodes which can be switched on and off at least as fast as 1 kHz are available which are more stable than amplifiers in this frequency range (Wollack, private communication), but the situation at W-band is less certain. In a total power radiometer scheme, there is some advantage to performing detection at the 2 GHz stage of the IF conversion process, since both gain and noise fluctuations are somewhat decorrelated from one part of the band to another; but this advantage may be reduced by the fact that the separation into 2 GHz bands comes after fiber optic transmission and other electronic processing, which may induce gain fluctuations which dominate the noise. Some reduction in gain fluctuation may also be achieved by running all but the early stages of amplification warm, which improves gain fluctuations by a factor of ~6, and by using GaAs devices in following stages where possible, since the gain fluctuations of GaAs devices are about 1/3 that of InP devices.

Finally, we note that the SIS mixer receivers will also be affected by gain fluctuations in the wideband HFET amplifiers employed in the 4-12 GHz region. These amplifiers will use gate widths of about 400 microns, in contrast to the 50 micron gate widths used at 100 GHz, and the gain fluctuations will be correspondingly reduced by about sqrt(400/50) = 3. Following stages of warm or GaAs amplification can also help. If we can reduce gain fluctuations sufficiently, the atmospheric variations in transparency and noise will dominate the radiometer noise; but this may be difficult to achieve in practice.

REFERENCES

1. Pospieszalski, M.W. et al., "Millimeter Wave Waveguide Bandwidth Cryogenically-Coolable InP HEMT Amplifiers," June 1997, IEEE MTT-S Microwave Symp. Digest, Denver, CO, pp. 1295-1288.
2. Wollack, E.J., "High-Electron-Mobility Transistor Gain Stability and its Design Implications for Wide Band Millimeter Wave Receivers," 1995, RSI,

vol.66, no.8, pp. 4305-4312.
3. Wollack, E.J. and Pospieszalski, M.W., "Characteristics of Broadband InP Millimeter-Wave Amplifiers for Radiometry," June 1998, IEEE Int. Microwave Symp., Baltimore, MD.
4. Weinreb, S., "VHF Pilot Signal Stabilized Monolithic Integrated Circuit Millimeter-Wave Radiometers," Prog. 1997 URSI North American Radio Science Meeting, p. 644.

Back to Chapter 5 TOC | Back to Chapter 5 Section 3

CRYOGENICS

L. D'Addario
Last revised: July 10 1998

See also:

MMA Memo No. 184, "Cryogenics options for the MMA," 98/09/20. General background information.

MMA Memo No. 185, "Review of Gifford-McMahon refrigerators for 4K," 97/09/25.

1. OVERVIEW

The MMA cryogenics design is presently very undetermined because, first, the cooling requirements of the receivers are not yet accurately known; and, second, because we intend to use the MMA development phase to achieve substantial improvements relative to the cryocoolers that have heretofore been conventional in radio astronomy. The large number of antennas and the remote location of the MMA imply that two kinds of improvement are highly desirable: power consumption and reliability.

Existing NRAO systems capable of cooling 1W to 4K use about 9.7kW from the power line; this is an efficiency of about .01%, compared to the ideal theoretical (Carnot) efficiency of about 1.2% and practical efficiencies achieved in other (much larger) systems of up to 0.2% (20x better). An improvement by a factor of 2 to 4 (to .02% to .04%) should be feasible in our size range, as described below. In addition, careful design to minimize the cooling requirements of the receivers should keep the 4K load below 0.5W and the loads at other stages below 10W.

Most cryocoolers in radio astronomy now use Gifford-McMahon refrigerators. These are highly developed commercially and are fairly reliable, but they nevertheless have parts that wear out and require overhauls at intervals of 12 to 18 months. In addition, random failures occur from leaks in joints and flexible hoses, from helium contamination, and in compressors. As a result, use of the traditional technology can be expected to result in about 2 shutdowns per year per cryocooler for maintenance or repair. Even with one system per antenna, this is more than one per week. Technologies now exist (for compressors, expanders, and hoses) that should, in principle, result in a large improvement in reliability. But in most cases the performance we need has not yet been demonstrated, and some components are not yet commercially available. Detailed investigations are planned during the development phase before making a final choice.

2. REQUIREMENTS

The receiver plan (see chapter 5 of this book) calls for dual-polarization coverage of 33 GHz to 950 GHz in 10 bands, with the lowest three bands (33-116 GHz) using HFET amplifiers as the initial stages, and the remaining 7 bands using SIS mixers. It is desired to have the HFET amplifiers as cold as 15K; colder would not lead to significant improvement, and temperatures as warm as 50K might be acceptable. The SIS mixers (at least for Nb devices) must be kept below

4.5K in the worst case, and the nominal temperature should be 4.0K or less. Each receiver will have a cold feed horn at the same temperature as its first stage, and optical components outside the dewar are to be minimized. (As we shall see, eliminating all external optics after the subreflector implies large windows at the lower frequency bands, creating substantial heat load.)

(The receiver plan might change, but the cryogenics plan given here is based on the receiver configuration just described.)

With the goal of keeping the cryogenic system as simple and hence as reliable as possible, configurations with heat sinks at only two cryogenic temperatures are being considered (as opposed to the three stages that are traditionally used to reach 4K). The coldest stage would then have to be at 4K, and the other stage might be at about 40K. Some commercially available Gifford-McMahon refrigerators achieve these temperatures. Table 6.1 shows the likely distribution of cooled components for the 2-stage and 3-stage cases.

Table 1: Temperatures of Cooled Components

Three-Stage Case	**Two-Stage Case**
Stage 1, 70K	Stage 1, 40K
Radiation shields	Radiation shields
IR filters at windows	IR filters at windows
Cable and WG heat sinks	Cable and WG heat sinks
LO multipliers or photodiodes	HFET amplifiers, RF and IF, stages 3--5 (34 ea.) LO multipliers or photodiodes
Stage 2, 20K	Stage 2, 4K
Cable and WG heat sinks	All feedhorns (10 ea.)
HFET receiver feed horns (3 ea.)	SIS mixers (14 ea.)
HFET amplifiers, RF, bands 1-3 5 stages (x6)	HFET amplifiers, RF and IF, stages 1 and 2 only (34 ea.)
Stage 3, 4K	
SIS mixers (7 ea.)	
SIS receiver feed horns (7 ea.)	
HFET amplifiers, IF, stages 1 and 2 only (28 ea.)	

For the three-stage case, an attempt has been made to estimate all contributions to the heat load for the 10-receiver system. The result is given in Appendix A. The calculation assumes that all receivers are in the same dewar and can use a common refrigerator, but splitting them into separate dewars would affect only the radiation shields, which contribute a relatively small amount. The radiation through each window is assumed to be mostly sunk to the first stage by a fairly thick (several mm) PTFE infrared filter. The filter performance assumed is based on the measurements reported in [1]. A separate window is assumed for each band, with the sizes taken from [2]. The thermal properties of coax cables and waveguides are taken from [3]. The photonic and multiplier LO options (see chapter on local oscillators) are treated separately, but each is assumed to include 100 mW of RF dissipation; the multiplier option has slightly more load on the first stage due to its input waveguides.

It is apparent from Appendix A that the dominant load on every cooling stage is from the 300K radiation incident on the windows. This is true for each receiver separately as well as for the grand total. (This depends on using a large number of layers of multi-layer insulation ["superinsulation"] to keep the radiation from opaque surfaces low.) Therefore, minimizing the window area is one of the best things we can do to minimize the load.

Without any special effort, the calculated loads are about 36W, 1.4W, and 0.7W on the three stages. Appendix A also shows some ways to reduce this. A great deal of the first stage load (22W) and second stage load (1.2W) can be eliminated if the large windows of the three HFET receivers could be reduced to 2.2in diameter (from 9.0, 4.5, and 3.2in, respectively) by using external optics to reduce the beam waist diameter. Slightly more could be saved by eliminating the 33-50 GHz band entirely. The calculation assumes that all SIS receiver windows are 2.2 in diameter, which is what's needed for the lowest frequency band; actually, smaller windows can be used at the higher frequencies. After taking these steps, the largest remaining loads are from HFET amplifier dissipation and from conduction through coax cables. If only one SIS receiver is in use at one time, and if the bias is turned off for the other IF stages, significant saving results. (Note that the HFET receivers remain on, and could be used simultaneously.) If the IF coax from 4K to 20K is lengthened to 20cm (at some sacrifice in loss), further savings occur at 4K. With all of these measures, the loads become 6.5W, 0.27W, and 0.37W, respectively.

It should be emphasized that these are theoretical results, and are very preliminary. Whether they represent reality remains to be seen.

3. OFF-THE-SHELF TECHNOLOGIES

i. Two-stage Gifford-McMahon.

(to be written, see also MMA Memo #185.).

Table 6.2: Commercial Two-Stage GM Refrigerators for 4K					
Mfgr	**Model**	**Load@4.2K**	**MinTemp**	**System Price**	**Input Power**
Daikin[1]	CSW210	0.8 W	3.0 K	$35k	6.7 kW
Sumitomo[2]	SRDK415	1.5	3.2	41k	7.5
Sumitomo[2]	SRDK408	1.0	3.1	39k	7.5
Sumitomo[2]	SRDK205	0.5	3.1	24k	3.4
Leybold	4.2GM	0.5	3.4	37.5k	6.5
Boreas[3]	B100	0.9	3.1	38.5k	3.0

Model	1stStg	Q@T	P1,atm	P2,atm	He flow	Eff/Carnot[4]
CSW210	35W @41K	21.8	7.1	85 Nm^3/h	.041	.0084
SRDK415	30W @40K	?	?	?	.040	.0141
SRDK408	37W @40K	23.4	6.8	80	.041	.0094
SRDK205	4W @40K	?	?	30	.018	.0104
4.2GM	50W @50K	22.1	7.5	88	.044	.0054
B100	2.5W@80K	21	1.0	20	.023	.0211

Notes:

[1] Represented in U.S. by APD Cryogenics, Inc.
[2] Represented in U.S. by Janis. SRDK405 is new, promised in 4Q97.
[3] Not GM, but a proprietary hybrid cycle, partially recuperative.
[4] Efficiencies relative to Carnot, first with specified loads on both stages, then with zero 1st stage load but same power consumption.

---

ii. **Hybrid three-stage:**

2GM + JT.

4. ADVANCED TECHNOLOGIES

i. **Pulse Tube Refrigerators.**

(to be written; some notes:)

- Generally not commercially available (but 1-stage 10W@77K now avail from Iwatani)
- Orientation dependence may require tilted installation
- Cyclic, producing temperature fluctuations
- See refs [5--7] and other works cited therein.

ii. **Sterling-Style Compressors.**

- Produces cyclic pressure variation without valves, suitable for pulse tube or Sterling coolers

- Requires compressor close to cooler, but can be tipped
- Compressor cycle must be synchronized to refrigerator cycle for Sterling; compressor drives cycle for pulse tube
- Much more efficient than remote compressor with valves, as in GM, because of work recovery during expansion
- Until recently, these devices were very expensive and had relatively short lifetimes, with mainly military applications. But this is now changing. New compressors have clearance seals and flex bearings, with nothing subject to wear; low cost production methods under development by several companies. Further investigation needed.

iii. **Flexible Hoses.**

This is a simple one, and not really a matter of "advanced technology." A common cause of failures on NRAO's existing telescopes is the development of leaks in flexible hoses carrying high-pressure helium around the rotating axes of the antenna. These are corrugated metal hoses of stainless steel or brass. No manufacturer specifies the flex life of such hoses, yet it is well known (e.g., [4]) that the fatigue failure rate drops sharply to zero at a known value of peak stress for each material. By design, we should keep the stress below this level. This can be done by increasing the bend radius and/or decreasing the pitch of the hose corrugations (both at increased initial cost) as much as necessary. During the development phase, we will confirm this design approach experimentally. It should be possible to design for a flex life of $>10^6$ cycles, which should exceed the life of the array.

5. PRELIMINARY DESIGN

- Single dewar and cryocooler for all receivers.
- Maximum window diameter 2.2 inches; smaller for receivers above 245 GHz.
- Other thermal design choices (component placement, wire and transmission line types and lengths, etc.) consistent with Table 1 and Appendix A of this chapter.
- Three stage cryocooler with performance:
  Stage 1 <=80K at 12W
  Stage 2 <=20K at 1W
  Stage 3 <=4.0K at 0.5W

- Cryocooler is hybrid, with stages 1 and 2 using pulse tubes and stage 3 using a J-T expander. No moving parts at cryogenic temperatures.
- Compressor is conventional oil-filled rotary for PT stages, special dry rotary for JT stage.
- Predicted power consumption: 2.5 kW.

6. REFERENCES

[1] J. Lamb, "Infrared filters for cryogenic receivers," EDIR #290, 23 Apr 1992 (NRAO, Green Bank).
[2] J. Lugten and J. Welch, "A suggested receiver layout for the MMA antenna," MMA Memo #183, 15 Sep 1997.
[3] S. Weinreb, "Cryogenic performance of microwave terminations, attenuators, absorbers, and coaxial cable," EDIR #223, Jan 1982 (NRAO, Green Bank).
[4] T. Baumeister, ed., *Marks' Standard Handbook For Mechanical Engineers*, eighth edition, chapter 5. New York:McGraw-Hill, 1978.

[5] C. Wang et al., "A two-stage pulse tube cooler operating below 4K." *Cryogenics*, vol 37, pp 159--164, 1997.
[6] G. Thummes et al., "Convective heat losses in pulse tube coolers: effect of pulse tube inclination." *Cryocoolers 9*, R. Ross, ed. NY:Plenum Press, 1997.
[7] Zhu, Wu, and Chen, "Double inlet pulse tube refrigerators: an important improvement." *Cryogenics*, vol 30, pp 514--520, 1990.

Appendix A: Heat Load Calculations

		4K (mW)	20K (mW)	80K (mW)
Radiation Shields				
1st stage:	50cm H x 50cm D (1.18m^2)			
	e=0.1 both sides, 10 layers MLI			1211
last stage:	e=0.1 @ 80K, e=0.9 @ 4K, 900 cm^2			
	5 layers MLI (0 layers => 21mW)	1.7		
(See Cryo Engr, Fig 7.23 p 404.)				

HFET Receivers -- 3 each

IR absorbing filters at 80K between windows and feeds (see EDIR 290).
Amplifiers and feeds at 20K.
Output at RF to 300K via waveguide; mixers at room temp.
Waveguide 10cm long to 80K, then 10cm to 300K.
Bias via #36 BeCu wires, 30cm to 80K or 20K, then 30cm to 300K:
1.267e-8m^2; 4-300K:190, 4-80K:20.3, 4-20K:1.21 uW-m.

		4K (mW)	20K (mW)	80K (mW)
90-116	6 stages, 1.2v*25mA total, 30mW/amp, x2 pol		60	
	WR10 WG, SS 10mil, Cu 100uin, x2 pol		40	75
	Bias wires, (2/stg+LED+com)*2pol=28ea		1.8	16
	Window 3.2in dia 5.1887e-3m^2		110	2354
70-90	4 stages 1.2 15mA x2 pol		36	
	WR12 WG		24	94
	Bias wires 20ea		1.3	11
	Window 4.5in dia .01026m^2		218	4654
33-50	4 1.2 15mA x2 pol		36	
	WR22 WG		24	166
	Bias wires 20ea		1.3	11
	Window 9in dia .04104m^2		872	18620
SUBTOTALS, HFET RECEIVERS			1424.4	26001

SIS Receivers -- 7 each

IR absorbing filters at 80K.
Feeds and SIS mixers at 4.0K.
Single-stage HFET amplifier integrated with each balanced mixer
(2 amplifiers per dual-sideband mixer), 4-12 GHz, 1.2V 5mA.
Additional 4-stage IF amp at 80K.
Photonic mixers at 80K, waveguide or quasi optical to SIS
depending on band, separate optical fiber into each, one per

polarization per band, level control via external optical attenuator; OR

Multiplier chain at 80K, waveguide in from 300K, one per polarization per band, level control via d.c. bias on last multiplier, outputs as above. 100mW drive, maximum of 4 operating simultaneously.

125-175 GHz			
Window 2.2in	52		1113
1st IF stage x2sidebds x2pol @6mW	12		
Wires to 4K 2/SIS 2/HFET 2pol com spare = 26	1.76		14.7
LO waveguide WR8 SS 10mil Cu 0.1mil 10cm	6.7		
IF coax 4ch SS 10cm->80K, 20cm->300K	24		148
2nd IF amp x2sidebds x2pol @ 18mW			72
Photonic mixer dissipation			100
Multiplier			
Input WG WR12 10cm			94
Dissipation			100
Wires to 80K 10/ampl 2/photomix|mplr = 44ea			25
SUBTOTAL THIS BAND	96.5		1567[m] 1473[p]
175-245 GHz			
245-320			
320-416			
416-510			
602-720			
787-950			
TOTAL SIS FOR 7 RECEIVERS ASSUMED IDENTICAL	675.2		10369[m] 9711[p]
GRAND TOTAL, shields + HFET + SIS	677	1424	36370[m] 35712[p]
POTENTIAL SAVINGS:			
Turn off bias to IF amplifiers in all but one SIS receiver	-72		-432
Windows for top 4 bands (>320 GHz) to 1.1in dia	-156		-5843
IF coax 4K->80K increase to 20cm long	-84		
Windows for all HFET receivers to 2.2in dia (requires use of ext lens or mirrors)		-1044	-22289
Delete 33-50 GHz band		-933	-18797
Both of the last two items		-1157	-23579
GRAND TOTALS WITH ALL THE ABOVE SAVINGS	365	267	6516[m] 5858[p]

[m] -> with multiplier LOs (only one driven)

[p] -> with photonic LOs (only one driven)

MMA Project Book, Chapter 7

LOCAL OSCILLATORS

Darrel Emerson
Last revised May 27 1998

Introduction

There are currently two independent local oscillator development programs. The first (see below) is a photonic local oscillator, which involves beating two IR lasers together to produce a mm-wave or submm-wave phase-locked difference frequency. The second is a more conventional system, involving frequency multipliers. If the photonic scheme is successful, it offers considerable savings in cost, reliability and complexity to the MMA project; however, it is as yet an unproven system and so NRAO is currently pursuing both approaches in parallel.

---

1. **Photonic Local Oscillator System**
2. **Mm-wave Multiplier Local Oscillator System**

PHOTONIC LOCAL OSCILLATOR

John Payne
Bill Shillue
Last revised July 13 1998

1. Introduction

The local oscillator system for the MMA presents a great challenge for the instrument builders. The generation of a pure frequency with high phase stability in the frequency range of 90-900 GHz at each of forty antennas, and preserving the phase relationship between antennas for long (perhaps hours) periods of time, is perhaps the most difficult part of the instrument. Although, in theory, the task could be completed with components available today by adopting the conventional route of a phase-locked oscillator at a frequency of around 100 GHz, followed by multipliers, the cost and complexity are daunting. There have been recent developments in the so called "conventional" techniques that suggest that the reliability and cost of this approach may be greatly improved by the application of new beam lead diodes and MMICs, and we are pursuing this approach as described in Section 2 of this Chapter.

However, recent advances in laser diode technology and optical fiber transmission systems raise the possibility of a local oscillator system for the MMA, based on the mixing of two optical signals separated by the required local oscillator frequency. Such a system could be realized using mainly commercially available components, resulting in significant savings in both cost and manpower when compared to the conventional approach of a phase locked oscillator followed by passive multipliers. This approach is so attractive that we have mounted a development effort to investigate the feasibility of such an approach.

Several groups have worked on systems similar to this. The phase-locking of the beat note between two infrared lasers to an external microwave standard, with the spectral purity required of the MMA, was first demonstrated many years ago and is now regarded as routine. (For references and more details, see MMA memo #200). Beat notes of up to several THz have been demonstrated with a cooled fiber-coupled photomixer at power levels that appear to be marginally adequate for supplying the LO to the SIS receivers on the MMA. New detector fabrication techniques hold the promise of increased power levels in the wavelength range of the MMA.

The potential advantages of such a system may be summarized as follows:

1. The majority of the components needed for the realization of the proposed scheme are commercially available. The communications industry has a huge investment in optical fiber systems, and the system outlined here exploits these fairly recent developments. We can be certain that intense development in this area will continue.
2. All of the frequency synthesis components of the local oscillator system may be situated in a laboratory environment remote from the array. At the antennas, only some leveling electronics and a photomixer are required. In terms of serviceability and reliability, this is regarded as a great advantage.
3. The receiver interface is greatly simplified. Due to bandwidth requirements, the usual

Martin-Puplett quasi-optical LO injection scheme will not be appropriate. LO injection using conventional methods with waveguides entering into the cryogenic enclosure (for each receiver band) would involve a relatively high loss and would complicate the thermal design of the receiver. In contrast, all that will be needed in the photonic system is one optical fiber into the receiver dewar resulting in negligible heat load. Vacuum feed-throughs for fiber are fully developed commercially.

4. There is a great reduction in complexity .
5. The proposed system eliminates the need for the usual microwave harmonic mixers.
6. The real cost promises to be far less than a conventional system.
7. In the conventional system the passive multipliers following the fundamental oscillator introduce additional amplitude and phase noise. Investigation is needed here but it seems possible that noise on the fundamental oscillator may be multiplied by the square of the multiplication ratio. In the case of multiplying from 100 GHz to 800 GHz, phase noise enhancement by factor of 64 may well be involved which could prove unacceptable for use in the array. It may be that such an effect is absent in the case of the photonic system although work is needed to settle this question.

Details on the following are given in MMA Memo 200 , Photonic Local Oscillator for the MMA.

1. MMA LO requirements.
2. Description of proposed system.
3. Theory of photo-mixing.
4. Expected performance.
5. References.

2. Critical Issues.

There are many issues to be decided in the implementation of such a system. These decisions are made more difficult by the rapid development in the microwave photonics field. At the present time the tentative choices that we have made are as follows:

1. Wavelength of Operation.

   There are two so-called " communication windows " in fiber optic communications at present: the 1.3 micron window in which fiber dispersion is minimum, and the 1.5 micron window in which fiber attenuation is minimum. The 1.5 window seems to be in favor with industry at present, and at least one prominent manufacturer has announced the discontinuing of components for the 1.3 micron band. Also, although it is not yet known if dispersion will be a problem, a special zero-dispersion fiber is available at 1.5 micron for a nominal extra cost. Therefore, we have, provisionally, adopted the 1.5 micron band for our development.

2. Choice of lasers.

   The power output and spectral purity of lasers in the 1.5 micron band is improving rapidly. There are two types of lasers that seem to have the potential to be satisfactory for our application. These are the external cavity diode laser and the erbium doped fiber laser. The erbium doped fiber laser has greater power output and narrower line width than the external cavity laser but phase locking of these lasers has not been demonstrated, at least to our knowledge. The external cavity lasers are easily phase lockable to an external microwave reference and we have conducted measurements on a pair of rented lasers and reached the conclusion that the spectral purity is adequate for our purposes. It may well be that our initial choice of external cavity diode lasers is not the optimum one . As part of our development program, we intend to test other types of lasers as the state-of-the-art advances.

3. Choice of detectors.

The photo detector is the key element in the proposed scheme. Although beat frequency detection of up to several terahertz have been reported, the power levels have been marginal for our application. There are several types of photodetectors in use at high frequencies, and we have chosen to adopt the approach pioneered by UCLA: the velocity matched photodetector. A description of this and other detectors in use are given in MMA Memo # 200. We have a contract with UCLA to initially develop a photodetector producing 100 micro-watts of power over the 75-110 GHz band. The design is such that we believe that it will be scalable to the higher frequencies.

3. Development Course

As with many development projects the time scales are uncertain but the following is an estimate of the times for the various tasks that have to be completed to bring the development project to completion.

	Task	**People**	**Completion Date**
1)	Set up two external cavity lasers phase locked to a microwave ref. Measure phase and amplitude noise at beat frequencies up to 18 GHz.	1 EE	8-98
2)	Repeat (1) with Erbium doped fiber lasers.	1 EE	10-98
3)	Continuing investigation of round-trip photonic phase calibration.	½ E	5-00
4)	Broadband phase/amplitude calibration (link to Memo #?).	½ E	5-00
5)	Design and build the optical comb generator.	1 EE + ½ Tech	2-99
6)	Test comb generator up to 80 GHz using commercial detectors.	1 EE + ½ Tech	5-99
7)	Test and evaluate 100 GHz UCLA photo-detector	1 EE + ½ Tech	9-99
8)	Test and evaluate 230 GHz UCLA photo-detector	1 EE+ ½ Tech	12-99 (Note 1)
9)	Complete laboratory test using two 230 GHz photodetectors.	1 EE + ½ Tech	5-00

LOCAL OSCILLATORS: MULTIPLIER SYSTEM

Richard Bradley

Last revised July 14 1998

2. Local Oscillator System with Modern Multiplier Chains

1. LO Specifications

Frequency Requirements:

Table 7.1. Possible MMA receiver and LO frequency plan [1]

Receiver Type	Frequency Range (GHz)	Frequency Ratio	LO Range (GHz)	LO Frequency Ratio
HFET	33-50	1.52	55-64	1.16
HFET	68-90	1.32	90-104	1.16
HFET	90-116	1.29	76-94	1.24
SIS	125-175	1.40	137-163	1.19
SIS	175-245	1.40	187-233	1.25
SIS	245-320	1.31	257-308	1.20
SIS	320-416	1.30	332-404	1.22
SIS	416-510	1.23	428-498	1.16
SIS	602-720	1.20	614-708	1.15
SIS	787-950	1.21	799-938	1.17

Power Requirements: Approximately 100 microwatts per receiver.
Amplitude Noise: One Kelvin per microwatt.
Phase Noise: Three degrees over time scales greater than one second.
System Compatibility: Phase lock to the distributed microwave reference signal; compatibility with fringe rotation functions; provide proper frequency monitor and control; provide power leveling control.
Other Issues: System should have high reliability; overall cost should be minimized.

2. Discussion of Specifications

The conventional approach to generating local oscillator (LO) power for millimeter and submillimeter wave heterodyne mixers is to generate power at a lower frequency using a suitable phase-locked source, and to convert this power to the desired commensurate frequency using a nonlinear diode such as a varactor in a frequency multiplier circuit. Although useful for single- dish telescope receiver systems, the conventional approach, using current state-of-the-art components, is highly impractical for large array-type radio telescopes for which manageable cost and high reliability are important factors. A summary of the MMA LO specifications, current state-of-the-art, and development plans are presented. See [1] for more information.

Frequency Requirements

A proposed band plan for the MMA is shown in Table 7.1. The first three columns indicate the type of receiver, either HFET for the lower frequencies or SIS for the higher frequencies, the RF band, and the RF band-delimiting frequency ratio defined as fmax/fmin for each band. The highest frequency band is a future possibility and is not planned to be implemented as part of the initial construction. Assuming an IF band from 4-12 GHz, columns four and five show the LO tuning range required and the LO band-delimiting frequency ratio defined as fmax/fmin for each LO range. A prime feature of this plan is that all LO frequencies above 60 GHz can be derived from two phase-locked sources: #1 covering the range 60-90 GHz, and #2 covering 80-120 GHz. It should be noted that since the noise of SIS mixers may increase with IF frequency, the sensitivity of the radiometer for spectral line observations may be improved somewhat if only the lower portion of the IF band is used. The LO tuning range required to achieve optimum sensitivity may therefore be greater than that shown in the Table 7.1.

Power Requirements

A specification for the LO power level is derived from the pump power required by the SIS mixers which is approximately 1 microwatt. In the worst-case scenario where only single-ended SIS mixers are used, a waveguide or quasi-optical LO coupler, having a coupling factor of -20 dB, will be required to combine the LO with the RF signal. As a result, the amount of LO power required at the input of the receiver will be approximately 100 microwatts. An estimate of frequency conversion efficiencies that form realistic yet challenging goals for new broadband, fixed-tuned, planar frequency multiplier designs is given in Table 7.2. The first three columns give the LO tuning range from Table 7.1, the driving source along with its tuning range, and the multipliers needed. Columns four and five give the multiplier efficiency and output power for a driving power of 50 mW.

It is proposed that each receiver of a dual-polarization system be equipped with a separate multiplier chain that is driven by a common source. The minimum required output from the single phase-locked source will therefore be 100 mW. Such pairing of the multiplier chains with the receivers has the following benefits: 1) switching of the LO source at 100 GHz rather than at the higher LO output frequency minimizes the losses associated with the long waveguide runs inside the dewar, the vacuum window, and the waveguide switch, 2) multiplier chains could be tied to the cryogenic refrigeration system to improve conversion efficiency and increase the varactor lifetime, and 3) an LO leveling circuit, perhaps using the SIS mixer current in a servo loop while adjusting the bias current on the frequency multipliers, could be incorporated into each multiplier chain. Given the expected performance of the multipliers, the available dynamic range is shown in column six of Table 2. This dynamic range can be increased substantially if balanced mixers are used [2], since the RF and LO ports will then be separated, eliminating the need for the -20 dB LO coupler.

Table 7.2. Estimated efficiencies required for multiplier chains (50 mW drive level assumed).

LO Tuning Range (GHz)	Drive Source & Tuning (GHz)	Multiplication Factor	% Conversion Efficiency	Output Power mW	Powe Dyna

137-163	#1 68-82	X2	30	15
187-233	#2 93-117	X2	20	10
257-308	#2 85-103	X3	5	2.5
332-404	#2 83-101	X2,X2	20,10	1.0
428-498	#1 71-83	X2,X3	30,3	0.45
614-708	#2 102-118	X2,X3	20,3	0.30
799-938	#2 99-118	X2,X2,X2	30,15,5	0.010

Amplitude Noise

The specification for LO amplitude noise is to meet an acceptable value for the gaussian noise that will be added to the front-end noise of the SIS receiver. The contribution of LO noise to the HFET front-ends will be negligible. The mixer LO noise manifests itself as noise sidebands associated with the CW source, but far enough away from it that the noise will ultimately appear in the RF passband of the receiver. It has been suggested that the LO amplitude noise contribution to the noise temperature of a single-ended SIS mixer be limited to one degree Kelvin, and with a typical LO pump power of 1 microwatt per mixer, a specification of 1 Kelvin per microwatt is therefore defined. A relatively low-Q bandpass filter, centered about the signal frequency, can be used to reduce this noise if needed. If balanced mixers are used, the specifications for the filter can be relaxed in proportion to the LO isolation that is provided, typically on the order of 10 to 20 dB, or perhaps eliminated entirely.

Phase Noise

The dominant contributor to the phase fluctuations encountered by the MMA will be atmospheric fluctuations along the line of sight of the instrument. Two distinct methods [3] are currently being considered for phase calibration: 1) Fast Phase Calibration (FPC) and 2) Radiometer Phase Correction (RPC). Based upon the requirements of these proposed phase calibration methods, as well as the need to use holography to measure the surface features to the desired accuracy of better than 8 microns (wavelength/100), the resulting phase stability specification for the electronics, as defined by the MDC Phase Calibration Working Group, is 3 degrees over time scales greater than 1 second. This specification will be used as a guideline for MMA LO development. However, specifications for allowable phase fluctuations on time scales shorter than one second have not yet been defined. Integration-and-dump times, associated primarily with mosaicing operations, will be the defining factor for this specification.

Reliability and Cost

Reliability is an important issue not only because of the number of components required but also due to the remoteness of the observing site [4]. Reliability can be greatly enhanced by using all-electronic tuning and by replacing the fragile point contact varactor with the more rugged planar varactor. Due to the relatively large current densities in varactors, anode temperatures can reach well over 100 degrees C above ambient, thus compromising the long-term lifetime. The lifetime can be increased indefinitely through the use of cryogenic cooling which is typical in modern receivers and therefore should not increase the cost of the LO. The cost of building frequency multipliers is rather large due to the current complexity of the micro fabrication required. This cost can be reduced substantially at the circuit design stage by using monolithic (MMIC) technology [5], minimizing the machining operations required, and reducing the need for close tolerances during machining steps so that efficient duplication can be achieved. Finally, the higher-frequency multipliers should be designed as cascaded components of doublers and triplers for interchangeability.

3. State of the Art

In order to improve upon reliability and decrease cost, the limitations of the two basic components in the current LO system, namely the oscillator and the frequency multiplier, must be carefully examined. For the power source, the mainstay is the Gunn-effect oscillator which has been used successfully for many years because of its adequate output power, inherently low amplitude noise characteristics, and electronic fine tuning making it well suited for phase-locked circuitry. However, for large array applications, its usefulness is somewhat compromised since the coarse tuning is accomplished through mechanical adjustment of a high-Q resonant cavity. It is an expensive task to make this mechanical adjustment automated, accurate, repeatable, and reliable. Also, such cavity systems can suffer from unwanted moding which results in narrow frequency bands in which the output power can drop to very low levels. The mechanical tuning is limited in range as well, and, hence, several Gunn oscillators are needed to cover a given waveguide band. The maximum operating frequency of second harmonic Gunn oscillators is about 150 GHz, and so to reach millimeter and submillimeter wavelengths, frequency multipliers are required. State-of-the-art multipliers are limited in performance because of several factors, including: narrow instantaneous bandwidth requiring mechanically-adjustable tuning structures that may reduce reliability, low conversion efficiency leading to difficulties in power distribution, use of point-contacted varactors which are mechanically fragile structures, and intricate mechanical details making component assembly rather difficult. Overcoming these limitations is essential if conventional LO systems are to be made practical for the MMA.

All-electronic LO tuning has the advantage of improving reliability for array systems at a modest cost. The most useful all-electronic power source up to 50 GHz is the Yttrium-Iron-Garnet (YIG) tuned FET oscillator (YTO). A YTO can be tuned over a very broad band and it can easily be phase-locked to a reference source. A chain consisting of a YTO, followed by wideband, fixed-tuned frequency multipliers and low noise power amplifiers, forms a viable alternative to a Gunn oscillator chain. Wideband monolithic HFET power amplifiers are becoming increasingly more common up to 100 GHz, primarily due to current military and commercial demands for systems operating in this frequency range. However, the development of wideband frequency multiplier technology has lagged behind in development. Advances in this area will determine the success of future YTO-based millimeter and submillimeter wave LO systems.

Over the past few years, the single most important factor influencing future frequency multiplier development has been the advent of versatile computer-aided design packages enabling the design engineer to analyze complex electromagnetic structures, create and simulate detailed equivalent circuit models, and predict semiconductor transport properties, all to a high degree of accuracy. For the first time, the nonlinear dynamics of the varactor, the electrical properties of the semiconductor package, and the embedding circuitry of the multiplier can be analyzed together as a complete frequency multiplier circuit. Very successful millimeter wavelength multiplier circuits have recently been developed by

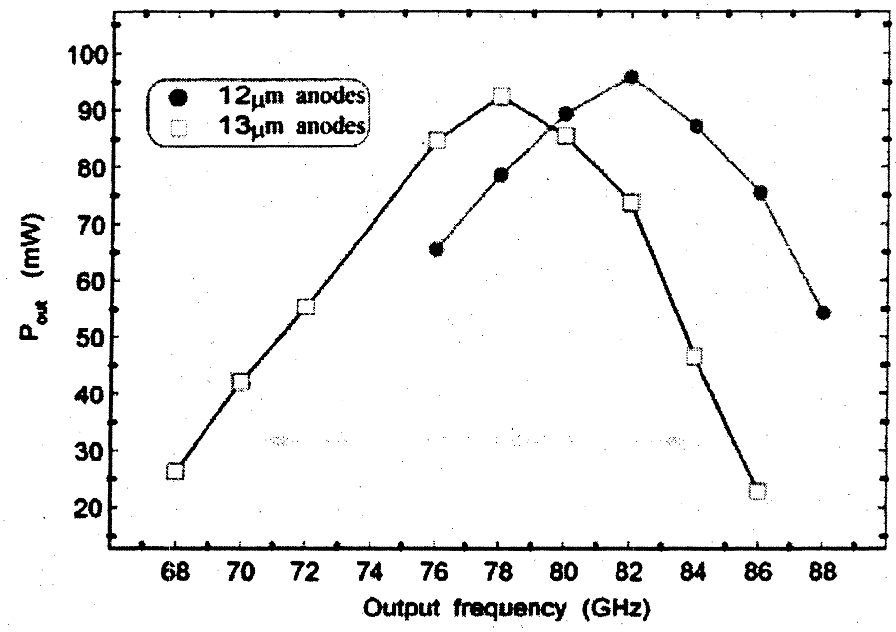
Porterfield et al. [6] and Erickson [7] based on calculations using modern computer-aided design tools. Upon applying such tools, one begins to understand the reasons behind the limitations of existing multiplier designs, thus opening the door to exploring new approaches and techniques never before possible in order to meet the stringent demands placed on the LO system by the MMA specifications.

4. Development Plan

The conventional LO development plan is divided into four technical development areas:

1. frequency multipliers using discrete planar varactors,
2. frequency multipliers using monolithic circuitry,
3. LO phase-locked source development and evaluation, and
4. functional prototype LO development.

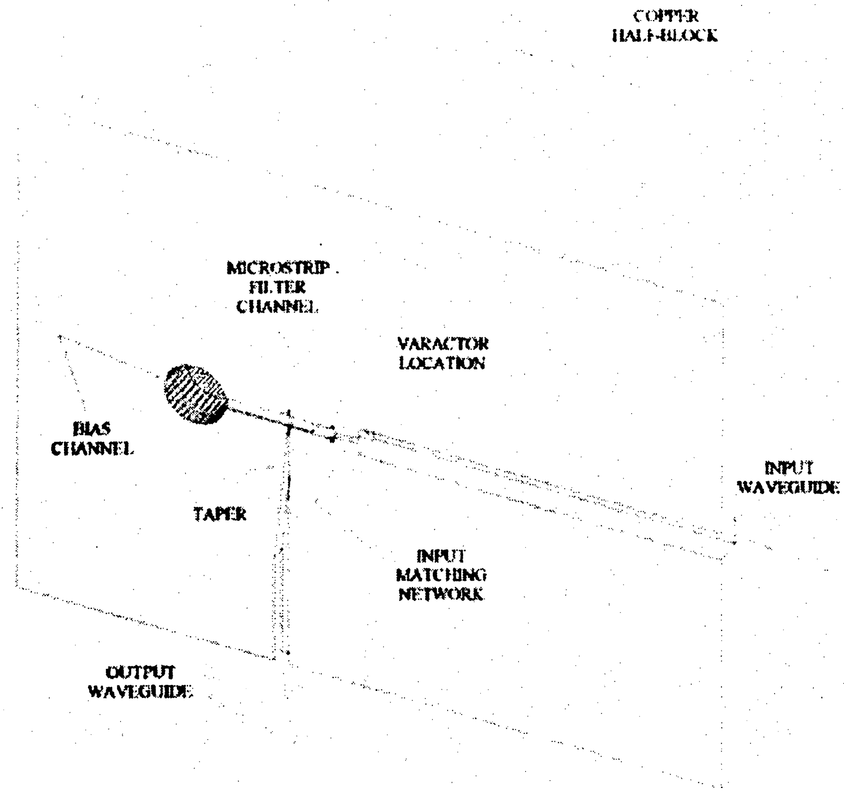
The first deals with broad-band, fixed-tuned frequency doublers used to extend the phase-locked loop LO system to cover the 137-163 GHz and 187-233 GHz bands. Frequency doublers for these bands will be based on the highly successful 40/80 GHz design [6] which uses a balanced planar varactor chip from the Semiconductor Device Laboratory of the University of Virginia. The measured results are shown in Fig. 7.2.1 for room temperature operation. The peak efficiency increased to more than 60 percent upon cooling the doubler block to 20 K. There are currently two new designs already in progress: 55/110 GHz and 110/220 GHz for use in a 690 GHz heterodyne tipping radiometer [8]. These designs will become the first iteration of the MMA designs. Figure 7.2 shows a sketch of the 110/220 GHz block. Future iterations will be concerned with increasing the output power of the doublers and increasing the operational bandwidth as well as making the designs easier to fabricate. Designs using discrete planar varactors are limited to about 250 GHz because the size of the chip package becomes electrically large and therefore the multiplier circuit becomes exceedingly more difficult to tune properly over a wide bandwidth.



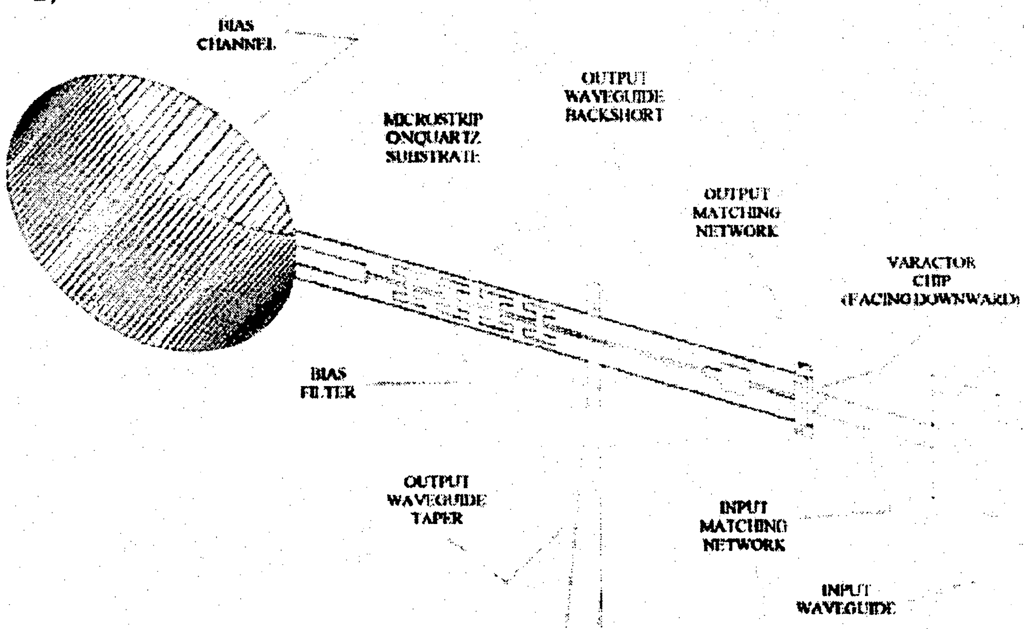
Click to zoom

Figure 7.2.1 Measured output power as a function of frequency for the 40/80 GHz doubler [6].

A)



B)



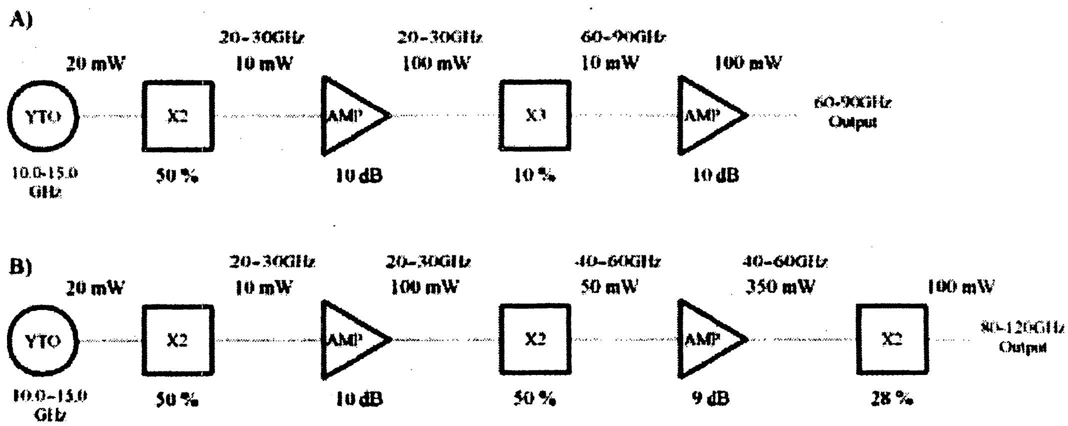
Click on images to zoom

Figure 7.2.2 Sketch of 110/220 GHz doubler.

A) one-half of the split block design and B) details of microstrip circuitry.

The second development area will focus on the doublers and triplers needed to extend the PLL LO up to 700 GHz. Because of the short wavelengths involved, these designs will require some level of circuit integration (MMIC). Due to the high development costs of MMICs, sharing ideas and designs between groups is prudent and so we are exploring areas of common need among projects such as the MMA, FIRST/HIFI, etc. Recently, a collaboration was formed between the NRAO-CDL and NASA-JPL for MMIC multiplier development, where our experience in multiplier circuit design and electromagnetic simulation will be coupled with the MMIC fabrication experience of JPL. Two initial designs are currently underway: a 200/400 GHz balanced doubler and a 220/660 GHz balanced tripler.

The third development area will concentrate on the two phase-locked LO systems used to drive the frequency multipliers. A block diagram of the RF oscillator is shown in Fig. 7.2.3. The 100 mW, 60-90 GHz oscillator consists of a 10-15 GHz YTO followed by a doubler to 20- 30 GHz, a 10 dB power amplifier, a tripler to 60-90 GHz, and a second 10 dB power amplifier. The 80-120 GHz oscillator consists of a YTO, doubler, and power amplifier to 20-30 GHz as in the previous case, but here it is followed by a doubler to 40-60 GHz, a 9 dB power amplifier, and a high power doubler. Both systems will require appropriate interstage filters and isolators. The bandwidth of the phase-locked sources may be limited by the availability of power amplifiers since research groups requiring such amplifiers are forced to share designs due to high development costs. It is not clear at this time how the MMA LO will be affected by amplifier availability.



Click to zoom

Figure 7.2.3 All-electronic oscillator for
A) 60-90 GHz and B) 80-120 GHz.

Understanding the noise contribution of the various approaches to generating LO power at 100 GHz is important. The type of LO system that will meet the array noise performance specifications will have an enormous impact on the entire MMA system design, both in the degree of complexity and the overall cost. Furthermore, based on the measurements, a decision will be made regarding the highest frequency that can be included within the phase-locked loop; this decision will impact millimeter-wave harmonic mixers.

To study PM noise, a standard phase noise measurement system will be constructed. This system will be based on the complete HP phase noise measurement system, including the necessary mixers to translate the 100 GHz signals down to the operating range of the instrument. A stable reference oscillator and distribution system should also be purchased. For AM noise studies, a 100 GHz SIS mixer, which is currently available at the NRAO-CDL, may be used. However, a balanced Schottky system, dedicated specifically for AM noise measurement, would be quite useful, since the system will not require cooling and the balanced nature of the design results in the LO noise directed to a separate IF port for ease of measurement. Two YIG-based phase-locked oscillator/ amplifier/ multiplier chains as well as phase-locked Gunn-effect oscillators will be compared on the basis of both AM and PM noise.

Finally, the fourth development area will center on the "functional prototype" which is defined here as a complete LO system, ready for installation into the prototype MMA antenna system. We will build upon the breadboard version developed earlier; various modifications will be done to provide for locking to the distributed microwave reference signal, insure compatibility with fringe rotation functions, provide proper frequency control, and provide power leveling control. Size reduction and modular packaging will also be performed. Close coordination with the front-end, system, and control groups will be essential.

REFERENCES

1. Bradley, R.F. "Wide Bandwidth, YIG-Based Sources and Modern Frequency Mult: Local Oscillator System," NRAO MMA Memo No. _____, May, 1998.

2. S-K Pan, Private Communication.

3. D. Woody, M. Holdaway, O. Lay, C. Masson, F. Owen, R. Plambeck, S. Radford, ¿ MDC phase calibration working group report," NRAO MMA Memo 144 (http://www.cv.n: 1995.

4. P. Napier and J. West, "High Altitude Medical and Operations Problems and So: Millimeter Array," Proc. SPIE, 3349, 1998.

5. R. F. Bradley and R. J. Mattauch, "Planar monolithic Schottky varactor diode frequency multipliers," Technical Report RL-TR-92-187, Rome Laboratory, Air For(Griffiss Air Force Base, NY, June 1992.

6. D. W. Porterfield, T. W. Crowe, R. F. Bradley, and N. R. Erickson, "A high-p(millimeter-wave balanced frequency doubler," submitted to IEEE Trans. on Microw¿ Techniques, Jan. 1998.

7. N. Erickson, "Wideband High Efficiency Planar Diode Doublers,? Ninth Intern¿ Space Terahertz Technology," Pasadena, CA March, 1998.

8. R. F. Bradley and S-K. Pan, "690 GHz Tipping Radiometer: A Design Survey," I Division Technical Note 171, Aug 28, 1995.

SYSTEM DESIGN OVERVIEW

Richard Thompson
Last revised July 13 1998

1. INTRODUCTION

The part of the system considered here includes the signal paths from the outputs of the front ends to the digital samplers, and the local oscillator (LO) system, but it does not include details of the front ends, the digital delays or the correlator. The purpose of this chapter is to provide a general description of the signal paths from the receiving front ends to the correlator, and the distribution of local oscillator and monitor-and-control signals to all parts of the system as required.

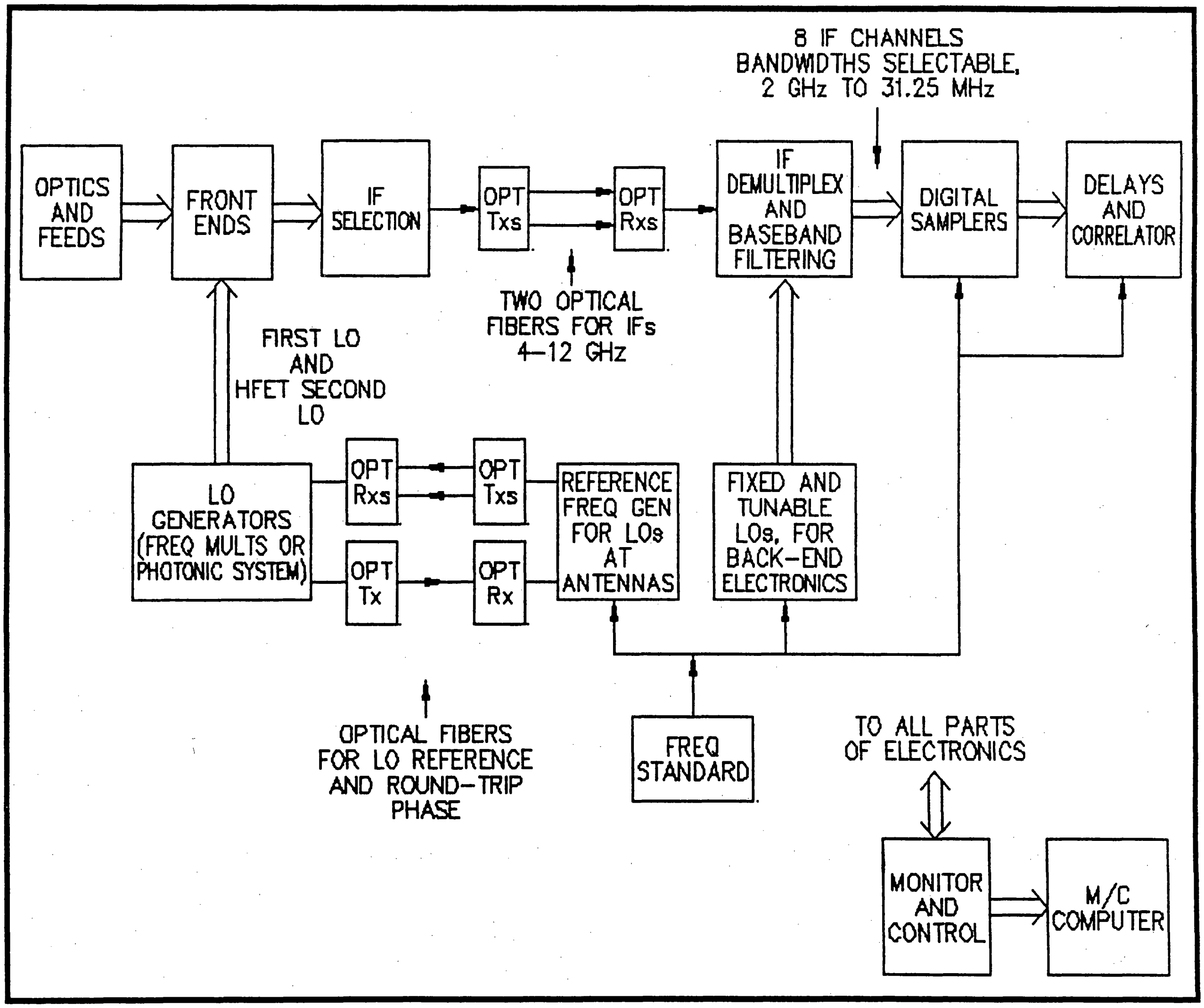


Figure 1: Simplied block diagram of the overall system

A simplified block diagram of the overall system is shown in Fig. 1. Components at the left-hand side of the optical fibers in the IF and LO paths are located at the antennas and those at the right-hand side are in the Electronics Building which is located close to the array. At each antenna there are ten receiver front ends which receive signals from Cassegrain feeds. The input stages use HFET (heterostructure field-effect transistor) amplifiers at frequencies from 30 to approximately 115 GHz and SIS (superconductor-insulator-superconductor) mixers for higher frequency bands up to 950 GHz. Two orthogonal polarizations are observed simultaneously, so there are two IF outputs from each HFET front end. For each SIS front end there are four outputs since the mixers will also provide separate outputs for the upper and lower sidebands. The IF signals transmitted from each antenna to the Electronics Building where the correlator is located will consist of two bands each 8 GHz wide, one for each polarization. The 8-GHz bandwidth is the widest that is deemed practicable considering limiting factors which include the need to keep the noise temperature of the SIS-mixer input stages as low as possible, and the processing capacity of the correlator (MMA Memo. 166, Memo 194). The IF band is tentatively chosen as 4-12 GHz to keep frequencies low while also keeping the relative bandwidth less than two octaves to allow satisfactory performance for components such as the IF hybrids which are used in the sideband separating mixers.

During the IF processing the signals are digitized and the insertion of compensating time delays and the cross-correlation of the signals from pairs of antennas is performed digitally. An important question is the location of the analog-to-digital (A/D) conversion in the data streams from the antennas to the correlator, in particular, whether the signals are transmitted from the antennas in analog or digital form. A/D conversion at the antennas considerably increases the complexity of the electronics at the antennas, which is undesirable because maintenance is most easily performed for equipment located in the Electronics Building. The stability of the frequency response of the overall analog system is an important factor in the performance of the array, and with analog transmission this includes the frequency response of the fiber optic system. However, it is very difficult to predict the degradation in performance, if any, that would result from analog transmission. The two 8-GHz-wide bands from each antenna can be transmitted as analog signals on two fibers. For digital transmission of the IF signals, with four-level digitization (two bits per sample), the total data rate from each antenna would be 64 Gb/s. Transmission of digital data would thus require more bandwidth than is needed for the signals in analog form. The present decision is to use analog transmission for the prototype test array, as described here. A more extensive discussion of the analog/digital transmission question can be found in MMA Memo. 142.

At the Electronics Building the two 8-GHz-wide IF bands from each antenna are split into four baseband channels, each 2 GHz wide. These are passed through filters that allow the bandwidth to be varied in factors of two, digitized, delayed and cross-correlated. A frequency standard and various synthesizers provide the reference frequencies for various local oscillators and for timing of the digital units. For the LOs at the antennas, standard frequencies are generated at the Electronics Building and transmitted to the antennas on optical fibers. A separate fiber is provided for measurement of the round-trip phase. Appropriate phase switching and fringe-frequency phase rotation are inserted through the LO system. A monitor and control system interacts with all parts of the electronics.

2. SIGNAL PATHS AT THE ANTENNAS

Figure 2 shows the system at the antennas. Note that the diagrams indicate the signal flow by showing mainly filters, mixers and switches, and other components will be found in the detailed design descriptions. Each of the four IF outputs from each SIS-mixer front end goes to one of four 1x10 switches. An IF band spanning 4 to 12 GHz (that is, a bandwidth of 8 GHz) is used for each of the two opposite polarizations received. An SIS mixer with a 4 GHz IF bandwidth and good noise temperature has been demonstrated (Padin et al. 1996) using an IF stage integrated with the mixer. It is believed that 8 GHz bandwidth is an achievable goal within the construction period of the MMA. However, in the shorter term the noise temperature of SIS front ends matched to a 4 GHz bandwidth may be better than those matched to the full 8 GHz bandwidth. Thus, instead of two 8-GHz-wide bands it may be preferable to use four 4-GHz-wide bands from each SIS front end (that is, both polarizations for upper and lower sidebands) in order to maximize the sensitivity. By means of the four 1x2 and two 1x3 switches shown in Fig. 2, the input to each optical transmitter can be selected to come from either one 4-12 GHz front end output or two 4-8 GHz front end outputs, one of the latter being converted to 8-12 GHz for transmission on the fiber. It is hoped to simplify this switching arrangement during the development phase since, in practice, the deviations of the IF passband from flatness can be expected to increase with the number of interconnections.

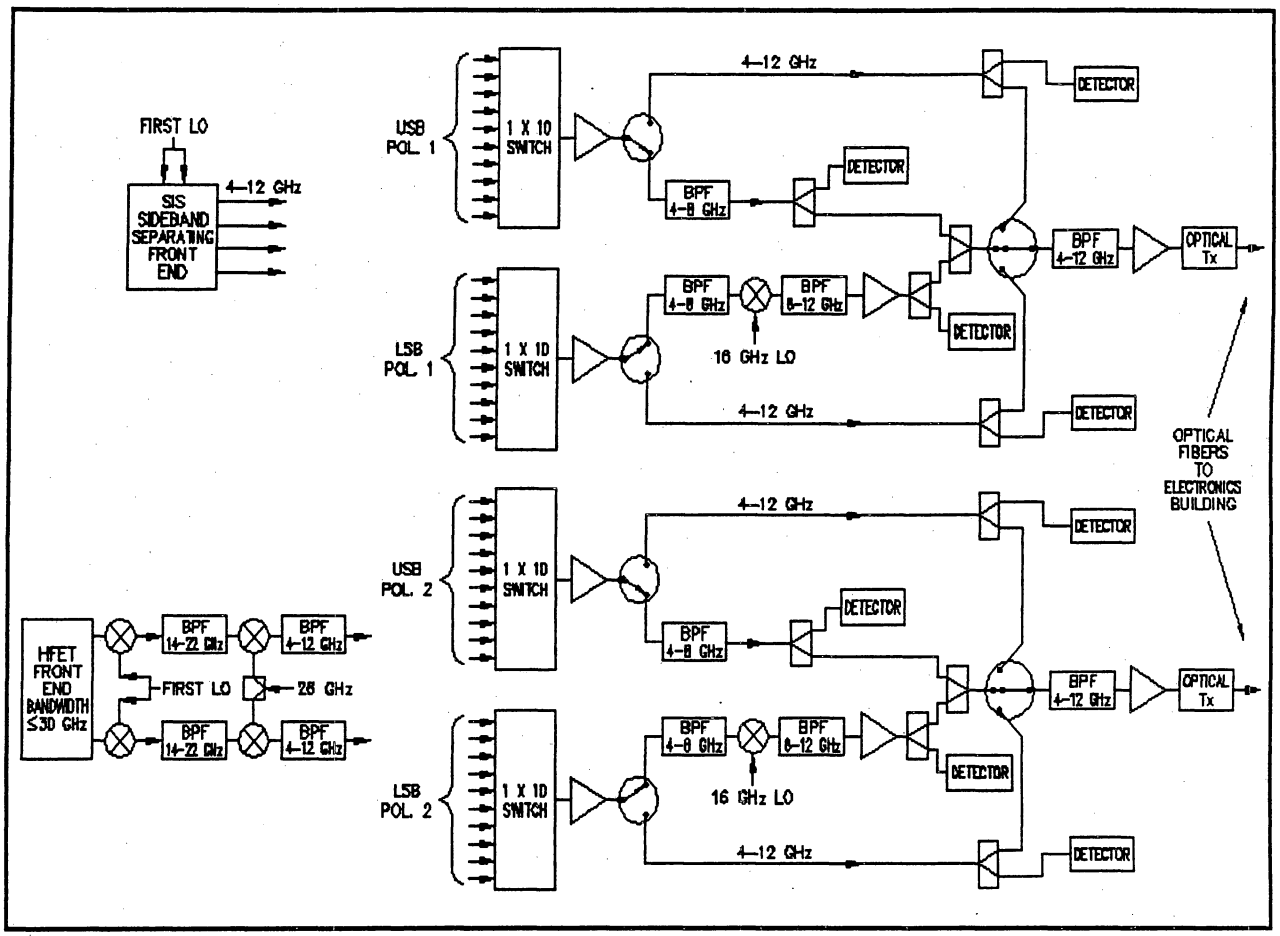


Figure 2: System at each antenna

The maximum bandwidth of any of the HFET front ends is approximately 26 GHz. The first IF band for the HFET front ends is 14 to 22 GHz, so there is a frequency gap of 28 GHz between the upper and lower sideband responses at the mixer that converts to the first IF. Thus for most HFET front ends it should be possible to avoid image responses if a filter is included to confine the amplifier response to the nominal band. However, two frequency conversions are required to get from the HFET input frequency to the 8-12 GHz IF band. Note that the IF band 14-22 GHz is chosen to be about as low in frequency as is possible to use while avoiding direct feed-through to the 4-12 GHz band. For reasons of cost it is preferable to keep intermediate frequencies as low as possible, but the IF bandwidths required place lower limits on the frequencies because of the need to avoid image responses, etc.

The arrangement of switches shown in Fig. 2 is one of several possibilities and the final system design will have to take into account the availability and cost of different types of switches, and the effect of their loss and reflections on the flatness of the IF response. Some modes of operation of the array may require continuous switching between two front ends at intervals as short as 10 s, so mechanical switches with limited switching-cycle lifetimes would not be suitable. The switches should be mounted within the front-end Dewar to minimize the number of coaxial lines that have to pass through the Dewar wall. The switches should be cooled to 15K to minimize the heat leakage into the cryogenically cooled components. It is not clear that coolable GaAs switches with sufficient bandwidth are available. Thus, an alternative which is being considered is the use of power combiners to replace the switches. The power combiners would be cryogenically coolable and front ends not in use would be switched off by removing power from their output stages. Since there would be a 10 dB loss in signals passing through a 10-way combiner, a corresponding increase in IF gain would be necessary.

3. SIGNAL PROCESSING AT THE ELECTRONICS BUILDING

Each of the 4-12 GHz IF bands is demultiplexed into four 2 GHz-wide bands when received at the Electronics Building, and these are all converted to a frequency of 2-4 GHz, as shown in Fig. 3. Demultipexing to bandwidths of 2 GHz allows the use of a clock frequency of 4 GHz for the samplers, which is within the state of the art but requires some development. Also the problem of keeping the overall bandpass characteristic uniform within approximately ~1 dB is simplified since the narrower 2-GHz bandwidths can be considered individually. The 2-4 GHz IF bands then go to individual Baseband Converter units, of which four are required per polarization for each antenna. A switching network allows the four IF bands for a given polarization to be connected to four Baseband Converter inputs in any manner desired. For example, all four Baseband Converters may be connected to one 2 GHz-wide IF band if it is desired to study four narrow lines that all lie within the same 2 GHz-wide band.

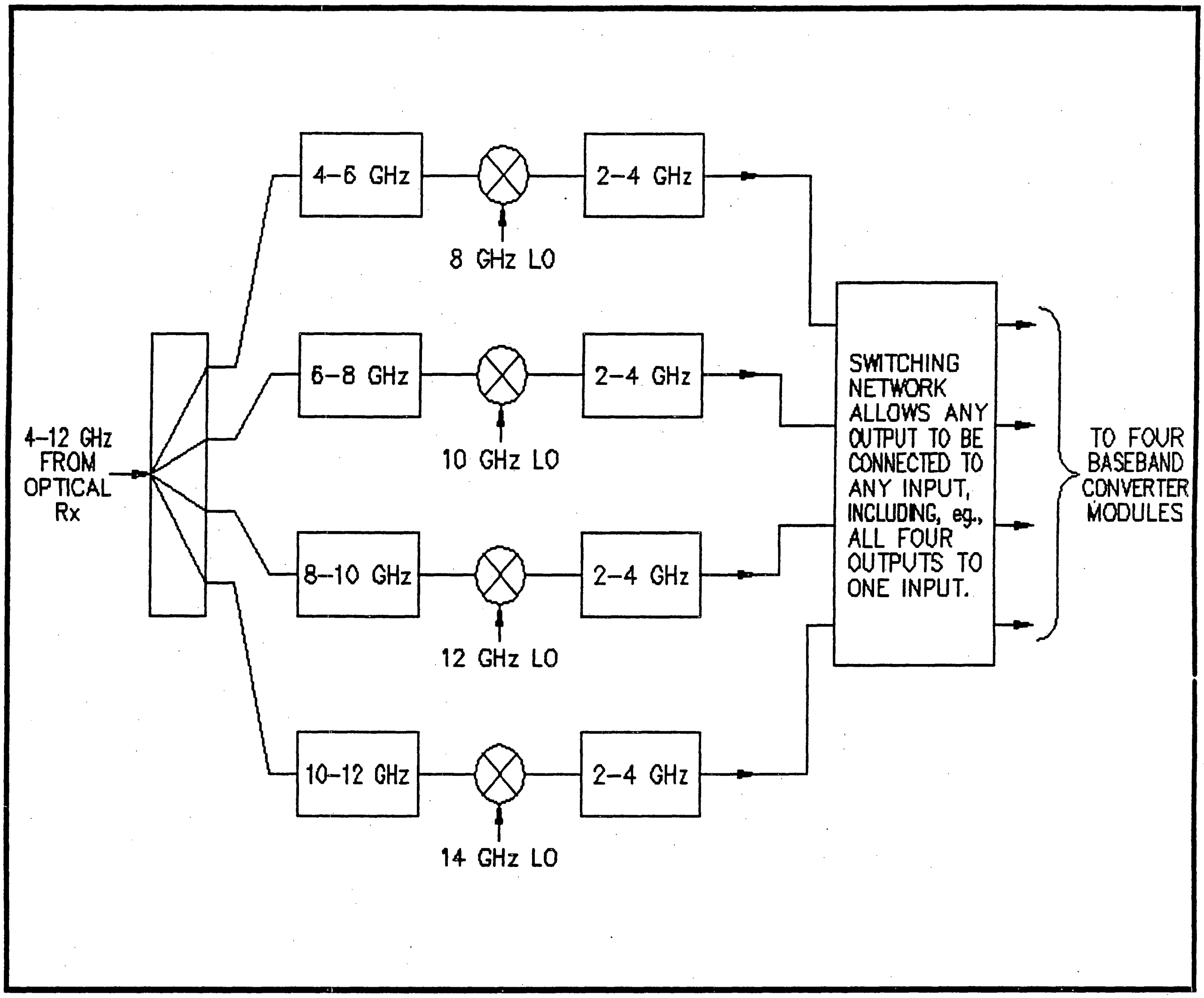


Figure 3: Demultiplexing the 2-GHz bands

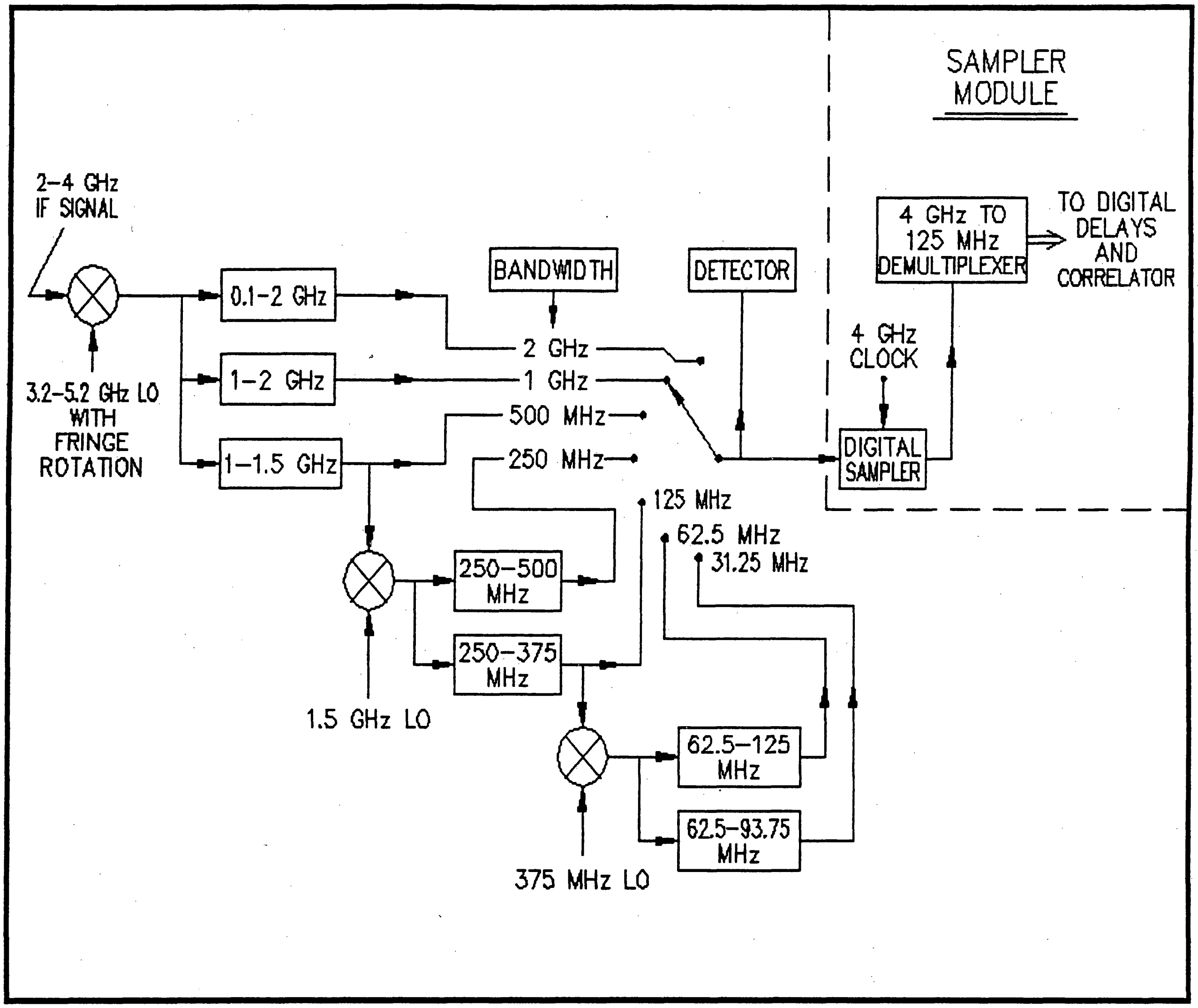


Figure 4: Baseband Converter

In the Baseband Converter unit in Fig. 4, the input signal at 2-4 GHz is converted to the baseband range and filters are provided for selection of bandwidths from 2 GHz down to 31.25 MHz in steps of a factor of two. Just ahead of the Baseband Converters the signals pass through the 2-4 GHz filters shown in Fig. 3. When converting the full 2 GHz bandwidth to baseband using a 4 GHz LO, frequencies above 4 GHz can get converted to frequencies at the low end of the baseband range. To minimize this effect the 2-4 GHz filters in the Baseband Converters should have a sharp response at the 4 GHz edge. The low frequency cutoff of the 0.1-2 GHz filter is also intended to help reduce such unwanted frequencies. The 1-2 and 1-1.5 GHz filters are far enough away from the low end of the 0.1-2 GHz baseband to be free from such unwanted effects. Similarly, the two frequency bands following each of the 1.5 GHz and 375 MHz LOs are well above zero frequency and will not include unwanted components from spectrum fold-over at an LO frequency. The reason for the variable bandwidth is that in the correlator the frequency resolution is increased as the total bandwidth to be processed is reduced. The 3.2-5.2 GHz LO at the first frequency conversion is tunable to allow the response of the baseband converter to be set at any part of the 2-GHz-wide IF input band. Thus for observations of spectral-line features the baseband responses can be adjusted in frequency and bandwidth to suit the characteristics of the line involved. In designing the Baseband Converter, it was decided to use a system involving a number of frequency conversions with filters chosen to reject the unwanted sideband responses, rather than using a sideband-separating mixer because the latter, although simpler, does not provide sufficient rejection of unwanted sidebands. Unwanted responses from other lines should be reduced by at least 40 dB. Table 1 lists the bandwidth of each filter, the frequency range of the passband as seen at the output of the first mixer (the one with the 3.2-5.2 GHz LO), and the tuning range of the LO which just covers the 2-4 GHz input band in each case. Image responses fall above 4 GHz in all cases.

Table 1. Filters in the Baseband Converter

Bandwidth	Filter response as seen at output of first mixer	LO tuning range
1.9 GHz	0.1-2.0 GHz	4.0 GHz
1.0 GHz	1.0-2.0 GHz	4.0-5.0 GHz
500 MHz	1.0-1.5 GHz	3.5-5.0 GHz
250 MHz	1.0-1.25 GHz	3.25-5.0 GHz
125 MHz	1.125-1.25 GHz	3.25-5.125 GHz
62.5 MHz	1.1875-1.250 GHz	3.25-5.1875 GHz
31.25 MHz	1.1875- 1.21875 GHz	3.21875-5.1875 GHz

The fractional bandwidths of the filters, other than the two widest ones, are in the range 0.4 to 0.67. Note that if the chosen output band comes from one of the three widest filters, the input of the 1.5 GHz LO signal to the module should, if necessary, be turned off to prevent pickup of this LO frequency in the earlier filters. Similarly, unless one of the two narrowest bandwidths is being used the 375 MHz signal should be switched off. The filtered outputs are sampled at a 4 GHz clock rate and digitized. The 4 Gb/s digital data streams from the samplers are each demultiplexed into 32 bit streams at 125 Mb/s and then go to the delay and correlator system. Redundant samples for the bandwidths of 1 GHz and less are removed as necessary at an appropriate stage. Eight Baseband Converter units are required for each antenna.

An alternative to the Baseband Converter module in Fig. 4 is the digital filter scheme shown in Fig. 5. Here the IF signal is sampled at its full bandwidth, and then the filtering is implemented digitally. Each 4 Gb/s output stream of the sampler is multiplexed into 32 bit streams at 125 Mb/s, which is the clock rate for the correlator and would also be used for the digital filters. Filter design is based on the FIR (finite impulse response) type, and discussed in more detail in MMA Memo. 204. The incoming sample stream is fed through a two-bit shift register, the length of which is tentatively chosen as 128 samples. Each stage of the shift register is assigned a number, referred to as a tap weight, by which the sample in the particular stage is multiplied. At each clock cycle the samples are multiplied by the tap weights corresponding to their positions in the shift register, and the products are summed and form one sample of the filtered output signal. At the next clock cycle each sample moves to the next stage of the shift register and the multiplication by tap weights and addition are repeated. It will be seen that in the time domain the series of summed products represent a sampled convolution of the input IF signal with a function represented by the series of tap weights. Thus, in the frequency domain the output stream represents the IF input spectrum multiplied by the Fourier transform of the tap weight function. The tap weights can be stored in a random-access memory (RAM) and different sets of values can be entered to provide different filter characteristics as required.

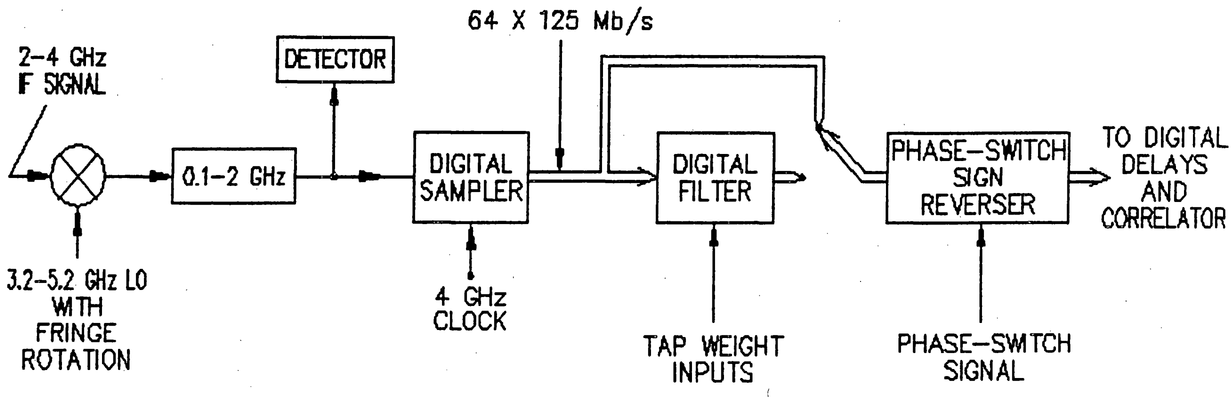


FIG. 5 Digital Filter Unit

Figure 5: Digital Filter option

The fact that the sampled data are represented by only four levels results in an important simplification, since the multiplication of one data sample by a particular tap weight can result in only one of four different values, and these can be stored as a look-up table in a RAM. The multiplication is then replaced by selection of the appropriate value. Note that as the filter bandwidth is halved the output sample rate is also halved, which, in effect, allows the data to be recirculated through the shift registers, doubling the effective number of tap weights.

The advantages of digital filtering are that the accuracy of the phase and frequency responses depends mainly on the timing accuracy of the digital circuitry, which allows greatly reduced dependence on temperature stability and better matching of responses for different IF signals. The ability to change the tap weights allows one set of digital filter circuits to provide many different responses. The use of digital transmission of the signals from the antennas would be simplified, since digital filtering allows the filters to remain at the correlator, and only the samplers need be moved to the antennas. Thus stability and flexibility are enhanced, and there should also be some cost saving. A disadvantage of digital filtering is that the output samples resulting from the tap-weight multiplication and addition are represented by multi-bit numbers. The number of bits then has to be reduced to whatever the correlator will accept, which is only two bits as presently conceived. The truncation of the data results in the quantization loss occurring twice, once at the sampler and once in the truncation. This additional loss need not occur when the full IF bandwidth is used since the digital filter can then be bypassed. A proposed digital filter implementation, as described in MMA Memo. 204, would require development of a custom, large-scale, integrated circuit but this would not so large and complicated as that for the correlator.

The digital filter will be pursued by computer simulation, followed by chip design. the goal would be to have a tested design for the array construction phase, but for the test antenna the more conventional design of Fig. 4 will be used.

4. TOTAL POWER OBSERVATIONS AND SIGNAL LEVEL CONTROL

In the IF system at the antennas, (Fig. 2), detectors are shown which measure the IF signal level in the 4-GHz-wide and 8-GHz-wide bands at the inputs to the optical transmitters. In the Baseband Converter (Fig. 4) and digital filter module (Fig. 5) a detector is shown which measures the level of the signal going to the digital sampler. These detectors perform two functions. First, they can be used for total power measurements. Those at the antenna would be used for wide-band continuum measurements. For total power measurements in spectral line mode the autocorrelation outputs of the correlators are required. Note that if SIS mixers with the full 8 GHz bandwidth prove feasible, then at the antennas it is possible to make total power measurements on all four 4-12 GHz outputs of an SIS front end simultaneously, thus making use of twice the bandwidth that is available in interferometer mode.

The second use of the detectors is to check the IF signal level at two points where it is particularly critical. One of these is the inputs to the optical transmitters in which nonlinearity can occur if the level is too high, and loss of signal-to-noise ratio can occur if the level is too low. The other point is the input of the digitizing sampler where it is necessary to know the rms input noise level in terms of the A/D reference levels. A variable attenuator or a variable-gain amplifier (not shown in the figures) will be included in the signal path ahead of each detector. These gain controls and the detector outputs will be accessible through the monitor and control (M/C) system. Thus various control schemes including adjustment in discrete steps, full ALC (automatic level control), sample-and-hold, etc. can be implemented through a control computer. If on-the-fly mapping or subreflector nutation are used in observing, then it may be necessary to measure and control the gain for a particular reference direction of the antenna beam.

5. PHASE SWITCHING AND FRINGE ROTATION

SIS mixers incorporating sideband separation are being developed for the MMA, but it is expected that it will be possible to achieve only 10-15 dB of isolation between the sidebands. This is sufficient to eliminate the noise from the unwanted sideband to a satisfactory degree. However, spectral dynamic range at least as high as 40 dB is desirable, so it is also necessary to include additional sideband separation at the system level. This is often performed by means of by 90 phase switching, but there are advantages to using a fringe-frequency offset scheme suggested by B. G. Clark. In either of these schemes phase adjustment must be applied to the first LO. In correlator arrays it is also generally useful to incorporate 180 degree phase switching to reduce effects such as unsymmetrical offsets in the sampler reference levels and spurious responses from unwanted signals that infiltrate the IF stages.

To explain the frequency offset scheme for sideband separation, consider first just two elements of the array. The fringe frequency can be brought to zero, that is the fringes can be stopped, by adding a frequency offset to the first local oscillator at one antenna, and a smaller frequency offset to a later LO. In the MMA we plan to use the 3.2-5.2 GHz LO shown in Figs. 4 and 5 for the second of these offsets. The offsets compensate for the variation of the delay between the time that a wave front from a source under investigation reaches one antenna and then the other. Suppose that, in addition to the fringe-stopping phase shifts we also add a frequency f_1 to the first LO of one antenna. Then the fringe frequency at the correlator output becomes equal to f_1 for both sidebands. Now if we subtract f_1 from the frequency of the LO with the second fringe-stopping offset, then the result will be that the fringe frequency at the correlator becomes zero for one sideband and $2f_1$ for the other sideband. If $1/(2f_1)$is large compared with the averaging time at the correlator output, the fringes will average to a very low residual leaving the full response for the sideband for which the fringes were stopped. The averaging can be made most effective by choosing f_1 such that there are an integral number of fringe cycles in an averaging interval. One can select either sideband to be the one that is retained by either adding or subtracting f_1, as appropriate, at the later LO. This is illustrated in Figure 6.

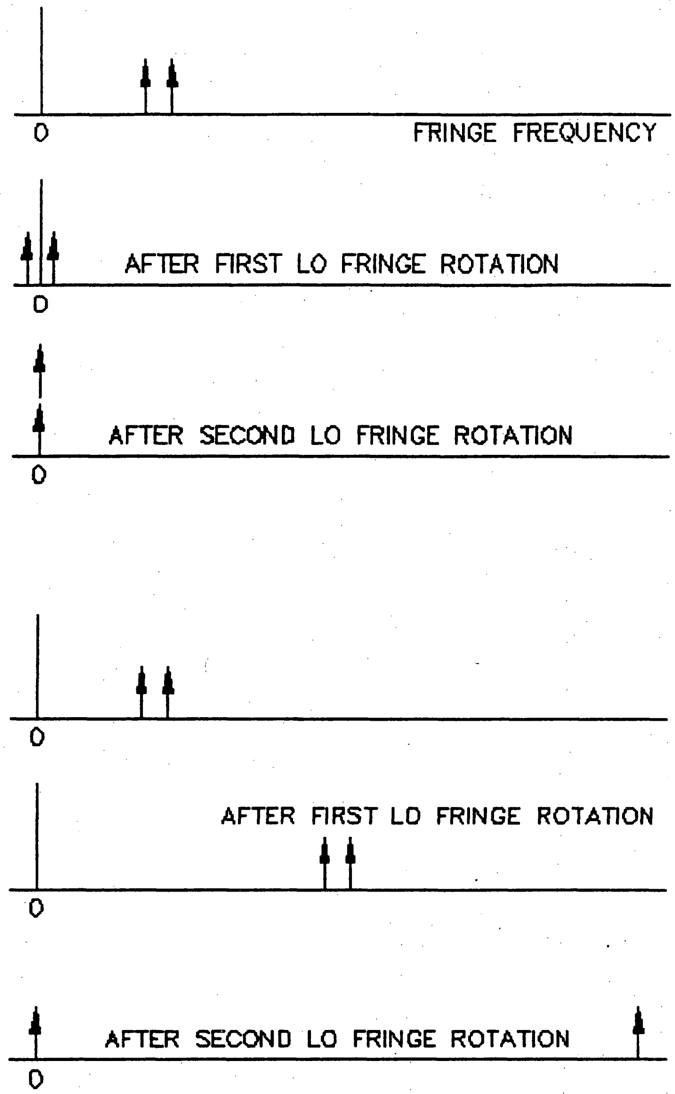


Figure 6: Fringe rotation and the sideband separation scheme

The upper half of the diagram shows how the fringes of both sidebands can be stopped in a double sideband system. The natural fringe frequency of the two sidebands is shown by the two arrows at the top. An offset on the first LO is then used to move the fringe frequencies so that they are centered on zero: at this point the fringes can be envisaged as the projection on the real axis of two contra-rotating vectors in the complex plane. An offset on a later LO brings the two fringe frequencies to zero. The lower half of the diagram shows how the fringe frequency can be brought to zero for one sideband and, for the other sideband, brought to a high frequency that is removed in the time averaging at the correlator output.

In applying the above scheme to an array of N_a antennas, the offset frequency must be different for each antenna. This is easily achieved by making the offset equal to N_1 for antenna N, where N takes values from 0 to (N_a-1). Although this sideband separation process results in the loss of one sideband, it fits the MMA requirement well because the SIS mixers that will be used will be designed to separate the sidebands, but will provide only 10-15 dB of rejection. The fringe offset scheme described here is required to increase the suppression of the unwanted sideband which has already suffered limited rejection at the mixer. Note that the reason for using the sideband separating mixers is that they suppress the noise from the unwanted sideband, which the LO offset scheme does not do. Thus the sideband separation in the mixer increases the sensitivity by a factor that depends upon the antenna temperature, and may be as high as two. The variable frequency LOs are required for fringe stopping, whether or not the above scheme is used for suppression of an unwanted sideband. Thus no additional hardware, but some software, is required to implement this method.

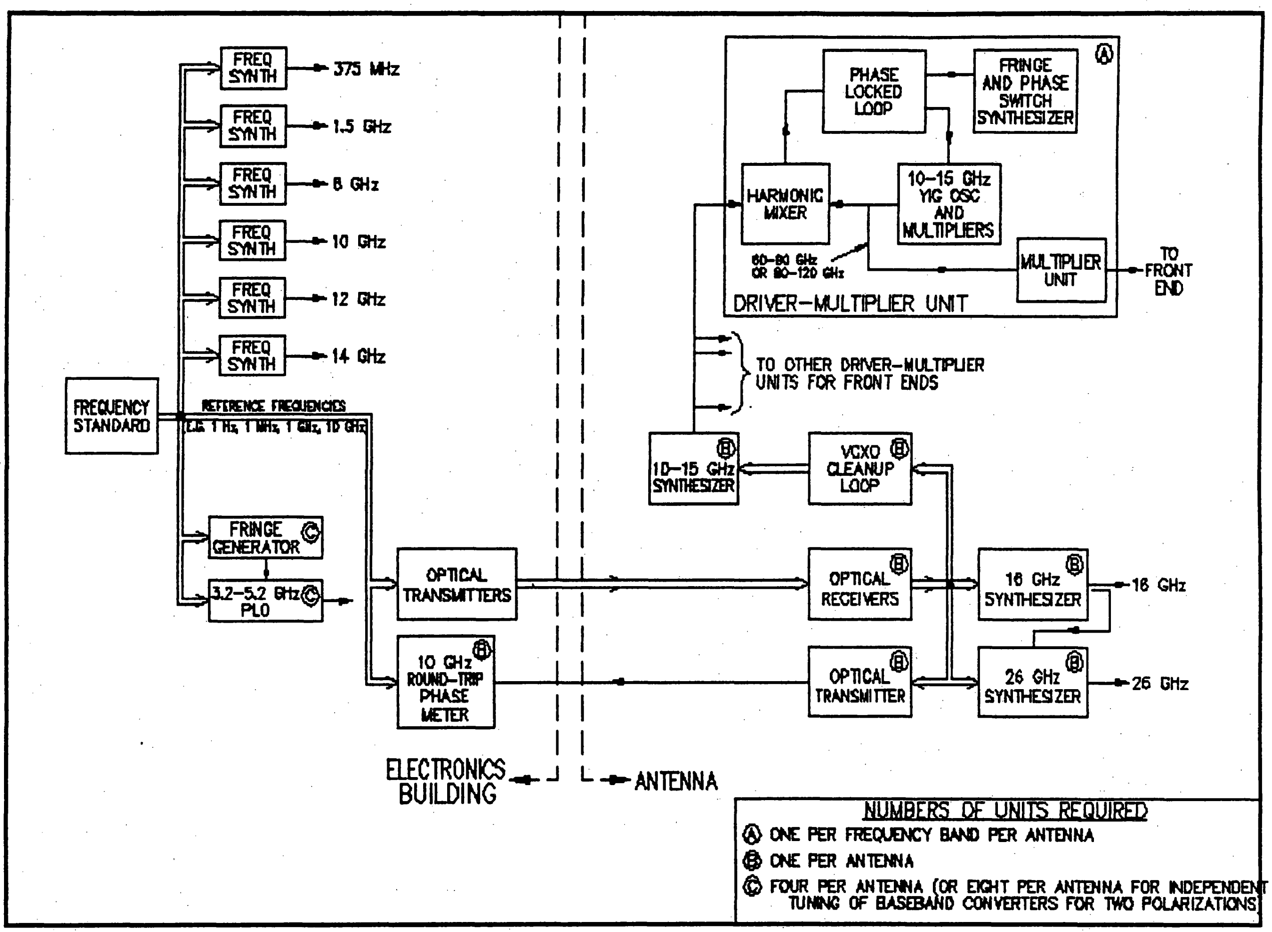
Phase switching by 180 degrees is required to eliminate responses to possible spurious signals and to correct for nonsymmetrical errors in the quantization thresholds of the samplers. It is necessary that the switching waveforms between any two antennas be orthogonal. Switching sequences based on Walsh functions allow this condition to be achieved with a smaller range of transitions per second than would be possible with squarewaves. The phase switching should be inserted at the first LO and removed after digitization by sign reversal in the 125 Mb/s data streams. The timing of the Walsh function transitions should be such that for the signals from all antennas the transitions are in phase at the input to the correlator. After the IF signals are digitized, a variable digital delay is used to compensate for the variation of the arrival time of a wave front at the different antennas, and this should be taken into account in the timing of the transitions when the Walsh functions are inserted at the antennas. Maintaining the correct timing can require some extra hardware and software when using very short averaging times and/or long baselines. If the phase switching is found to be necessary only to correct for non-symmetrical quantization thresholds, then it should be possible to replace it by automatic control of the threshold levels. For four-level sampling, the dc level of the sampled waveform would be adjusted to equalize the total numbers of positive and negative samples, and the quantization levels would be adjusted to equalize the numbers of positive and negative high-level counts. Counters for the four sampling levels would provide inputs to digital-to-analog converters to provide the dc levels. Such a scheme may be examined as part of the two-element prototype testing.

The LO fine tuning required for introducing fringe frequency offsets may also be used for fine adjustment for Doppler shifts, to allow data taken at different times to be combined. Thus the range of the offsets provided for fringe rotation should be as wide as the narrowest tuning interval provided by any other LO, to provide complete frequency flexibility. Alternatively, Doppler shifts can be removed by interpolation of the spectra at the correlator output.

6. THE LOCAL OSCILLATOR SYSTEM

A scheme for the local oscillator system based on the use of conventional frequency multipliers is shown in Fig. 7. A frequency standard provides inputs for various synthesizers and other units.

The double line between blocks indicates multiple signal paths carrying several frequencies, such as 1 Hz, 1 MHz, 1 GHz, and 10 GHz. The final values for these reference frequencies will be chosen to take account of the design details of the synthesizer units for which they are required. The synthesizers at the top left of the figure provide the two fixed frequencies (375 MHz and 1.5 GHz) required by the Baseband Converters (Fig. 4). These two frequencies would not be required with the digital filtering scheme in Fig. 5. The four synthesizers immediately below them provide the four LO signals required by the Demultiplexer units in Fig. 3. In the lower left part of Fig. 5 there are units designated 3.2-5.2 GHz synthesizers. These produce the 3.2-5.2 GHz LO signals with frequency offsets for fringe rotation and sideband separation that are used in the Baseband Converters in Fig. 4. Each of the 3.2-5.2 GHz Synthesizer units receives a signal from a fringe generator which uses a number-controlled oscillator or similar device to produce a signal at some convenient low frequency, such as 1 MHz, with the required offset for fringe rotation and sideband separation. This frequency is then combined with the synthesizer output.



Click on image to zoom

Figure 7: Local oscillator scheme using conventional frequency multipliers

The other part of the local oscillator system in Fig. 5 is concerned with the LOs required at the antenna. In the lower right area, fixed-frequency synthesizers at the antenna produce 26 GHz for the second LO for the HFET front ends and 16 GHz to convert 4-8 GHz IF bands to 8-12 GHz. The 10 GHz reference frequency is returned to the Electronics Building on a separate fiber to provide a round-trip phase measurement. The phase of the returned 10 GHz signal is measured with respect to the outgoing signal to monitor changes in the total path length of the fiber between the Electronics Building and an antenna. Thermally induced path-length variations usually result more from the fiber runs up the antennas than from the longer buried sections. It can be assumed with some degree of accuracy that all fibers going to an antenna will be in a similar environment and suffer similar thermal expansion.

In Fig. 7 a system of multipliers for the first LO system is in the upper right area. The unit labeled "10-15 GHz YIG oscillator and multipliers" (in the Driver-Multiplier unit) produces the 6^{th} or 8^{th} harmonics of a 10-15 GHz YIG oscillator (60-90 or 80-120 GHz), and this is amplified to a level of no less than 50 mW. This frequency is locked to the corresponding harmonic of a synthesizer that uses a fixed reference frequency which has been filtered by means of a phase-locked loop containing a crystal oscillator. The phase-locked loop for the YIG oscillator uses a phase reference that contains the required frequency offsets and phase changes for the fringe rotation, sideband separation, and phase switching. The 60-90 GHz and 80-120 GHz signals are sufficient to drive multipliers that can supply the first LO for any of the front ends above 60 GHz. For lower frequency front ends, an LO signal can be brought out from an earlier stage of the multiplier chains. Note that the frequencies in the multiplier chains are too high for switches other than waveguide type and these are deemed undesirable because of expense and also because experience shows that they do not always reset with sufficient mechanical precision, resulting in reflections and phase errors. Thus a separate Driver-Multiplier unit is required for each of the ten front end bands at each antenna. However, the output of the fringe and phase switch synthesizer, shown in the Driver-Multiplier unit, can be switched and shared between the Driver-Multiplier units at an antenna. It is expected that the multiplier scheme will be used for the two-antenna test array to be constructed during the development phase of the project.

A photonic LO system is being investigated as an alternative for the high frequency first LOs in the MMA (MMA Memo. 200). Two optical signals that differ in frequency by the first LO frequency are transmitted to the antenna on a single fiber. These are then combined in a photodiode resulting in enough power to drive an SIS mixer. The lasers that generate the optical signals for each antenna are located at the Electronics Building. If the photonic scheme proves practicable, it will greatly simplify the electronics since the requirement for frequency multipliers producing power levels sufficient to drive the mixers will be eliminated. Also, switching the LO signal to different front ends can be done using an optical fiber switch, and will eliminate the requirement of a separate Driver-Multiplier unit for each frequency band at an antenna. It has been suggested that a similar scheme should be used to produce correlated phase-calibration signals at the antennas. If this proves successful the round-trip phase measurement should be moved from the LO system to the calibration system.

7. DISPERSION IN THE OPTICAL FIBER

The optical transmitters for the IF signals will use external modulators to avoid causing frequency modulation of the laser. The optical frequency of the laser at 1300 nm wavelength is 2.3 x 10^{14} Hz, and a 2 GHz-wide IF signal spans 0.011 nm in wavelength. Assume that the laser wavelength is within 20 nm of the zero-dispersion point of the fiber, so that the dispersion is no more than 1.5 ps/nm.km. For standard fiber the zero dispersion point is near 1300 nm. Then for 25 km of fiber, which is approximately the longest fiber run from an antenna to the Electronics Building, the time difference for frequencies at the edges of a 2 GHz-wide band is no more than 0.41ps, which corresponds to 0.3 degrees of phase at 2 GHz. This is the effect of the dispersion on the intrinsic bandwidth of the IF signal. One can conclude that any resulting loss in coherence between an IF band that has traversed 25 km of fiber and one that has not can be neglected. Now consider the effect of variation of the laser wavelength, which is a function of temperature. If the laser wavelength changes by 0.013 nm, that is by 1 parts in 10^5, and again the dispersion is about 1.5 ps/nm.km, then the time for traversal of 25 km of fiber changes by 0.49 ps. This can be compared with the minimum increment in the compensating delay in the system which we will take to be 1/32 of the reciprocal bandwidth at the sampler (the same as in the VLA), which is 15.6 ps. Thus again the error is small enough to be tolerable. Note that if we used a laser wavelength of 1550 nm, the dispersion in standard fiber would be approximately 15 ps/nm.km, that is, a factor of 10 greater. For IF transmission the dispersion would still be just about tolerable, but one would not want it to be much greater.

The effects of variation in the effective length of the fiber on the LO signals must also be considered. A change in delay of 0.49 ps, resulting from a drift in the laser wavelength, corresponds to a phase error of 0.4 turns (144 degrees) at 800 GHz. The effect of a slow variation of this magnitude over several hours would be removed by frequent switching to a calibrator source, which is required in any case to correct for atmospheric effects. If one used standard fiber and a laser wavelength of 1550 nm the same laser wavelength drift would result in 4 turns of phase change at 800 GHz.

The loss in standard single mode fiber is about 0.35 dB/km at 1300 nm and about 0.2 dB/km at 1550 nm. Also, optical amplifiers are more readily available at 1550 nm. However, the dispersion is about 15 ps/nm.km at 1550 nm, whereas it is near zero for a laser wavelength of 1300 nm. Dispersion-shifted fiber in which the zero-dispersion point is shifted to 1550 nm is also available. An enquiry to Optical Cable Corporation reveals that the price of dispersion-shifted fiber is about 50% more than standard fiber, but that they could supply multi-fiber cable with, say, one fourth of the fibers of dispersion-shifted type, for a correspondingly smaller cost increase. Thus the advantage of using dispersion-shifted fiber for the local oscillator reference signals should be examined.

Variations in the effective length of a fiber can be monitored by a round-trip phase measuring system. Round-trip phase systems have been satisfactorily demonstrated in fiber optic transmission, using either the same fiber as the outgoing signal, or a separate fiber (Webber and Thacker 1990), for the returned signal. Here again the effects of the wavelength stability of the lasers and the dispersion in the fiber on the overall accuracy should be considered. Tests to determine the best way to design and implement a round-trip phase measurement to improve the instrumental phase effects for the MMA should be part of the development phase of the project.

8. ACKNOWLEDGMENTS

Members of the MMA Systems group who have contributed to the discussions on which this system plan is based include D. S. Bagri, J. E. Carlstrom, B. G. Clark, L. R. D'Addario, D. Emerson, R. P. Escoffier, P. J. Napier, S. Padin, J. D. Romney, R. A. Sramek, D. D. Thornton, J. C. Webber, and W. J. Welch.

9. REFERENCES

S. Padin, D. P. Woody, J. A. Stern, H. G. Leduc, R. Blundell, C.-Y. E. Tong, and M. W. Pospieszalski, An Integrated SIS Mixer and HEMT Amplifier, *IEEE Trans. Microwave Theory Tech.*, **44**, pp. 987-990, 1996.

J. C. Webber and D. L. Thacker, Phase Distribution on Fiber Optic Cable, *Report prepared by Interferometrics, Inc. for Naval Research Laboratory*, May 3, 1990.

Monitor & Control: Introductory Remarks

Darrel Emerson
Last modified July 16 1998

Introduction

The Monitor & Control requirements and specifications have not yet been defined in detail. The following two chapters (8.2b and 8.2c) outline two alternative approaches to the M & C system design. The precise choice of M & C system has implications throughout the electronic hardware and realtime software development of the MMA; reaching a timely decision on the choice of M & C design is a high priority.

MMA MONITOR AND CONTROL

Larry D'Addario
Last revised July 15 1998

1. DESIGN CONSIDERATIONS

This is a list of high-level issues that affect the choice of communication hardware and protocols; the type and distribution of computers; and the design of real-time control software. Each is followed here by some discussion. Decisions based on the answers to these questions are given in the next section.

1. What is the size/complexity of a "device" that must be separately recognized (addressed) and handled (programmed, monitored) by the master computer? By definition, we say that lower level devices, if any, are either not directly programmable nor monitorable, or they are handled by autonomous lower-level computers.

    Discussion: This is a crucial decision that must be made early, since it sets important aspects of the design of many devices, not just of the monitor/control system. It determines, for example, whether a given piece of hardware requires an embedded computer or instead can rely on a higher level device (or the master computer) to handle any complex control. If a local computer is required, the overall policy determines whether it can be a simple microprocessor running one fixed task, or a "real computer" that requires an operating system.

2. What data rates are required at each device? What is the worst-case timing accuracy requirement for delivery of a control signal to its destination?

    Note that it is not sufficient to have enough capacity to handle the long-term-average aggregate data rate. Each control signal must be delivered on time. If the requirements are tight, then one way to meet them is to have a hierarchy of control computers, where local ones have little to do but can meet close timing tolerances and higher level ones operate leisurely on big buffers. However, such a system is complex. At the other extreme, a single master computer might be able to meet all requirements. We cannot decide the appropriate arrangement until the requirements are known quantitatively.

3. What is the allocation of "intelligence"? Should we require some minimum level of computational ability in every device, or can some (or most) devices be allowed to be completely "dumb"? (These choices interact closely with those of item 1 above.)

    If the system design assumes a sufficiently high minimum level of intelligence in all devices, then it relieves all computers (except those imbedded in devices) of the burden of dealing with low-level, device-specific functions. This makes the design and maintenance of the software in those computers much easier. However, it imposes a heavy burden on the device designer to build the embedded processor and its software, no matter how simple the device might be. Since devices will be designed by various engineers at different laboratories, a wide variety of implementations may result, and all must be maintained.

4. How will development and maintenance be supported in the absence of a complete monitor and control system? Control signals must be provided and monitor signals recorded and displayed not only during normal operation of the array but also during development and

testing. An individual device must be testable in the laboratory without the master computer or any subordinate computer of the M/C system. A collection of devices forming a subsystem, or a complete antenna's hardware, should be separately testable without support from the master computer (which might be busy with software tests or otherwise unavailable).

This is a practical problem which has plagued several of our large telescope projects. There are several possible approaches. One is to provide duplicates of the M/C computer system and its software (all levels) at each development laboratory, and to provide at least one duplicate for each level at the site along with switching that allows the duplicate to be used for parts of the system and the main for other parts. Another approach is to support testing via separate computers (e.g., laptops) that need not have the same architecture, operating system, nor code as the main M/C system, but that support the same physical interface(s) to devices, subsystems, and antennas. Such computers can then be substituted for the M/C system whenever it is convenient to do so, and their software can be tailored to the testing requirements rather than to operational requirements.

5. The answers to the above questions should imply:
    a. For each device, how much embedded computing capability is required. This can be expected to vary widely among devices, according to their complexity.
    b. Whether intermediate-level computers are needed between the master computer and some devices, and if so how tasks should be divided between these and the master.
    c. The communications rates required, and therefore the kinds of physical links that would be appropriate.

However, there are new questions that will arise in this process. A device or logical function can be classified according to whether it is associated with one antenna, associated with a subset of antennas, or common to all antennas. Some things (like the correlator outputs) are organized by interferometer baseline; others may be associated with a subarray of antennas whose membership is time variable.

Comment: The proper handling of subarrays has been one of the most difficult issues in the design of the VLA and VLBA control systems. For the MMA, even greater flexibility is required. It is therefore important that this be taken into account early in the design.

6. What should be the topology of the communication network? It's assumed that the master computer will be in a control building of some sort. Should there be a separate path from there to each antenna (star configuration), or a single party line for all antennas (linear or ring configuration), or something in between? Should every *station* have its own connection, whether occupied by an antenna or not? Even if there is a separate physical path to each antenna, it is still possible to have a single logical network, such that all elements receive all messages but they respond only to those appropriately addressed. This approach requires much more communication bandwidth and is in this sense wasteful, but it may result in substantial simplification of software and some hardware.

2. STRAW-MAN LOGICAL DESIGN

In this section, we describe one way of organizing the MMA monitor and control (M/C) system.

We attempt to take into account all of the design considerations mentioned in the previous section, and to explain each design choice. In many places, the choices are arguable and alternatives are possible. This description omits many details, including the sizes and types of computers and the physical mechanisms of data transmission. The latter will be examined in the next section.

A. Organization Of Tasks

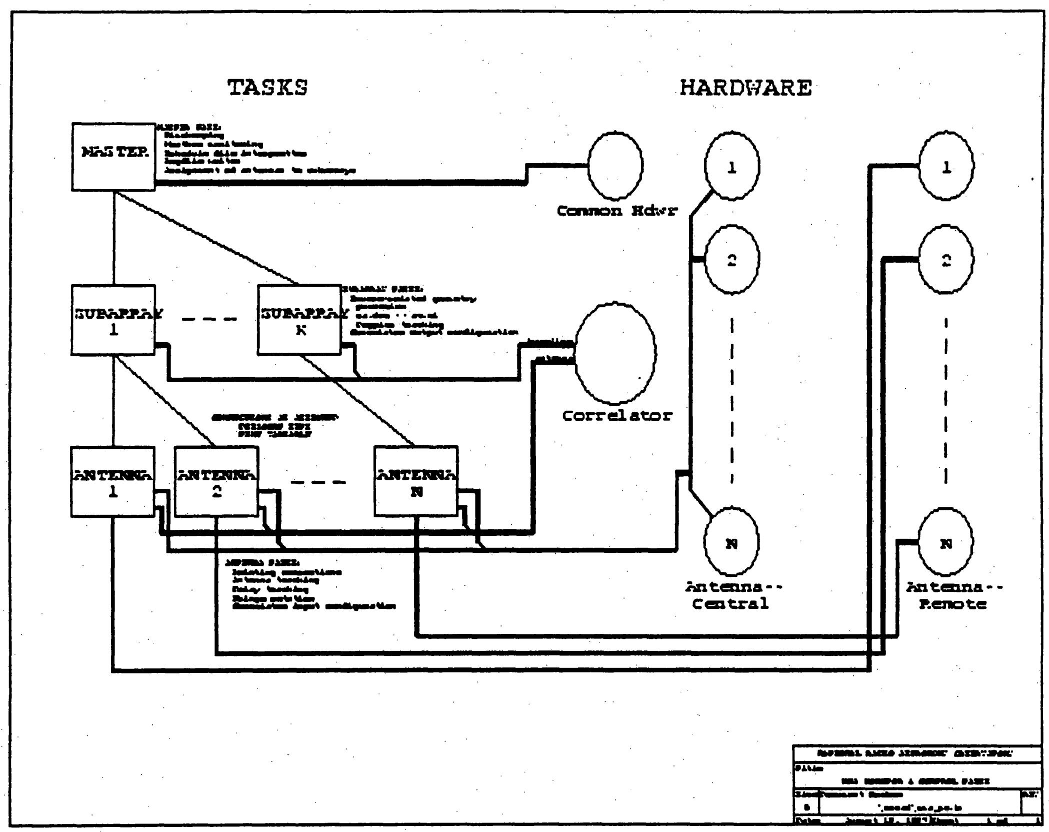


Figure 1. (click here for Figure 1 postscript version) shows a breakdown of the monitor and control system by tasks that must be performed in software, along with communication paths required between each such task and the hardware devices of the telescope. It is important to understand that this does not specify the assignment of these tasks to computers. They might all run in the same computer, or they might be distributed; if distributed, the computers might be in the same chassis or rack, or they might be spread out across the array. These things can only be decided after the required tasks and their interactions are understood, so that is all that Figure 1 is intended to show.

Tasks are organized into three levels: One Master Task, which is concerned with issues common to all elements of the array; a Subarray Task for each active subarray, which includes a group of antennas currently participating in the same observation; and an Antenna Task for each antenna. Some of the jobs performed by each type of task are listed in the figure.

The controlled hardware is divided into three classes: antenna-related, baseline-related, and common. "Antenna-related" means that there is exactly one copy for each antenna; some of these devices are located at the central building and some are located at the antenna. The correlator is treated specially; it contains the only baseline-related hardware, but it also contains antenna-related hardware (like delay lines). "Common" devices include such things as weather monitoring, central building environmental control, and the LO reference system. Here we assume as little as possible about the hardware details.

Some of the following is based on our experience with the VLA and VLBA, but the MMA has a special requirement that did not exist for those telescopes: total power (non-interferometer) observations. In this mode, measurements by accurate square-law detectors at IF or baseband must be sampled and recorded. In the design to be described here, this requirement is handled separately from the other monitor and control signals, and is discussed in section C below.

B. Design Choices

The following items are numbered the same as the Design Considerations in section 1 of this chapter, but they are taken up in a different order.

1, 3, 5(a): Device Complexity. "Devices" are allowed to have a wide range of complexity and built-in "intelligence." There is no requirement for some minimum processing capability, and many devices may be completely "dumb." By a dumb device I mean one that sets its state to that given in a coded instruction immediately upon receipt of the instruction, without further processing. A somewhat intelligent device might execute instructions at specified future times, or interpolate between instructions, or derive its new internal state from a combination of the present instruction and its current state. The degree of intelligence and its method of implementation are at the discretion of the device designer, except for certain cases that I'll discuss below. In all cases, device intelligence is considered "embedded" -- part of the hardware -- and therefore not part of the monitor/control system.

This design choice has some important consequences. First, device designers need not build fancy algorithms into their hardware unless there's a strong need to do so. They can depend on computers of the M/C system (including the master computer) for needed intelligence, limited only by available communication rates. Construction of simple, dumb devices is encouraged. On the other hand, use of embedded microprocessors (no operating system) for well-defined, repetitive local control and monitoring is possible without restrictions. Connections from such processors to their controlled hardware and the design of their firmware can be tightly tailored to the device's requirements without concern for general M/C protocols, and development of different devices can proceed independently in various laboratories. For very complex devices, in particular the correlator, a true computer (with

OS) may be embedded. Such a computer is still part of the device and not part of M/C; its software should conform to some project-wide standards, but all that I require here is a well-controlled interface to M/C.

With this approach, allocation of intelligence must be made on a device-by-device basis. A detailed examination of likely devices and their requirements is given in Table 1.

2(a): Data Rates. Data rates can be estimated from Table 1, and from the VLA and VLBA experience reported in Appendix A. These are summarized here:

	rates in bytes/sec			
	VLA	VLBA	MMA est (Table 1)	Proposed Capacity
Antenna control	125	200	21[3]	2000
Antenna monitoring	384	20	3[3]	incl
Central control	50[1]	100k[2]	231[4]	1M
Central monitoring	77[1]	2.3k[2]	10[4]	incl

[1] Not including correlator.
[2] Entirely due to correlator.
[3] Per antenna.
[4] For 40 antennas.

What is striking from these estimates is how low the data rates are. This will be discussed further below. The high control rate for the VLBA correlator comes from the possibility of complete reconfiguration every 2 sec; most of this traffic should be considered internal to the correlator. The monitor rate for the VLBA correlator is nearly all related to the tape machines. The proposed capacity for the MMA is at least 100x greater than the estimates, and at least 10x greater than required by either of the existing telescopes. How this capacity should be implemented is covered in a later report, but note that the antenna communication could be handled by one VLBA MCB per antenna, and that the central communication could be handled by an ethernet link to the correlator and one MCB for all other devices.

TABLE 1. DEVICES

		uP	Ctrl	Mon
1.1 At each antenna				
ACU	position, status	Y	10/1*	10/60
Subreflector focus adj (?)	maybe not implemented	N	4/120	2/600
Subreflector nutation control	maybe not implemented	Y	4/60	2/60
Cryogenics	mostly monitoring	N	2/rare	100/600
Front ends (approx 6 each)	mostly monitoring	N	12/rare	180/600
1st LO switching, monitoring		N	4/60	4/60
1st LO tuning	if conventional	N?	20/60	30/300
Fringe rotation, phase sw	if conventional	Y	10/1*	10/60
IF switching		N	4/60	4/60
2nd, 3rd LO	mostly monitoring	N	0	8/600
IF level attenuators, detectors		N	16/60	16/10
Optical transmitters	mostly monitoring	N	8/rare	16/600
General (environmental monitoring, safety, etc.)		N	0	32/600
1.2 By antenna, at central building				
1st LO tuning, status		N	20/60	30/300
Fringe rotation, phase sw	if photonic	N	10/1*	10/60

Cable length mon (round trip)		Y?	2/rare	4/10
Last LO tuning (BBC)		N	32/60	32/60
Bandwidth selection (BBC)		N	16/60	16/60
Digitizer mode	maybe part of corr.	N	16/60	16/60
1.3 Common, at central building				
Timing standard	mostly monitoring	N	0	20/600
Reference signal generation		N	0	10/600
Weather instruments		N	2/rare	80/60
1.4 Correlator				
1.4.1 Antenna-related				
Input configuration		Y	64/60	64/600
Delay tracking		Y	32/1*	0
1.4.2 Baseline-related				
Output configuration		Y	4K/60	4K/600

uP -> Device contains some intelligence in a microprocessor that we program. Does not include any uP that is part of a bus interface, nor any that is embedded in purchased hardware. In the case of the correlator only, it could be a "real computer" (with OS).

Each item listed is potentially a separate device, although in many cases several will be implemented in a single module or some similar assembly and addressed as one device.

Ctrl, Mon: [Total address space, bytes]/[period, sec]. For items that might be set for each change of observation, a period of 60 sec is used. Devices that must be updated periodically are marked * and are discussed separately. A minimum monitoring period of 10 min is used, even for devices where less frequent monitoring would be sufficient. There will be some special modes in which certain rates would be substantially different from these estimates, but not by enough to affect the design. All antenna-related rates are per antenna. Four IF channels and 8 baseband channels are assumed.

In both existing telescopes and in this design for the MMA, the communication rates are dominated by the periodic updating that occurs during an integration: antenna pointing, fringe rotation, and delay tracking. The traffic for setting up an observation and for routine monitoring is negligible (but consideration must also be given to meeting timing deadlines, as discussed below). The estimates are based on a complete new setup every 60 sec, but the resulting traffic would remain negligible if this were 10 times more frequent. For some specialized observations, some setup items (but never all) might have to be changed much more frequently, but this is unlikely to have a significant effect on the overall rate.

The data rates have been kept particularly low by specifying a period of 1.0 sec for each of the periodically-updated devices. This requires that those devices include the ability to interpolate/ extrapolate between updates, and that the update protocol provide sufficient data to allow this to be done accurately. Accuracy requirements are analyzed in Appendix B [not yet completely written, but where we hope to show that extrapolation via cubic polynomials is sufficiently accurate for all natural sources and for most satellites].

2(b): Timing Accuracy. It is not sufficient to have enough communication capacity to support the average data rates. In addition, we must ensure that no command is delivered too late and that no monitored value is sampled at the wrong time. One way to do this is to communicate only via large buffers, sending all commands well in advance and transmitting all monitor data long after its acquisition. If the buffers are large enough, then steady transmission at the average rate is sufficient. This puts all the burden for accurate timing on the devices, which must have separate knowledge of the absolute time and the ability to extract commands from the buffer and execute them at the correct times. While this is feasible for complex devices, and may be the only practical method for devices with very tight timing requirements, we choose not to require it of all devices, thus allowing many devices to remain simple. Instead, within reasonable limits, the monitor/control system accepts much of the responsibility for accurate timing.

Somewhat arbitrarily, we adopt the following specification: The monitor/control system (including computers, data transmission channel, and local interfaces) shall be able to deliver messages to devices within +-0.1 msec of known absolute times, and to deliver periodic messages where jitter in the period is no more than +-.05 msec. Receipt of such messages can be used to trigger execution of a command or sampling of monitor information. Devices for which this accuracy is sufficient need not implement any internal timekeeping. Devices requiring more accurate timing will need the ability to store messages and act on them at an internally-triggered future time. In addition, the M/C system shall be able to deliver a limited number of messages [number TBD] to different devices within +-2.0 msec of the same time. Devices which must be synchronized and for which this accuracy is sufficient need not implement any internal synchronization; those which require tighter synchronization must support it by separate timing signals. [The numerical values in this specification are subject to revision during the detailed design. The numbers given here are believed to be achievable without undue hardware or software complexity.]

3, 5(b): Distribution of Computers. In view of the low data rates, and also because some antenna-related devices are at the central building, we choose to assign all the tasks of Fig. 1 to computers at the central building. This means that there are no computers at the antennas except those embedded in devices. At present it appears that most devices at the antennas do not need an embedded processor, and that those that do (antenna drive, fringe rotator) need only a single-task microprocessor with no operating system.

We will not specify now the number or configuration of the central computers. There will certainly be a master computer running the Master Task, and perhaps it can run all the other tasks too. Individual processors are now very powerful; and robust, multi-tasking, real-time operating systems are readily available. A multi-processor arrangement with all processors in the same chassis and with high-speed interprocessor communication should be considered one computer.

This choice has some important consequences. It means that complex control jobs, such as optimizing the tuning of a receiver, must rely on algorithms implemented in the central complex rather than locally. But if the communication bandwidth is sufficient, this is a not a restriction; it provides valuable simplification of the software development work. Any complex control requiring bandwidth or having timing constraints beyond those of the M/C system would have to be embedded in the controlled device, but there seem to be no such cases. The VLA works well without any antenna computers, and the VLBA's antenna computers do not do any complex control.

In particular, control algorithms involving feedback are allowed, provided that they can operate within the M/C system's timing specifications. That is, the computer may generate a command whose value depends on recently-acquired monitor data. (Such algorithms were prohibited in the VLBA M/C design, primarily to eliminate the need for software timing specifications.)

4, 6: Development and Maintenance. Here we want to consider how an individual device or group of devices can be tested via its M/C interface when the complete M/C system is not available. This is especially important during development, when a single device may be the only part of the telescope on hand in a given laboratory. To answer this thoroughly requires that choices be made for the topology of the communication network(s) and the structure of the device interfaces, so we now specify some aspects of these.

At each antenna, all devices communicate via a common bus or network. Messages carry addresses to specify the affected device, and all devices see all messages. The exact type of bus will be considered in section 3, below. Each antenna bus has a separate physical connection to the central building.

At the central building, the antenna-related hardware (except perhaps that in the correlator) uses the same type of bus. There are several possible configurations. Figure 1 shows a separate instance of the bus for each antenna's remote devices, and another instance for all antennas' central devices. Another possibility is to have the central and remote devices of each antenna on the same bus. It is also possible to have a single logical bus for the whole array, with each message seen by everyone, but this would require a bus bandwidth of 2000*N B/sec to achieve 2000 B/sec to each antenna, or 80kB/sec for N=40 antennas; with addressing and overhead, a raw rate of nearly 2~Mbaud would probably be needed. If such a speed is possible at reasonable cost, then the central hardware and some software are simplified. However, tight synchronization becomes more difficult since two messages can never be simultaneous, not even to different antennas. Pending further study of possible bus types, we assume an independent connection to each antenna, both physically and logically.

For the correlator, a different interface is probably required. Whereas it requires a sophisiticated computer as its embedded processor, it can support a faster and more complex interface such as ethernet. Please note that we are only discussing the control/monitor interface here; the output data interface, where drastically higher rates are needed, is beyond the scope of this report.

Returning now to the question of operation independent of the M/C system, it should be clear that any computer having the appropriate bus interface can be substituted for the M/C system's computer(s) on any instance of the bus. Also, the presence or absence of one or more devices on the bus does not affect the ability to communicate with another device. In particular, a single device can be operated in the laboratory over its M/C interface by a computer that need not resemble the M/C computer used to operate the whole array. Moreover, an entire antenna can be unplugged from the bus and plugged into a local test computer; then all its devices can be operated from the local computer. This is the arrangement that I envision for development and for on-site testing and maintenance. The test computer is likely to be a portable PC (laptop), especially for field use.

Some programmers may object to the scheme just suggested on the grounds that using different computers and somewhat different software during development and operation leads to wasteful duplication of software effort and to confusion and errors by people who must switch from one system to the other. But experience shows that the opposite is the case. The requirements for laboratory testing and for troubleshooting software are often very different from those for operational software. Many engineers and technicians are quite capable of writing their own test software, and being freed from dependence on a programmer and from any constraints of the operational system allows them to complete their jobs much faster. Indeed, the need to control devices in a simple way in the lab will arise long before the need to run the whole telescope or even one antenna. Furthermore, with a little care and discipline, significant amounts of code can be shared between the test and operating environments (especially low-level functions that are tightly tied to the hardware design) without imposing significant constraints on either.

C. Total Power Data

The presently-favored technique of total power observing is on-the-fly data collection with the antenna scanning across the source as rapidly as possible. Some calculations indicate (F. Owen, private communication) that this would require sampling intervals as short as 0.4 msec at 800 GHz. At 12b per measurement and 2 channels per antenna, this implies a raw data rate of 60kb/s/antenna. If this were included with general M&C data, it would be the dominant traffic.

A sufficiently fast bus to accommodate all such data may be feasible, but there are major logical differences from general M&C data that make a separate transmission method desirable. The latter requires a polled, random-access protocol, with guaranteed delivery or at least detection of failures. The master computer selects devices and functions by address, sends short commands or monitor requests, and expects a response. This implies (a) some overhead for addressing, data integrity checking, and acknowledgements; and (b) for remote devices, a transaction rate limit caused by the round-trip time. (See Part III of this series for further discussion.) The total power data, in contrast, is a continuous stream in one direction; each measurement need not be separately solicited by the master. Indeed, such a master-slave protocol would prevent us from reaching the desired 0.4 msec sampling interval if the detectors are at distant antennas. (With 30km of cable and v/c=0.67, the transaction rate would be limited to 3350/sec, regardless of the raw bit rate.) Therefore, a different logical protocol should be used for the total power data.

The above considerations apply whether the rapidly-sampled total power detectors are located at the antennas or at the central building, a choice that depends on various system design considerations that are not yet resolved. If they are at the antennas, it is possible that the two logically-separate channels (general M&C and total power) might be multiplexed onto one physical medium for the link to the central building; this is a question for the detailed design.

3. PHYSICAL COMMUNICATION PATHS

In this section, we consider the physical mechanisms for transfering data between the monitor/control computers and devices.

It is possible to treat the various classes of devices (antenna, central, common) differently and to communicate with each by a different mechanism. The relatively large distance to an antenna is an important consideration. It is also possible to have a M/C computer at each antenna that communicates with devices there by one mechanism and that communicates with a central computer by a different mechanism. However, we begin by assuming that there is only one mechanism (type of interface) for all devices, and that there is only one, central M/C computer for real-time operation (although the latter might include an ensemble of processors).

Various published communication standards are examined for their applicibility to our requirements.

A. Design Considerations

1. **Distance.** The basic MMA will have antennas located several km from the central building, perhaps as much as 5 km. Cable routing could cause the longest run to be as much as 10 km. Extensions to the MMA may eventually involve much greater distances. This imposes some constraints on the kind of real-time control that is possible, regardless of the physical mechanism of communication. If timing precision smaller than the worst-case propogation time is needed, then we must either buffer transactions at the receiving end and execute them according to a precise local clock, or we must take account of the propogation time on the transmitting end. In protocols that require acknowledgment of one transaction before the next one is started, the rate of transactions is limited by the round-trip propogation time; unless long messages are used, this limits the data rate.

   We take it as a requirement that any antenna should be movable within a 20 km radius of the central building without any major modifications to its M/C system, including software. Allowing for a velocity factor of 0.67 in (optical) cables and a 50% increase for indirect cable routing, this gives a maximum propogation time of about 150 usec each way, limiting the transaction rate to 3350/sec

2. **Data Rates.** As shown earlier, average data rates for telescope control and monitoring are typically a few hundred bytes per second or less. In our present straw man design, the total rate for 40 antennas is 1201 B/s or 9608 b/s. Allowing 200% for addressing and other overhead (which assumes a very inefficient protocol), this becomes 30 kbaud for the entire array. For any one antenna, the rate with overhead is just 600 baud. More careful estimates will be needed for a real design, but these cannot be very far off. We should design for rates well above the final estimates in order to allow for errors, growth, accurate timing, and software flexibility. Even so, the required rates remain many orders of magnitude below those supported by modern data communication mechanisms.
3. **Control hierarchy.** In our application, it is essential that the central computer be the master of the array. This means that mechanisms and protocols that allow only one master and many slaves are the most appropriate ones. In some such systems (e.g., the IEEE 488 bus), it is possible for "mastership" to be passed from one device to another,

but we have no need for such a feature. Some also allow messages to be initiated asynchronously by a slave (typically generating an interrupt at the master, who must then respond); although we could use such a feature if it is available, doing so would increase the software complexity and it is clearly not required. This means that systems designed for "peer to peer" communication are not particularly appropriate, nor are master/slave systems with unneeded features. we do not exclude the possibility of adopting a system and not using some of its features, or using it in a simpler way than its designers intended; but the costs to us of its extra complexity should be fully evaluated and weighed against any benefits..

4. **Message length.** For nearly all of our devices, the required control and monitoring messages are intrinsically short. This includes, for example, antenna pointing updates (6 bytes), frequency settings (2 bytes), switch settings (a few bits), status monitoring (a few bits) and voltage monitoring (typically 12 bits). The natural time order of both command and monitor data involves skipping from one device to another, rather than having blocks of messages to or from each device. It would sometimes be possible to block the data by device so as to create longer messages; it would also be possible to establish device-specific protocols that update the complete internal state of the device each time or read out all available monitor data each time, resulting in longer messages. But each of these options increases the software complexity considerably. The ability to update or read out only one item (e.g., a switch setting or a voltage) without distrubing any others is an important software simplification.

   These considerations favor short messages, in which a few bytes of user data are transferred each time. Therefore, protocols in which the overhead per transaction exceeds a few bytes are not appropriate. Protocols that allow variable-length messages are useful and convenient, but not at the cost of significantly higher complexity or large overhead. Protocols with fixed-length messages are acceptable if the message length is appropriate (2 to 4 bytes seems ideal).

   On the other hand, there are exceptions. The correlator, especially, may require large blocks of data during the setting up of an observation. For this reason, it is worth considering a different mechanism for communicating with the correlator than that used for communicating with the antenna-related devices.

5. **Standardization and Support.** We prefer to adopt a communications mechanism and protocol that is already established, so as to minimize the design work that we must do ourselves. Various standards have been adopted and published by industrial groups and by organizations like IEEE and ANSI. By choosing one of these, we may be able to use hardware (chip level or board level) and software that is produced in large quantities and hence is available cheaply; and we may be able to avoid a large hardware/software development effort. In principle, we could also gain free or inexpensive support from user groups and vendors. This is to be contrasted with choosing a private mechanism and protocol (like the VLBA Monitor/Control Bus) where we are responsible for all development and maintenance. However, a private method remains under our control, and avoids our being forced into "upgrades" or changes imposed from elsewhere and based on considerations that may be irrelevant to us.

   No general statement about the best approach is possible. A detailed examination of each proposed mechanism/protocol is needed to evaluate its advantages and disadvantages. Published standards tend to be much more complex than we need because they are attempting to serve a wider range of applications; this is acceptable if the complexity has been mostly absorbed into existing chips and software, so that we can ignore it. But the process of tailoring a complex system to our needs, while drastically easier than developing the same system from scratch, may still be more work than in-house development of a system that is inherently much simpler.

B. Survey Of Possible Communication Mechanisms And Protocols

A large number of existing communication standards has been studied, and a brief summary of their characteristics is given in Table 2. The study has in no way been exhaustive, and it is entirely possible that the most appropriate existing standard has so far been missed. In most cases, the associated specifications have been initially drafted by ad hoc industry groups (or sometimes a single company) to meet the needs of a perceived market. This happens fairly rapidly (6-12 months), and the newest ones are still under development. After 2-3 years of discussions, if a reasonable concensus exists, the specification may be adopted and published by one of the standard-setting organizations like IEEE or ANSI.

Table 2: STANDARDS EXAMINED

Spec No.	Name(s) / info site	Data Rate User, max	CableLen max
IEEE P1394	FireWire, Serialbus www.firewire.org	100Mb/s	4.5m * 16
IEEE P1394.2	Serial Express www.scizzl.com/P1394.2	400Mb/s (future 1.2Gb/s)	?
ANSI/IEEE 1596	Scalable Coherent Interface (SCI) www.scizzl.com	200Mb/s	?
ANSI X3T11	Fiber Channel Standard (FCI) www.amdahl.com/ext/CARP/FCA	>1Gb/s	10km
IEEE 802.3	Ethernet wwwhost.ots.utexas.edu/ethernet	<10Mb/s	2.8km @5seg
IEEE 802.3z	Gigabit Ethernet www.gigabit-ethernet.org	~1Gb/s	2km
IEEE 1355	Std. for Homogen.Intercon. (HIC) stdsbbs.ieee.org/groups/1355	10-100Mb/s	NA
none	Universal Serial Bus (USB) www.teleport.com/~USB	<12Mb/s	5m * 6
NRAO 55001N1	VLBA Monitor/Control Bus (MCB) (see VLBA Tech Rep #12)	16kb/s (future 1Mb/s)	unlim.
ANSI 878.1	ARCNET www.arcnet.com	~2Mb/s	100km+

The remainder of this section will describe these existing standards. No attempt is made to cover all of them in detail. A brief statement is made about those considered unsuitable, along with the main reasons for that judgment. Then we give a more complete description of the three that appear to be the most promising: Universal Serial Bus, ARCNET, and NRAO-MCB.

B.1 Mechanisms Considered Unsuitable Foı MMA

B.1.1 IEEE P1394, "Firewire." Very fast, gross overkill for us. Length limit of 4.5m/hop prevents use for center-to-antenna links.

B.1.2 IEEE P1394.2, "Serial Express." Even faster than P1394, and might avoid length limit. Further study needed, but expected to be too expensive.

B.1.3 IEEE 1596, SCI. Designed for high-speed interprocessor communication by making remote devices appear to be memory-mapped. Goal is to keep latencies around 1usec, which requires >100MB/s raw speeds. This is really not a physical layer standard at all, as it can run over ATM or FibreChannel links. It's really an interface between a computer bus (e.g., VME) and any of various possible fast data transmission mechanisms.

B.1.4 ANSI X3T11, FibreChannel. Not examined carefully because it seems like gross overkill for our application. Having to provide 1Gb/s interfaces at hundreds of nodes would be too expensive to justify. Optimized for streaming large volumes of data, useful for disk drives. Said to use complex protocols with high overhead.

B.1.5 IEEE 802.3, Ethernet. Common Ethernet with 10Mb/s raw speed is well supported with chip-level and board-level hardware from many vendors, and it allows very flexible topologies. It is fast enough to have all devices of all antennas on one logical bus. However, its distance limit of 2.8 km is too small; and its method of resolving bus contentions (collisions) is statistical, so that latency is in principle unbounded. For real time control, we prefer bounded latency if possible, at least on transactions whose timing is important. These problems can be overcome, but at significant cost.

To achieve even the 2.8km distance requires 4 segments with 3 repeaters. The single-segment limit is 2km. This is the main reason that we can't use ethernet. These numbers apply to 10 Mb/s ethernet on fiber (10BASE-FL). For fast ethernet (100Mb/s), the maximum lengths are smaller. For Gigabit Ethernet, the standard is still in development, but in the present plan the longest single segment is still 2km. The distance limit could be overcome by creating a separate ethernet at each antenna and at the central building, then connecting each antenna net to the central net via fiber optic bridges. The bridges are expensive (~$1k each x N_ant x 2).

Also, Ethernet is designed for peer-to-peer networks, not master-slave as in our case. On the other hand, if we impose a higher-level protocol that requires all transactions to be initiated by a single node, then no collisions should occur and we have effectively a master-slave network. Ethernet itself does not guarantee delivery of any message; usually this is handled by higher level protocols. For our application, we would have to develop a hardware interface from ethernet to each device (e.g., a receiver or LO synthesizer), along the lines of the MCB Standard Interface. This would be complicated and expensive if it must support a general purpose protocol like TCP/IP, so we would probably create our own, much simpler protocol. We end up with a proprietary system (hardware and software) even though it is built on top of "standard" Ethernet. we conclude that the expense and complexity are too high, especially since there is no benefit relative to alternatives.

B.1.6 IEEE 1355, HIC.

B.2 MECHANISMS MOST PROMISING FOR MMA.

B.2.1 Universal Serial Bus (USB). [to be written]

B.2.2 ARCNET. [to be written]

B.2.3 NRAO-MCB. [to be written]

C. Conclusions And Recommendations

[This is a brief summary, to be expanded into much more detail.]

The NRAO-MCB, developed for the VLBA, is presently the best choice for MMA monitor and control communication. The standard interface needs to be re-implemented with modern components, and in the process various design improvments can be made. In particular, it is probably possible to double (or perhaps quadruple) the data rate from the present 1000 transactions per second (where one transaction transfers a 16b monitor word or command word); nevertheless, the present rate appears sufficient to support the entire MMA on a single bus, except for the correlator.

A few alternatives to the MCB are possible, including ARCNET and USB. Each of these has problems, but there might be acceptable solutions. They should be investigated in more detail before a final decision is made.

The correlator must have its own embedded control computer, and that computer can be expected to be fairly powerful and to be running a real-time operating system. Its monitor and control interface to the master computer, assumed to be in the same building, should use a general-purpose high-speed LAN such as Ethernet-TCP/IP. (Correlator output data is something completely different and is not considered here.)

High speed data from the antennas, especially total power detector readings, should be streamed to the central building on a dedicated link. This is logically different from all other monitor/control, since the data stream is generated autonomously rather than solicited by the master computer. The data is either ignored or recorded, according to the telescope's mode. The data should be treated as back-end output data, similar to the correlator's output, rather than monitor/control data.

APPENDIX A: DATA RATES FOR SOME EXISTING SYSTEMS

A.1 VLA

Pointing: 1 command/antenna/WG_cycle	24b*19.2Hz =	460.8 bps
Fringe rotation: ditto		460.8
Initialization: each scan, as needed	negligible	

Operator: manual commands, as needed negligible

Total command rate used ~1000
Total capacity: 4 to antenna, 2 to CB devices
per antenna per WG_cycle, 24 data bits each 2764.8

Monitor: 1 word/data_set/WG_cycle; 32b word;
max 5 DS at antenna (all used), 3 at CB (1 used);
word = 2*12b analog + 8 bit digital.
Total monitor data rate 6*32b*19.2Hz 3686.4
Total capacity 8*32b*19.2Hz 4915.2

For 28 antennas: Command rate used 28 kbps
Monitor rate used 103 kbps

All I/O was originally done by a single Modcomp II, using about 2/3 of its capacity (K. Sowinski, private communication).

A.2 VLBA STATIONS

No fringe rotation, but tape recorder control and monitoring are needed. Data rates still dominated by antenna pointing:

Az,El commands at 20 Hz rate, each requires
2 16b MCB transactions. 64b*20Hz 1280.0 bps

Operator interfaces, typ 10 trans./sec 160

Approx. 300 monitor addresses, average sampling period
30 sec (1 sec to 1 hour), 16b each 160

MCB capacity: 16b/transaction (command or monitor),
1.7 msec per transaction achieved at VLBA 9411

(Therefore, an estimated 25% of achieved capacity is used. Readout of fringe check buffer is limited by MCB capacity, but is done only occasionally.)

Theoretical MCB capacity:
56 kpbs raw, 5 bytes/transaction incl overhead,
11 bits/byte (55 total bits for 16 user bits) 16.3 kbps
(This should be achievable with a better interface implementation.)

A.3 VLBA CORRELATOR

MCB is used for control and monitoring of the 24 playback drives, but not for closing the tape speed loop.

HCB (invented for the VLBA correlator) is used for other control data, including configuration loading and geometric model loading into low-level microprocessers. It is also used for acquiring playback system status data:

HCB capacity 1.0e5 B/s
=========

HCB use:
Model downloading: 1080 B/station.
Typ 120s/download (ground VLBI) 180 B/s

```
                10s/download (space)                  2160
        Max    5s/download, all 20 stns              4320
              0.5s/download, 1 stn
    Re-configuration:  ~1e4 B/station worst case.
        Beam switching or band switching might
        require reconfig every 2s ->                 2.0e5 B/s
    Playback drive monitoring:  1156 B/stn.
        All 24 drives checked each 12s                2312 B/s
```

If rapid reconfiguration is needed, it dominates the traffic and can exceed the bus capacity in the worst case. Otherwise, the traffic is always below 7% of capacity, and usually below 3%. (Data from Steve Blachman.)

"Model" is separated into delay and rate parts. Four "delay models" are sent, one for each potential delay center; and 8 "rate models" are sent, one for each baseband channel. Each of the 12 requires 90 bytes.

APPENDIX B: ACCURACY REQUIREMENTS FOR ANTENNA, DELAY, AND PHASE TRACKING

A.1 Antenna Pointing

The VLA and VLBA antennas operate on position commands via a Type II servo. Each commanded position is considered to be effective immediately, and the servo drives the antenna so as to keep the error as close to zero as it can. With such a system, if position commands are sent at a rate much slower than the loop bandwidth, then the actual position will be nearly piecewise-constant (stair step tracking); and if commands are sent much faster than the loop bandwidth, then the position will change smoothly. In the latter case, the position error will be held very near zero for any constant velocity within the drive system's speed range, but will be proportional to acceleration according to delta a = $(d^2a/dt^2)\ T^2$ where T is the servo's closed-loop time constant.

Accurate tracking then requires that updated position commands be sent whenever the position changes by the allowed tracking error (provided that stair step tracking is acceptable), or whenever a small fraction of T has elapsed, whichever is slower. In the latter case, it is also necessary that T be small enough that the acceleration error does not exceed the allowed error. For sidereal rate tracking, the worst-case acceleration error is $(.001\text{arcsec/s}^2)T^2$ or only .001 arcsec at T=1s; it is therefore negligible even for rather slow servos, and for this reason many radiotelescope antennas can use time constants ~1s or slower. This is the case for the VLA and VLBA, which then use update rates of 9.6 Hz and 20 Hz, respectively.

However, for the MMA much faster servos will be needed to support rapid position switching. The antennas are expected to be quite stiff, with minimum strutural natural frequencies ~12 Hz. It may then b possible to design a stable servo with a time constant as small as 0.1s, so that a strategy of sending position updates many times per time constant requires an update rate of 50 to 100 Hz. I assume that this will be done, but now consider partitioning the work of computing these updates among available computers.

Consider tracking at 10 times the sidereal rate (7.272e-4 radian/s) and requiring each update to be computed to an accuracy of 1/20 of the beamwidth (3.125e-6 radian for an 8m antenna at 600 GHz). If the angle vs. time function is represented by its truncated Taylor series over each interval of length tau, then Taylor's theorem shows that a first-degree polynomial (using the value and 1st derivative at the beginning of the interval) gives an error less than (2.7e-7 rad).tau and a 2nd degree polynomial give an error less than (6.41e-11 rad). tau^2. Thus, the required accuracy is achieved for tau<12s using a first degree polynomial and for tau<221s using a 2nd degree polynomial. The ratio of the required polynomial update interval to the servo update interval is thus several orders of magnitude, justifying a two-level scheme of polynomial computation and polynomial evaluation.

We therefore adopt the following straw-man design. The antenna controller is a device that receives, from the central computer over the common bus, an updated position and its derivative for each axis every tau seconds. It also receives a separate signal that provides an accurate absolute-time mark every tau seconds. Each update of position and derivative is computed for the time corresponding to the *second* time mark after it is received. Between any two time marks, the controller evaluates the (unique) 3rd degree polynomial that matches the received data for those time marks; the evaluation is done at the servo update rate, approximately 100 Hz, and the results are supplied as position commands to the servo.

The above calculations show that this design would support accurate tracking at 10 times sidereal with tau>>100s. But such a value of tau would be too large for many purposes, including rapid position switching, because no updates are possible more frequently. We therefore adopt tau=1s. This allows extremely accurate tracking and an arbitrary change of trajectory every 1s, and it allows the bus data rate from the central computer to each antenna to be kept low.

Other schemes are possible, and some might be somewhat simpler. For example, sending only the position every tau seconds and requiring spline-function interpolation in the controller would work, but then trajectory changes would be effective only after the time span of the spline (at least a few times tau). Another possibility is to do the interpolation in the central computer and to send position updates to each controller at the servo update rate, thus simplifying the controller. The bus rate is then much higher, but there exist bus designs that could support it. However, the antenna controller probably needs an embedded microprocessor for other reasons, and giving it the interpolation job is a negligible added burden, so that is the approach selected here.

A.2 Delay Tracking

[to be written, along similar lines]

A.3 Phase Tracking (Fringe Rotation)

[to be written, along similar lines]

Alternative MMA Monitor and Control

Jeff Hagen
Last revised July 15 1998

The previous chapter describes the problem of Monitor and Control at the MMA well.

Here is a summary in simple terms:

- The paper suggests a master/slave design where each antenna is specifically controlled from a central computer.
- This might be accomplished by a tree network of custom designed hardware.
- The paper suggests a relativly low overall baud rate, 2k bytes/sec but a relativly high timing accuracy, 100 usec +- 50 usec.
- It examines a host of high speed digital protocols suggesting a direction to develop the appropriate master/slave protocol.

That paper proposes good arguments for such a solution:

- We did it that way before on the VLA and VLBA so we know it will work and we have the benefit of experience.
- If we build our own, we will get exactly what we specify and have full control of the design without the effects of industrial trends.
- If we build our own we will save money because we will only build what we need and not need to pay for extraneous hardware and features.
- Past experience suggests ease of development in the laboratory.

Proposed here is an off-the-shelf solution which meets the above constraints. I will argue each point and then make a specific suggestion.

- We did it that way before on the VLA and VLBA so we know it will work and we have the benefit of experience.

Times have changed. In the days of the VLA design there was no other cost-effective solution. General purpose computers were big and expensive. Microprocessors were limited and cumbersome. What was done was a wonderful piece of engineering. General purpose technology was nowhere near being able to meet the needs. A few pieces were there so NRAO made a development project to build the custom hardware required to meet the specification. It was a tough job which pushed the limits of existing technology. As I will show, today's off-the-shelf technology can easily meet the needs of the MMA monitor and control. Its not a hardware development project. We simply need to buy the right standard equipment and hook it up.

- If we build our own, we will get exactly what we specify and have full control of the design without the effects of industrial trends.

This is true. Along with that means we will not have options toward upgrade. We won't be able to change our mind in the future. We will never be able to to relax the master/slave constraint. Any change will force a full redesign. I hope we can think of everything now. Do we really want to make inflexibility part of the design? Why design in hardware constraints that might be advisable but not necessary? How about parts? I understood parts availability was an issue at the VLA. It certainly is at KP. With general purpose computers and parts, changes and upgrades are easier. We need expertise to pick the correct devices but that is quite a bit simpler than designing our own from scratch.

- If we build our own we will save money because we will only build what we need and not need to pay for extraneous hardware and features.

This is just not true. There is no way you can argue against the economies of scale. Using off-the-shelf equipment will get us microprocessors that are too fast and interfaces that we might never use. There is no reason to scale this down. You cannot build and redesign custom hardware more cheaply than you can buy mainstream equipment. In fact the reverse may be true depending on the device. Also, engineering talent stays in the mainstream too. We can more readily hire people with experience. We should only build our own when we push the technology envelope. The correlator is a good example of this. M&C is not.

- Past experience suggests ease of development in the laboratory.

There is a huge need for engineering test facilities. For the first time, the laboratory test computer can be the same off-the-shelf computer with the same interface as the production equipment. With an MCB style interface, engineers get to fight over the use of the test box, or they have no option but to build their own test chassis creating more custom use-once electronics. Using general purpose computers, engineers can build and test their hardware in an environment they understand. Rather than building a custom test chassis, only an additional off-the-shelf interface is needed. The software they run can be anything, it does not need to depend on the final production software.

The proposal.

My solution is simple. The general purpose computer I would choose is an industrial embedded PC. Run an additional telephone line to each antenna and install a clock for timing (more below). The computer would commuinicate via dedicated modem. At the control building, the master computer talks to a standard array of terminal servers with modems. I could spec-out the current logical choice but it doesn't matter. The overall baudrate spec of 2k bytes per sec is less than current modem speeds. This is mainstream everyday technology. It has every capability that the homemade design has, but it is totally expandable and changeable. It does not require a high-level research project.

Timing is an important issue. Running NTP (Network Time Protocol) to set the computer's clock would only be accurate to tens of milliseconds and is not within the specification. How about a GPS receiver? Microsecond accuracies are easily attainable with very cheap equipment, and higher accuracies with more work. This is new to arrays, but this is easy, cheap, mainstream technology. $1100 per node would produce what I have in mind. Look at: the 'Totally Accurate Clock' (TAC) by Dr. Tom Clarke, or TAC2 or Model 100 GPS Clock for a detailed techincal description.

This is a highly debatable issue. Should GPS technology be deemed too unreliable, we would be forced, out of true necessity, to build an accurate distributed timing pulse. Such a device would be a small subset of the full blown reinvention of networking technology proposed in the previous paper.

The proposed computer network does not guarantee any message transaction speed. Without any clock or timing pulse the the network would be limited to accuracies of 100's of milliseconds (or longer) in response. We need to carefully design the software to remove the rigid time constraints from the telescope commands. A good example of this would utilize the Taylor series to approximate a polynomial for azimuth and elevation control as suggested in the previous paper. (This idea is at the heart of the 12m servo as well.) There is no need to go crazy with this idea. For slow devices or events, command/acknowledgement will work fine.

Now we have a computer at each telescope. Up to here, I have basically proposed a modern version of the VLBA system. Let me go further. The VLBA uses the MCB for intra-antenna M&C with the master in the control building with the clock and telephone. Once again there is no need for a homegrown MCB bus as there might have been a few years ago. Connectivity is standard cheap and easy. Depending on complexity of the device, a "controller" could be another embedded PC on an intra-antenna ethernet. It might be a set of digital IO signals or analog voltages or a serial port. We should choose a standard technique for each interface but the hardware design should not force a solution, it should be the result of our own discipline in making careful choices.

Consider the future. The network I propose is the cheapest possible off-the-shelf solution. Like a lot of areas, fiber optic technology is expanding. I presume that if we use fiber optics in other areas of the array there would be free fibers at each antenna. It would then be possible to upgrade our connection speed and response at a minimum cost and disruption. It might be that this is irrelevent and makes no real advantage. The point is we would have a considered option.

This proposal makes maximum use of existing modern technologies. It allows for unknowable future expansion. We should limit the use of custom hardware to solutions that are not possible any other way.

APPENDIX

Definition of Requirements

Jim Schroeder
7/15/98

Before any solution can be prescribed, requirements must be clearly defined. At present, not all parameters are defined. For the moment, let?s concentrate on communications required to control and monitor the antennae themselves. By looking at information gleaned from MMA Project Book, Chapter 8, Section 2, and Chapter 4, we can assume certain communications are required for operation of MMA antennae. The following tables give an estimate of the number of bytes needed to be transferred for each operation, and the typical period between such transfers. The final row of each table is a simple summation of bytes for all transactions, with the minimum transfer rate of the table. This then gives the absolute upper limit of the possible monitor or control data transfer rate.

Item of Interest to be Controlled	Bytes	T (Sec) update rate
ACU (position and rate)	10	1
Subreflector focus adjustment	4	120
Subreflector nutation control	4	60
Front ends (x6)=	12	600?
1st LO switching, monitoring	4	60
1st LO, tuning	20	60
Fringe rotation	10	1
IF switching	4	60
IF level attenuators, detectors	16	60
Optical transmitters	8	600?
Cable length	2	600?
Last LO tuning	32	60
Bandwidth selection	16	60
Digitizer mode	16	60
Weather instruments	2	600
Antenna input configuration	64	60
Antenna output configuration	4096	60
Delay tracking	32	1
Power distribution switching	2?	60?
Rcvr cabin HVAC temp set point	1?	60?
Total bytes of control data/minimum time (sec)	4355	1

Table 1. Items to be controlled at each antenna. Note that numbers followed by question marks are guesses.

Item of Interest to be Monitored	Bytes	Update period T (Sec)
ACU position and rate (velocity and direction)	10	60
Motor currents	1?	1?
Servo power supply voltages	1?	1?
Subreflector position	1?	1?
Subreflector focus adjustment	2	600
Subreflector nutation control	2	60
Limit switch status	1?	1?
Cryogenics	2	600?
Front ends (x6)=	180	6
1st LO switching, monitoring	4	60
1st LO, tuning	30	30
Fringe rotation	10	60
IF switching	4	60
2nd LO	4	60
3rd LO	4	60
IF level attenuators, detectors	16	10
Optical transmitters	16	60
Environmental and HVAC performance	32	60
Cable length	4	10
Last LO tuning	32	60
Bandwidth selection	16	60
Digitizer mode	16	60
Weather instruments	80	60
Timing standard	20	60
Reference signal generation	10	60
Antenna input configuration	64	60
Antenna output configuration	4096	60
Total bytes of monitor data/minimum time (sec)	4641	1

Table 2. Items to be monitored at each antenna. Note that numbers followed by question marks are guesses. Apparently, the 2nd and 3rd LOs will also be monitored, but no data size or periods are available. The values in this table are guesses.

LO, IF and SIGNAL TRANSMISSION SYSTEM

Dick Sramek
Last revised May 27 1998

Below are the base level specifications for the IF system. Values for many of these parameters will only be available after the next level of design is begun. Many additional specifications will be developed as the design progresses.

1. General

The input signal interface for the IF system is at the receiver package. One or more stages of frequency conversion will take place in the receiver package. The output signal interface of the IF system is at the digital samplers.

Top level block Diagram: Figure 9.x

2. System elements

Number per antenna

IF bandpass selector 2

FO Transmitter 2

FO Receiver 2

IF demultiplexer 2

Baseband converter (analog) 8

At this time, these system elements do not necessarily correspond to hardware modules.

3. IF bandpass selector

Description: Provides total power measurement of the RF bandpass and places two 4 GHz bandpass signals in the 4 to 12 GHz band.

Block Diagram: Figure 9.x

Number of input signals: 2

Input Frequency range: 4 to 12 GHz or 4 to 8 GHz

Input Level: ?? dBm to ?? dBm

Number of output signals: 1

Output Frequency range: 4 to 12 GHz

Output Level: ?? dBm to ?? dBm

Image rejection: ?? dBm

Detector 1, Freq range 4 to 12 GHz

Detector 1, input power ?? dBm to ?? dBm

Detector 1, output voltage ?? mvolt

Detector 2, Freq range 8 to 12 GHz

Detector 2, input power ?? dBm to ?? dBm

Detector 2, output voltage ?? mvolt

Detector 3, Freq range 4 to 8 GHz

Detector 3, input power ?? dBm to ?? dBm

Detector 3, output voltage ?? mvolt

Square law linearity

LO frequency 16 GHz

LO power ?? dBm

4. **FO Transmitter**

Description: Input signal goes to ALC and modulates digital FO transmitter.

Block Diagram: Figure 9.x

Number of input signals: 1

Input Frequency range: 4 to 12 GHz

Input Level: ?? dBm to ?? dBm

Output wavelength: 1550 nm

Output level: ?? dBm

Spur free dynamic range ??

5. **FO Receiver**

Description: Input signal on FO cable goes to pin diode receiver and amplifier.

Block Diagram: Figure 9.x

Input wavelength: 1550 nm

Input level: ?? dBm

Output Frequency range: 4 to 12 GHz

Output Level: ?? dBm

Receiver noise figure: ??

6. **IF demultiplexer**

Description: Input signal is split, filtered, and down converted into four 2 to 4 GHz sub-bands.

Block Diagram: Figure 9.x

Input Frequency range: 4 to 12 GHz

Input Level: ?? dBm to ?? dBm

LO frequency 16, 10, 12 and 14 GHz

LO power ?? dBm

Number of outputs 4

Output Frequency range: 2 to 4 GHz

Output Level: ?? dBm

7. **Baseband converter (analog)**

Description: Input signal is 2 to 4 GHz which is mixed with LO which applies fringe rotation. Output is final bandwidth sent to the digital samplers.

Block Diagram: Figure 9.x

Input Frequency range: 2 to 4 GHz

Input Level: ?? dBm to ?? dBm

LO frequency, with fringe rotation 3.2 to 5.2 GHz

LO power, with fringe rotation ?? dBm

Secondary LO frequency 1.5 GHz and 375 MHz

Secondary LO power ?? dBm

Output frequencies 0.1 to 2 GHz

1 to 2 GHz

1 to 1.5 GHz

250 to 500 MHz

250 to 375 MHz

62.5 to 125 MHz

62.5 to 93.75 MHz

Output power levels: ?? dBm

MMA CORRELATOR

John Webber
Ray Escoffier
Last revised July 15 1998

1. Introduction

This section describes the proposed correlator for the MMA. The design described here is for a lag correlator with a system clock rate of 125 MHz. The system architecture described has been chosen as the best tradeoff to produce high reliability, robust operating margins, a minimum number of integrated circuits, and a minimum number of cable interconnects (see MMA memo 166).

The correlator system envisioned for the MMA includes; the samplers (digitizers), possible digital filters, mode selection, delay lines, a data format conversion stage, cross- and auto- correlators, long term accumulation, and initial digital computer processing. Depending on the mode of operation, the output of the correlator could be in either the lag or frequency domains.

2. Specifications

This section gives a summary of the MMA correlator specifications.

```
  40 antennas (number may change)
   8 baseband inputs per antenna
   4 GHz maximum sampling rate per baseband input
   2 bit, 4 level sampling
  30 KM maximum baseline delay range
1024 hardware cross correlators (1024 lags and 1024 leads) per baseline
1024 auto correlators per antenna
   4 product pairs (RR, RL, LR, LL) possible for polarization
```

Bandwidths per baseband input range from a maximum of 2 GHz down in factor of 2 steps to 31.25 MHz. For 8 baseband inputs per antenna, this yields a maximum bandwidth per antenna of 16 GHz.

3. System Block Diagram

A simplified block diagram for the MMA correlator is given in Figure 1. This diagram presents a fairly conventional lag correlator except for the presence of the data format conversion stage.

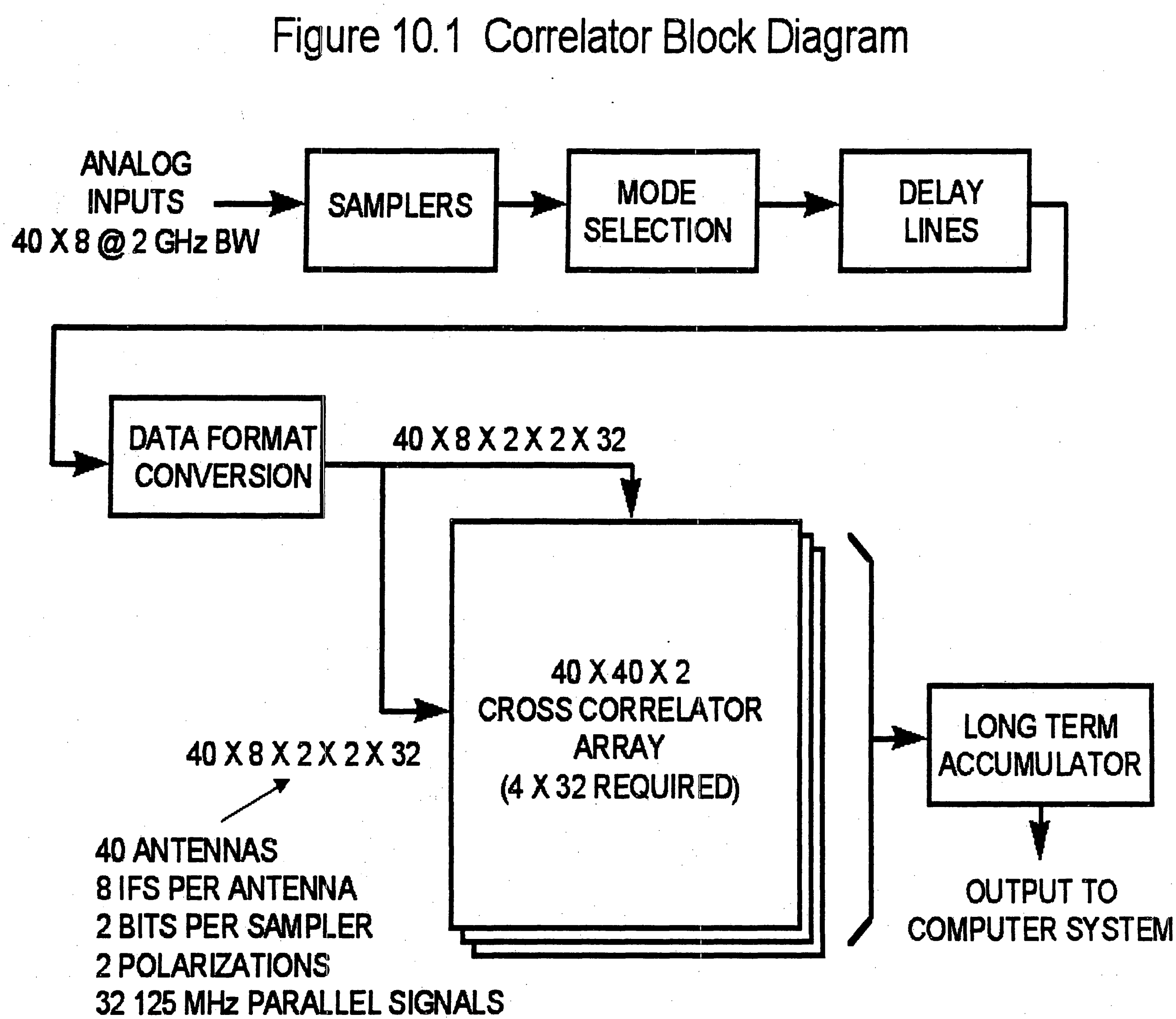


Figure 1: simplified block diagram

The analog outputs of the baseband system drive sampler inputs where 2-bit, 4-level sampling is done at 4 GS/second. If a digital filter is employed rather than analog baseband filters (see section 8), then the digital filter is conceptually included in the "sampler" block of figure 1.

Logic in the mode selection block routes sampler outputs into the delay system. When fewer than 8 samplers per antenna are being used, this stage will assure high system efficiency by replicating active sampler outputs into unused delay lines and hence into otherwise unused correlators where additional lags can be generated. In this way, maximum performance will be obtained for the observational mode desired.

Next, delay lines are provided to time the signals from the array antennas. The delay will be provided in very efficient high density RAMs. For a 30 KM delay range, 524,288 RAM bits per sampler output bit are required.

The data format conversion block seen in figure 1 will take the 32 parallel outputs of each sampler and, using RAMs, re-sort the samples. In this block, the 32 parallel outputs of a high speed sampler would be converted from each carrying every 32nd sample to each carrying short (about 1 msec) bursts of contiguous samples. If the N-wide parallel (2-bit) output of a high speed sampler (each output carrying every Nth sample) were to drive the correlators using a conventional architecture, an N-by-N matrix of correlators would be required to insure every sample is correlated with every other sample. For N = 32, this would mean a matrix of 1024 small correlators to correlate the output of every baseband input of every baseline.

By using the format conversion scheme, the 32-wide parallel output from a high speed sampler will be transformed into 32 parallel signals each carrying 1 millisecond time segments of contiguous samples that need drive only an N-by-1 array of correlators. This simplification in the correlator circuit requirements is obtained at the cost of an inefficiency of about 0.2% which results because the end bits in adjacent 1 msec time segments of samples will not get correlated with each other.

(Note that the conversion from a conventional N-by-N architecture to an N-by-1 architecture does not improve the spectral resolution performance of the correlator. The performance is set by the number of hardware correlators in the system. The conversion does, however, greatly simplifying the system wiring in that all N-by-N signals from two antennas do not have to be wired to closely spaced electronics, thus simplifying the wiring matrix driving the cross correlators as well as reducing the number of I/O pins required by logic cards and integrated circuits.)

An additional benefit of the format conversion strategy is that it allows the system the same advantage as a recirculating correlator: when the bandwidth being processed is reduced by a factor of 2, the number of lags the system is capable of generating goes up by a factor of 2. This results in a factor of 4 increase in frequency resolution for a factor of 2 decrease in bandwidth.

Still another advantage of the format conversion (by far the most important in the MMA correlator) is that it allows a minimum cable interconnect complex between the station electronics and the correlators. It also eliminates any requirement to interconnect correlator arrays in low bandwidth modes. Since the number of data interfaces between these two stages in the MMA correlator surpasses that of any other astronomical correlator system by a factor of almost 100, this aspect of the system architecture is most important.

Block diagrams (not shown) for the delay line and a data format conversion card are almost identical and it is possible that these two cards can be of a single design. This design would have re-programmable logic such as field programmable logic arrays that would be configured at power up for one or the other function.

The cross correlator matrix of figure 1 is used to correlate the sampler outputs of every antenna with those of every other antenna. At the intersection of any antenna X and another antenna Y in this matrix, there will be a correlator chip. This correlator will compute lag products for the XY baseline while the antenna Y and antenna X intersection of the matrix computes the baseline lead products. Auto correlation products for each antenna are obtained from correlators on the matrix diagonal.

In order to minimize further the delay-line to cross-multiplier cable interconnect, a very compact cross correlator matrix is essential. The proposed design for the MMA correlator places an entire 40 X 40 cross correlator matrix (handling a 1/32, 125 MHz data rate, slice of the decimated sampler outputs) for two baseband inputs of opposite polarization on a single printed circuit card. This PC card in addition is configured such that no signal drives more than one load. For the number of signals required on a 40 antenna system, this property permits an absolute minimum cable matrix since every signal out of the station electronics goes one and only one place, driving only a single load.

One disadvantage of the proposed architecture is that once the number of antennas for the array has been set, future expansion of the correlator beyond this number is not practical.

The proposed custom lag correlator chip has a dual 4-by-4 array of correlators (one for each of 2 polarizations). The chip can be programmed via a microprocessor supplied program word for its position in the matrix and to select one of three correlator configurations;

1. four short correlators to compute the lags of all 4 polarization products (RR, RL, LR, and LL).
2. two longer correlators to compute just the lags for the two polarization components (RR and LL).
3. a single long correlator to compute lags for only one of the two baseband inputs.

Estimated size of this custom correlator chip is in the 750,000 gate range.

In observations where fewer than 8 baseband inputs are being used, more lags can be produced by dedicating more than one correlator array to process the outputs of active baseband inputs. In this case, cards in the data format conversion stage will be used to form a virtual connection, the effect of which is to link two or more correlator arrays in series. The delayed input to the correlator chips that are to compute the higher level lags will be displaced in time the appropriate number of bits by offset RAM addressing in the data format conversion cards.

The data format conversion stage will also do the sample decimation for observations in which sample rates less than 4 GS/second are needed. Again, offset RAM addressing in this stage will generate offset delays for the computation of additional lags.

The long term accumulation block seen in figure 1 integrates the correlator outputs for the desired duration. The correlator chips will produce a total of 52,428,800 lag results to be accumulated. The parallelism factor, 32, allows the reduction of this number to 1,638,400 which when double buffered and spread across 32 long term accumulator cards will require integration storage of 102,400 results per card.

4. Performance

This section gives performance parameters for some typical operating modes of the MMA correlator. The MMA correlator will be programmable on a baseband by baseband basis and, hence, some baseband inputs may be processed in one mode while other baseband inputs are processed in other modes.

Sub-arrays will also be possible using the MMA correlator. The maximum number of sub-arrays for the MMA will not be determined by the correlator (that is, the MMA correlator will be able to support the maximum number of sub-arrays limited by other parts of the MMA).

There are assumed to be 8 samplers per antenna. The baseband inputs driving the samplers can consist of 4 dual polarization pairs or 8 independent inputs. Where baseband inputs come in polarization pairs, all 4 polarization cross-products may be computed. Each sampler is assumed to digitize at 4 GHz and hence to be driven by RF signals at most 2 GHz in bandwidth. The maximum bandwidth processed is thus 16 GHz split into 2 GHz pieces. Note that the analog baseband constraints of the planned MMA baseband processing system will impose limits as well.

The smallest division of lags in the projected correlator chip is 64 lags. Because of the architecture proposed, this will produce 64 lead and 64 lag channels and hence 64 spectral points per product. This smallest correlator division means that in the full-up configuration, all baseband inputs active at maximum bandwidth and all 4 polarization products being computed, 64 spectral points will be produced for every baseline, every spectrum. This gives a frequency resolution per spectral channel of 31.25 MHz.

Given the full-up performance as defined above, the number of lags that the proposed correlator can produce for a given experiment results from the following considerations:

1. If polarization cross-products are not required, a factor of 2 more lags (finer resolution) can be obtained. The particular configuration can be selected on an baseband pair by baseband pair basis.
2. If fewer than 8 baseband inputs are required, lags go up as 1 over the fraction of baseband inputs used (1/2 the baseband inputs, 2 times the lags).
3. If a lower bandwidth than 2 GHz per baseband input is required, lags go as 1 over the fraction of maximum bandwidth (1/4 the maximum bandwidth, 4 times the lags) until a factor of 32 is reached. After that, the number of lags stays constant. The particular configuration can be selected on an baseband by baseband basis.

Note that item 3 implies the characteristic described above that for each reduction by a factor of 2 in bandwidth, an increase of a factor of 4 in resolution is obtained (up to the factor of 32 limit after which the resolution improves by only 2 for each factor of 2 reduction in bandwidth).

Table 1 below illustrates some of the possible modes. The first four columns relate to the correlator proper. The columns relating to velocity range and resolution assume 90% of the analog bandwidth will be usable. (See MMA memo 194 for additional illustration of the MMA correlator performance.)

# samplers	bandwidth /sampler	cross-pol	channels /product	At 230 GHz, in velocity space: Total range km/s	Resolution km/s
8	2 GHz	yes	64	9,391	40.8
8	2 GHz	no	128	18,783	20.4
8	1 GHz	no	256	9,391	5.1
8	500 MHz	yes	256	2,348	2.5
8	250 MHz	no	1024	2,348	0.32
4	2 GHz	yes	128	4,696	20.4
4	1 GHz	no	512	4,696	2.5
4	500 MHz	yes	512	1,174	1.3
4	250 MHz	no	2048	1,174	0.16
2	2 GHz	yes	256	2,348	10.2
2	1 GHz	no	1024	2,348	1.3
2	500 MHz	yes	1024	587	0.64
2	250 MHz	no	4096	587	0.08
1	2 GHz	-	1024	2,348	2.5
1	1 GHz	-	2048	1,174	0.64
1	500 MHz	-	4096	587	0.16
1	250 MHz	-	8192	294	0.04

A specification for the output dump rate of the MMA correlator has not been set yet. However, the architecture thus far described is versatile in regard to dump rates by the very nature of the process. The short time segments of samples from the memory cards insure that very short fundamental integrations are always made. Thus, dumps of the long term accumulations can be made at natural intervals that are multiples of 1/32 and 1 millisecond. This means that the correlator hardware need never be the limiting factor in obtaining high dump rates; the down stream processing and storage medium used with the system would set the dump rate limit. Since this part of the system can be changed as processing and storage technology improves, the system will be able to keep up with this improvement.

5. Size and Power Requirement Estimate

Table 2 gives a preliminary count of the module and printed circuit card requirements for the MMA correlator:

Table 2

Item	# req'd	size	power req'd	racks
4 GS/S dual sampler	160	2-wide VLBA module	20 w	4 racks
mode card	40	6U euro card	20 w	with samples
delay line	320	6U euro card	50 w	10 racks
memory card	320	6U euro card	80 w	with delay line:
correlator card	128	9U euro card	300 w	8 racks
control cards	32		40 w	with other card:

long term accumulator	32	60 w	with correlators
totals		100 kw	22 racks

The power estimates given in the table above are based on the experience gained in the development of the GBT spectrometer. The biggest unknown at this time is the dissipation to be expected in the custom correlator chip, 12,800 of which will be required in the system. The GBT correlator chip dissipates about 5 watts with a clock rate of 125 MHz. Such a high chip dissipation in the MMA correlator would mean both high system power requirements and lower reliability because of the difficulty in removing the heat from the system at the high altitude site.

By using low voltage chip technology it is hoped that the custom correlator chip described in this document can be built with about a 2 or 3 watt power requirement. The chip represents about a factor 2 increase in the level of integration when compared to the GBT correlator chip (twice the number of transistors). By using a more modern process, with finer component features and low voltage technology, a smaller chip with lower power requirements should be possible. The smaller silicon size should also mean a higher yield in the manufacturing process.

Holography Requirements for the MMA

Darrel Emerson
Last modified July 14 1998

1. Introduction

Holography will be used to measure the first MMA dishes, soon after they are first installed at the VLA site. The requirements are a measurement accuracy (rms) of 10 microns, and a resolution on the surface such that several independent points are available for each panel. In practice, this means the dish should be sampled at about 10 cm intervals. This in turn means that a 10-meter dish will need to be measured by about a 100*100 array of points. In practice, most holographic maps will probably be made at 256*256 or 128*128 points, with occasional measurements at 64*64 points. There is no requirement for the data to be measured on a 2^N grid - in fact an odd number of points, giving a symmetrical grid with the center point on the bore-sight, is advantageous. For example, measurements on the 12 Meter Telescope used a grid 129*129 or 65*65 points.

Note that, unlike standard single-dish astronomical measurements, holographic measurements record complex pairs at each sample in the sky plane, rather than just a single total power value. This implies that Nyquist sampling is defined as (lambda/D) radians, rather than (lambda/2.D). D is of course the dish diameter, and lambda the wavelength at which holographic measurements are being made. The number of complex data points (e.g. 129*129) mapped in the sky plane, giving the complex antenna beam pattern, equals the number of data points (for example 129*129) describing the complex antenna illumination pattern. So, the total angular extent of the necessary beam map is simply calculated, once the holographic wavelength, dish diameter, and necessary sampling interval on the dish surface are determined. For example, if a wavelength of 3.33 mm (90 GHz) is chosen, lambda/D for a 10-meter dish becomes 0.000333 radians, or about 69 arc sec. If a grid 257*257 is needed, the total map extent becomes 257*69" or about 4.9 degrees. In practice a little oversampling is always necessary, by perhaps 20%; in this example a sampling interval (after gridding) of about 57" would be appropriate.

Holography will be carried out in 2 distinct modes. In both cases, the aim is to produce a complex beam pattern - that is to say a map of relative amplitude and phase - of the antenna being measured.

a. Single dish observations: the phase reference for the measurement of complex antenna pattern will be provided by a small feed looking towards the transmitter, behind the main dish feed, at the prime focus of the antenna. The antenna will be scanned back and forth over the source, to map its detailed, complex beam pattern

b. Interferometric observations: this will use a pair of antennas, with one antenna tracking the source and providing the phase reference for the second antenna. The second antenna will scan back and forth across the source, to produce a 2-D map of its own complex beam pattern.

Case (a), the single dish mode, is in general a little more complex because of the necessary calibration procedures. Case (b), the interferometric mode is closer to a normal, astronomical mode of observing. In what follows, only the single dish mode, case (a), is considered.

2. **Holography Hardware**

   1. Choice of observing frequency

      The precise frequency is not critical. Holography measures physical distances; the lower the frequency, the smaller the phase change corresponding to a given distance, and so the higher the signal-to-noise required. If the frequency is too low, diffraction effects (diffraction shadows around the feedlegs, and diffraction around the central antenna blockage) can become significant. The lower the frequency, the larger the area on the sky around the boresight which has to be mapped, for a given linear resolution on the dish surface; this may ultimately present difficulties to the antenna drive and control system, in order for it to be possible to make a sufficiently large map in a reasonable amount of time.

      The required signal-to-noise ratio and dynamic range requirement both increase inversely with frequency. At too low a frequency, the needed dynamic range can become a serious problem, and even small cross-talk between the two receiver channels (dish feed and reference horn) can introduce serious errors.

      If the frequency is too high, then ambiguities can arise in the measures of the dish surface; fundamentally, holography cannot distinguish between a patch on the dish surface lambda/4 too high, or a patch lambda/4 too low. At the frequency becomes higher, several factors cause the signal-to-noise ratio to become worse - receiver noise temperatures are higher, and the available signal power may become less. Interestingly, the antenna capture area of the reference feed does not vary with frequency; the beam solid angle needed for a given resolution on the dish surface decreases with the square of frequency, exactly as the beamwidth of an antenna with constant physical size. The physical size of the reference feed approximates to the limiting physical resolution of the final holographic map of the dish surface.

      The ideal frequency is probably around 90 GHz. Ideally a signal source would be in the far field, but this is not essential. Several groups have achieved excellent holographic measurements of mm-wave or submm-wave telescopes using an artificial beacon a few km distant. For the reasons mentioned above, frequencies below 30 GHz are probably not suitable. Unfortunately, no satellite beacon transmissions suitable for MMA holography have yet been identified (although the search continues.)

   2. Frontend

      The single-dish holographic measurements will be made, at least initially, with a prime focus receiver. For the duration of the measurements the holography receiver box will be mounted at prime focus instead of the normal subreflector. The main advantage of prime focus holography is that potential measurement uncertainties resulting from inaccuracies of the subreflector are avoided. The holography frontend will have two feeds; the first, mounted close to the true focal point, will illuminate the dish in the normal way. (Note however that for holographic measurements it is advantageous to over-illuminate the antenna; G/T optimization is not an issue for holography.) The second feed will point away from the dish, along the boresight of the antenna; this feed serves to provide the reference signal for the holographic system. The beamwidth of this reference feed should be somewhat larger than the maximum anticipated holographic map - for example, 5 degrees. Note the comment above that, in this receiver arrangement, the ultimate physical resolution on the dish surface approximates to the physical size of the reference feed, whatever frequency is chosen for measurements.

Both feeds have independent r.f. amplifiers (if available - this depends on the final choice of frequency) and independent mixers. The independent mixers will be fed from a common local oscillator source. The two resultant i.f. signals will be fed independently to the backend processor, which may be mounted in the control room. Temperatures, and hence phase drifts, in the two mixer and i.f. chains, with their cables, should be well matched to avoid measurement errors.

3. Backend

The measurements will be made on a CW signal. To optimize signal-to-noise ratio, a receiver bandwidth of perhaps 100 Hz will be used. The two IF signals will be filtered to a few kHz of bandwidth, then digitized directly. A DSP card will perform FFTs on the data samples to produce a spectrum with perhaps a few Hz resolution. The peak signal (amplitude and phase) will be chosen and stored for later analysis. Data from the main beam (the telescope feed) need to be sampled often enough to match the holography on-the-fly mapping rate. Something like 10 ms will probably be appropriate. The reference channel, with its much larger beam, should be integrated later in the data reduction software, to improve signal-to-noise ratio; this will reduce its effective data rate by about 2 orders of magnitude. However, at the raw data stage, it will also be sampled and stored at up to the 10 ms rate.

4. General Telescope System

In order for holography observations to be possible, much of the telescope system needs to be fully operational. The most critical area is the telescope pointing; this has to be well understood and reproducible before any holography observations can be attempted. Proven observing techniques sufficient to check the telescope pointing frequently during a holography map will be needed. The telescope control system must already support high speed mapping operations, especially the on-the-fly mode. The holography mapping mode will be a variation of conventional on-the-fly observing; for instance, boresight pointing, amplitude and phase calibrations will be needed throughout a holography map - every mapping row, or perhaps every few rows depending on the stability of the system. The ability to make, analyse and apply pointing measurements quickly, in the course of a holography map, is an important requirement.

3. Holography Data Reduction Software

The data reduction will take raw data from disk, which has been observed in the on-the-fly mode, and will ultimately produce a list of adjustments, calibrated in microns, for each adjustment point of each panel of the dish.

The raw data for one holography map will consist of from 30 to 513 map rows. Each row may have up to about 5000 data samples. Each sample may be a complex spectrum of one to a hundred points. Each data point will have associated co-ordinate information. The sampling along each row will be up to 10 times greater than the Nyquist rate for the telescope, while the sampling interval between rows will be perhaps 20% more frequent than the corresponding Nyquist rate. At the start and end of each map row, or in general after n map rows, there will be a calibration measurement taken on boresight. The basic observing grid will probably be in an azimuth-elevation system, with respect to the transmitter. The transmitter may be moving slowly (e.g. on a satellite) during the observations.

The steps in the data analysis will be:

1. From each point in the spectrum, the complex data will be interpolated to a regular, 2-D grid. Either before or after the interpolation, some algorithm will choose the strongest point of the spectrum, reducing the spectrum to a single complex number.

   This regular grid will be an antenna-based co-ordinate system, significantly different from the original Az-El offset co-ordinate system. Note that the FFT-pair relationship between antenna far field beam pattern and aperture illumiation is a function of the sine of the angular offset from boresight, rather than simply of angle. Since the holography map may extend over as much as 5 degrees, this begins to become a significant correction.

2. The gridded data will be calibrated in amplitude and phase, based on the boresight measurement at the beginning or end of each of the n map rows and assuming a gradual drift in gain and instrumental phase with time.

3. Phase corrections will be applied to the gridded data, to bring the antenna reference field close to the plane of the antenna surface. This is analagous to a refocus operation.

4. Amplitude and phase corrections will be applied to the gridded data, to allow for the complex antenna response of the holography reference feed.

5. Some tapering may be applied to the gridded data, to reduce the sidelobes of the point spread function after the FFT.

6. A Fourier Transform is made of the gridded, corrected data. Note that in general the grid will not be 2^m points, and will usually be an odd number, to put the antenna boresight on a grid point at the center of the field before the FT. After the FT, the data represent the aperture illumination pattern, and the aperture phase pattern, of the dish.

7. After the FT, some correction needs to be applied to allow for diffraction fringes from the edge of the dish, from the feed-legs, and from the central blockage. The shadowed areas will also need to be masked out.

8. A feed displacement correction needs to be applied. This will be a sum of:
   a. A least squares fit to a 2-D linear gradient across the phase map. This corresponds to a

systematic pointing error, if any, during the observations.

b. A fit to an out-of-focus term. This corresponds to an axial out-of-focus term. It approximates, but is not exactly, a quadratic distribution across the antenna.
c. Higher order aberrations, such as coma lobs causes by radial offsets in the holography feed mount.

9. The corrected phase map now corresponds to an estimate of the errors in the dish surface, normal to the wavefront. From the phase map, we need to derive the errors normal to the dish surface, at the panel mounting points. If feasible, some algorithm should take account in some way of the finite resolution of the holography map.

10. Taking account of some structural model of panel and backup structure deformations, a table of corrections needs to be calculated for use by the antenna adjuster crew.

During the holography measurement campaign, the maps will be observed at night while the temperature is stable, and differential panel adjustments will be made during the day. The overnight holography data needs to be analysed in time for the morning adjustment crew to take the panel adjustment correction tables.

It is also likely that unexpected problems will be found during the holography observations and data analysis. It is important that the software analysis system be sufficiently versatile that any of the above steps can be modified, or additional analysis algorithms can be applied, in a timely fashion. It should be possible to introduce some new step into the analysis with not more than about one hour of programming effort.

4. Work Plan

In order to accomplish the above plan, the basic steps are:

1. Define in detail the holography system specifications, including transmitter frequency, needed power and signal to noise ratios, frontend and backend requirements.
2. Continue the search for a suitable satellite beacon, which would complement measurements with a terrestrial transmitter.
3. Study the possibilities for single-dish holography with astronomical sources (e.g. SiO masers). Coarse resolution holography may be possible on these sources, enabling large-scale dish deformations as a function of elevation to be studied.
4. Define in detail the design for the holography hardware, including transmitter, receiver frontend, correlator, DSP etc.
5. Define in more detail, in collaboration with the AIPS++ group, the specifications for data analysis software.
6. Define in detail the interferometric holography requirements. Interferometric holography will offer much higher signal-to-noise ratio on a given signal source, simply because the full 10 m aperture, rather than a broadbeam reference horn, can be used to retrieve the phase reference signal. This will permit measurements using astronomical sources. However, more of the complete electronics system has to be operational reliably for interferometric holographic measurements to be useful - fringes have to tracked with high phase stability. In the much longer term, holographic measurements with the MMA will probably become exclusively interferometric.
7. Define the timescales for all the above, taking into account antenna delivery schedules etc..

MMA COMPUTING

Brian Glendenning
Last changed July 19 1998

1. Introduction

This document describes the computing environment required for correct operations of the MMA. It includes the real-time software to control and monitor the array, as well as other software required for the array to be operated as an astronomical instrument (for example, scheduling). Post-processing software is described in another chapter of this project book, as are monitor and control (M&C) hardware considerations and holography requirements.

The MMA will undergo two stages of development, a Design and Development (D&D) stage, to be followed by a construction stage, possibly with another partner. The brief of the computing group during the D&D phase is as follows:

1. Write all software that is necessary to operate and test the hardware created for the D&D phase of the MMA. In particular, software required to measure and test the antenna is of the highest priority. If necessary this software can be thrown away after the D&D phase, although it would be preferable if this was not required.
2. A paper design and implementation plan for all software that will be required during construction and initial operation of the MMA.
3. Prototyping of software elements required for MMA construction or operation whose design or implementation technology is problematical.

An important principle is that we should use other people?s software whenever possible, and if we can?t reused their software directly, we should attempt to reuse their design ideas. As a rule of thumb, for software *effort* ~*size*^{3}, so there is a tremendous incentive to fight the "not invented here" (NIH) syndrome unless adopting that software truly does do violence to the integrity or maintainability of the system.

We intend to write software that emulates critical hardware (for example, the antenna) before that hardware is ready. We have not yet decided whether to create a software-only emulation, or whether it should be plug-compatible with the hardware it is emulating.

This document is not a design document. No detailed analysis has been undertaken yet. In many instances the detailed requirements are not yet known. In general, this chapter contains some requirements at the highest levels, and some discussion of implementation choices that might be taken to fulfill those requirements.

The most important thing to identify for MMA computing is a mechanism to obtain detailed specifications for the software and hardware. While these will likely be self-generated from within the project for the real-time control aspects of the MMA, it is vital that a mechanism be found for obtaining astronomy-level specifications from the community of potential users of the array.

2. Computing Standards

Some project-wide standards for MMA computing will be required. It is important that these standards be useful, not merely bureaucratic.

We must decide if a formal software design methodology is useful; and if so whether a tool to support it is worth purchasing. While there are many choices for a design methodology, the most promising is probably the Unified Modeling Language (UML) developed by a number of software methodology luminaries. If we decide not to adopt a formal methodology, we need to define how we want to document the design (this alone might be reason enough to adopt a methodology).

We need to choose which operating systems the MMA software will be built upon. For the real-time operating system we shall almost certainly choose one of VxWorks or real-time Linux. The former has been used widely within the observatory (GBT, VLBA, Tucson) and is a mature, well-supported commercial product. It is also relatively expensive. RT Linux is currently being investigated in Tucson for two 12m experiments (focus and tracker computers). It is a much simpler RT environment (no debugger, for example), but offers the ability to tightly couple a full Linux programming environment on the other "side" of the RT interface. For the non-real-time systems, the obvious choice is between a version of Unix and Windows NT. I believe we will face fewer problems if we adopt Unix for the D&D phase. We might well want to revisit this choice before construction.

We need to choose computer languages for the software we write ourselves. A mixture of C and C++ for the control software (C for real-time, C++ for non-real-time) might be appropriate for most of the MMA backbone software. We might well use Java and/or a scripting language such as Glish or Tcl/Tk for GUI?s. We should be prudent and try to keep the number of languages we require to be as small as possible for maintenance reasons. Software that we adopt might require that languages be available that we do not ourselves use, for example, Spike (Miller, 1996), is written in Lisp.

Whatever languages we choose, it is vital that all software and associated documentation that we write which is required for operating the MMA be under the control of a version control system. This includes software which is required to test the MMA at the site (*i.e.*, not test software that is only run in the lab). Concurrent Versions System (CVS) is widely used and appears to be suitable. It supports distributed development.

It is desirable that the programmer-level documentation be extracted from the source code - if the documentation is near the source code it reduces the likelihood that the documentation will become stale. Many tools are available to do this.

We have to define what, if any, code acceptance procedures and automatic test procedures we want to adopt for the MMA software code base. At the minimum we should probably require that new code be reviewed for completeness of documentation, and that optional automatic unit-test procedures are available.

3. Real Time Components

The MMA computing design should minimize the software that runs in real-time systems. Experience shows that the real-time systems are replaced more slowly than general-purpose computers. In particular, as much of the data-handling system as possible should be in general purpose computers to allow the throughput of the MMA to be increased by replacing general computer infrastructure.

3.1 Monitor & Control

As described in the M&C chapter of this project book there is a fundamental design choice which has to be taken: does every antenna have a computer or not? This fundamental architectural question which must be decided early in the D&D phase since many decisions will flow from it.

There are a number of general features that the M&C system should support however which are independent of the above choice.

- The software is less brittle if there is a consistent interface to devices. For example, changes to the hardware interface should not necessitate software changes.
- Operator interfaces should not be considered a last minute detail. Unsuitable interfaces can limit the operational efficiency of the telescope, and are hard to retrofit. A system that can avoid displaying error cascades (where a faulty component induces many other errors), like the system implemented at OVRO is helpful.
- The ability to remotely monitor and control is extremely valuable (*e.g.*, Scott, 1998). Security is clearly a concern in such systems.
- The MMA will have to take a decision about whether monitor data is ephemeral, or whether it should to be archived to allow for long-term monitoring of the health of the instrument, and to allow astronomers to make post-observation modifications to their data based on what they find in the monitor data. Whatever the lifetime of the monitor data, a database to contain it will either have to be defined or bought. Although relational database systems appear to be natural for this application, there are efficiency concerns that should be investigated before taking this decision.

All of the above argue for a middleware level that isolate the above items from some of the hardware and communications details. One possibility is to use an Object Request Broker (ORB), probably The Ace Orb (TAO) which is a CORBA standard ORB which offers Quality of Service (QoS) guarantees to allow it to be used within a real-time environment. Although there is much to recommend this technology, it may be premature to adopt it.

Another possibility is the Experimental Physics and Industrial Control System (EPICS) which has been under development in the accelerator physics community for about 10 years, and has been used to good effect in various astronomy telescope control systems (*e.g.*, Gemini). It offers us a framework for the entire M&C software subsystem (including things like alarms, plotting, and monitor screens). It does however require that we use VxWorks for the RTOS.

Another possibility would be to use the device manager/RPC++ subsystems designed in for the GBT.

Whichever middleware solution is adopted or implemented, it is important that the hardware and software protocols are not so elaborate that they preclude engineers from "plugging" devices into something like LabView for testing in the lab. It is unlikely that engineers will want to deal with the details inherent in the full M&C software system.

It should be possible to read any value which can be set.

3.2 Correlator (Data Rate)

Perhaps the single most important number for the MMA software is the maximum data rate the MMA must support initially. The data rate of the MMA will be approximately:

$$10\,(N_{antenna}/40)^2 \times (N_{channel}/1024) \times (N_{pol}) \times (1s/t_{dump})\ \text{MBytes/s}$$

Rupen (1997) makes the case that on-the-fly (OTF) synthesis imaging of astronomical sources could quite plausibly use 25ms correlator dump times (for a 10m antenna), leading to data rates of 100?s of Mbytes/s.

It does not appear to me to be prudent to commit the MMA to be able to sustain such data rates. While they might be achievable through the use of enough parallel I/O channels, the entire software system (and supporting hardware) would have to contort itself to support such data rates, and the resultant system might well end up being brittle.

Clearly the correlator hardware should be designed in such a way that data can arrive at interface computers at these sorts of data rates, so that the data rate can be increased by changing interface computers rather than correlator internals.

It appears that data rates of a few 10?s of Mbytes/s should be achievable with available bus technology and disk arrays. If rates higher than this are required at first light, then during the D&D phase high-data rate solutions should be investigated at high priority (note that 10 Mbytes/s was the recommended maximum rate of the MMA Software Working Group (MMASWG) (Scott *et. al.*, 1996)).

3.3 Antenna

Sophisticated antenna control is critical for the D&D phase. For example, holography will impose stringent requirements on the supported single-dish modes.

Antenna control is another area which is strongly affected by how computers are distributed in the array. For example, with a computer in each antenna one would be more inclined to update simple antenna positions frequently, whereas with a central computer you would likely send commands less often, so the command would probably be more complex (*e.g.*, contain polynomials).

In any event, basic antenna control (point, track) appears to be an area where we can reuse code or designs, from other instruments, both from within the observatory and from without.

Since the single-dish modes for the D&D phase are in service of holography of the prototype antenna, and since that effort will be building upon the experience of NRAO/Tucson personnel, the single-dish modes should probably be built by replicating the "experience" built into the 12m observing scripts.

4. Data Production

A thorny question is whether the first data "emitted" (that is, available to other programs) by the MMA should be in FITS format or a local format.

There are good reasons to choose a local format. FITS is a fairly rigid data format. It can be awkward to fit data with different "shapes" together, and there is no agreed upon standard way to represent hierarchies in FITS. Quick debugging or display programs are arguably harder to write for FITS files than with other formats.

However there is a compelling reason to choose FITS. It keeps the system honest - it ensures that data which is read from the archive will not be second class data - it will contain all required information. (I assume there is no dispute that archived data will be in FITS format).

The fundamental atom of data production will be a multi-HDU FITS file containing the data for some time range. The data, synthesis and total power, should be written in the format defined Diamond *et. al.* (1997) augmented as necessary for the MMA. If total power data cannot be represented readily in this format, they should be written in the emerging SDFITS format.

These FITS data files will be created by combining data blocks from the correlator with monitor data. In keeping with the principle that the real-time systems be kept as simple as possible, creating the FITS files should be the responsibility of a general-purpose computer. If the monitor data volume is small compared to the volume of science data, we should consider embedding all monitor data for a time range in its FITS file.

The MMA should maintain a dictionary of all FITS keyword names and column names to ensure that they are used uniformly.

5. Offline Processing

It is important to ensure that MMA data can be reduced by astronomers at their home institutions by at least one software package. The MMA should obtain a commitment from the AIPS++ project that it will meet those needs. The MMA should provide to AIPS++ a set of requirements for the offline system, particularly those that involve new algorithm development or high data-rate or parallel processing. While the AIPS++ project will be responsible for implementing most processing (*e.g.*, mosaicing), very instrument-specific operations may be implemented by MMA staff - at a minimum, the filler.

Astronomers are of course not precluded from using other systems, or writing their own programs to do specialized processing.

6. Service Data Processing

A fundamental decision for MMA operations is how much data processing will be performed for the observer. At one extreme the observatory merely provides raw data and instrument calibrations, at the other extreme the observatory undertakes to provide the observer fully processed images which have been carefully scrutinized for flaws by humans.

The recommendation of the MMASWG is that final, or near-final, quality images be produced for the user by MMA software and staff.

This decision implies that a larger number of data analysts may be required for the MMA than for other NRAO instruments. This need may be alleviated if adequate automatic tools for data editing (*e.g.*, interference excision), data calibration, and image formation can be created. Automatic imaging of this sort is going to be investigated at NCSA during the D&D phase as an MDC project.

Automatic processing of data at the average data rate of the MMA will probably require parallel processing during the initial operations of the MMA. The MMA should obtain a commitment from the AIPS++ project that it will be able to utilize such hardware.

7. Scheduling

To maximize the scientific throughput of the MMA, dynamic scheduling will be required. In brief, projects will be observed only when conditions are favorable for that project (*e.g.*, phase stability and airmass).

As Holdaway (1997) describes, this implies that the schedule will be created very nearly in real-time using many criteria, including:

- Scientific priority of the project.
- LST.
- Environmental measurements:
  - Current opacity of the atmosphere (as a function of frequency).
  - Current phase stability.
  - Weather measures, especially wind statistics.

  A history should be kept for all these measurements so that the scheduler can make decisions based on, *e.g.*, how often these conditions are likely to recur before the next array configuration change.

- Required UV coverage (while the snapshot coverage of the MMA is sufficient for compact configurations, larger configurations require more hour-angle coverage (Holdaway, 1998)).
- Calibration will impose additional requirements, for example parallactic angle coverage for polarization calibration (Cotton, 1998). Calibration observations should be shared between projects when possible. However projects which require extra calibrations should be charged by the scheduler for them.
- Array efficiency (slew time, projects that share calibration, linked observations that must be executed at about the same time).

This implies that the raw data for a single project will consist of independently calibratable chunks, possibly spread over many days.

Fortunately for the MMA there is experience with implementing dynamic scheduling. Wright (1997), has implemented a scheduling system at BIMA which incorporates some measurements of the atmosphere into the scheduling process. The ESO VLT telescope is implementing a "Data Flow System" which appears to implement many of our scheduling requirements (Silva and Quinn, 1997). The STScI has released Spike, a portable scheduling engine which we may want to adopt (Miller and Bose, 1996).

We shall have to provide the user with proposal and scheduling tools. We will have to decide whether the user should merely specify science objectives (*e.g.*, point source S/N) and rely on the MMA system (software and personnel) to schedule the observation (including calibration), or whether the user should create a more explicit schedule of observations. If the user specifies scientific objectives the system has to be able to measure them, which is in general a difficult problem. Measuring the number of "good" hours spent on project targets and calibrators would be much more straightforward.

Although dynamic scheduling will probably be the most common scheduling mode, standard queue observing must be supported (*e..g*, VLBI, transient sources). In addition, there may be circumstances where real-time control of the array is required.

There could be feedback between the scheduling system and array configuration changes. In principle, configuration changes could be driven by the scheduler.

8. Archiving

It is assumed that all data products from the MMA shall be archived in perpetuity, and made available to researchers over the network. If the average data rate of the MMA is 1Mbyte/s, and if the data product volume is dominated by the raw data, then a few 10?s of TB per year will have to be archived. While this is considerable, it is also manageable (a few thousand double-sided, double-density DVD?s, for example). The choice of archival media and hierarchical storage system should be deferred until MMA construction nears completion.

The main archive repository will probably not be at the MMA site in Chile (the demands it will place on general networking infrastructure will be considerable). As noted by the MMASWG (Scott *et. al.*, 1996), if we cannot achieve sufficient network bandwidth to Chile we may need to ship physical media from the MMA site in Chile.

We might need to support replication of the archive at one or more locations to relieve, for example, transcontinental network congestion. Replication should be built into the architecture of the archive. Even if there is only one active data center, the data should be physically replicated and stored in an offsite location for data recovery and safety reasons.

Access to data is not sufficient of course. MMA observations are apt to be complex, with observations and calibrations spread out widely in time. Also, some or most observations will have images of varying sophistication associated with them. This implies that fairly sophisticated meta-information, largely gleaned from the observers proposal and the actual schedule, will have to be maintained to allow the desired data to be found in a straightforward way.

The archive will be used both for post-observation data retrieval, and to inspect data while observations are proceeding. In this latter case, email notification to the proposal authors should be implemented so they can inspect their data as it becomes available.

The MMA archive system appears to share many features with other astronomy archives. For example, the Sloan Digital Sky Survey (SDSS) archive has an interesting three-tiered archive structure, and a decomposition into an operational and science archive (Brunner et. al., 1996). Another example is the STScI archive re-engineering effort (Hanisch et. al., 1997). We should take advantage of such expertise.

The archive must protect proprietary data rights and not release data to the general community until time limits have expired. Mechanisms to enforce these rights must not be too burdensome.

Archiving is going to be investigated during the D&D phase at NCSA as an MDC project.

9. Acknowledgements

Although the above is still woefully incomplete, it would be even more so without a number of discussions I have had with various people. I would like to thank: Tony Beasley, Barry Clark, Tim Cornwell, Darrel Emerson, Jeff Hagen, Ron Heald, Mark Holdaway, Frazer Owen, Michael Rupen, Steve Scott , and Al Wootten for helpful comments and advice.

10. References

R.J. Brunner, I. Csabai, A. Szalay, A.J. Connolly, G.P. Szokoly, "The Science Archive for the Sloan Digital Sky Survey", ADASS V, 1996.

W.D. Cotton, "Polarization Calibration of the MMA: Circular versus Linear Feeds", MMA Memo #208, 1998.

P.J. Diamond, J. Benson, W.D. Cotton, D.C. Wells, J.D. Romney, G. Hunt, "FITS Format for Interferometry Data Exchange", VLBA Correlator Memo #108, 1997.

R.J. Hanisch, F. Abney, M. Donahue, L. Gardner, E. Hopkins, H. Kennedy, M. Jyprianou, J. Pollizzi, M. Postman, J. Richon, D. Swade, J. Travisano, R. White, "HARP - The Hubble Archive Re-Engineering Project", ADASS VI, 1997.

M.A. Holdaway, "High Level Computing Information Flow for the MMA", MMA Memo #167, 1997.

M.A. Holdaway, "Hour Angle Ranges for Configuration Optimization", MMA Memo #201, 1998.

Glenn E. Miller, Ashim Bose, "Portable Astronomical Scheduling Tools", ADASS V, 1996.

Michael P. Rupen, "The Astronomical Case for Short Integration Times on the Millimeter Array", MMA Memo #192, 1997.

Steve Scott, Ray Finch, "The New User Interface for the OVRO Millimeter Array", ADASS VII, 1998.

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

D. Silva and P. Quinn, "VLT Data Flow Operations News", ESO Messenger #90, 1997.

M.C.H. Wright, "DYNAMIC SCHEDULING: Implementation at Hat Creek", BIMA Memo 60, 1997.

Imaging Requirements for the MMA

M.A. Holdaway
Last modified July 13 1998

This list is quite incomplete. It leaves out everything that we already do routinely for the VLA, and leaves out all the things I didn't bother to think about. I mainly address mosaicing, which is because I've thought a lot about mosaicing, but also because mosaicing may well be the biggest topic in imaging which is significantly different from what we do at the VLA.

1. Simulation Capability

Full simulation capability: given THIS source structure, THOSE hour angle tracks (i.e., THIS observing strategy), THIS phase stability, THAT opacity, (add other errors as required), what will my sensitivity be like? How should I calibrate? What will my image quality be like? This should be a tool to aid in the proposal process, and available to the astronomer checking the imaging as well.

Visibility weights should reflect the current noise level (i.e., reflect both changes in T_{sys} and opacity).

2. Mosaicing

Mosaicing +total power; MEM based and CLEAN based algorithms Linear & Non-linear mosaics. We must spend a lot of effort on making the mosaics fast and efficient.

OTF Single Dish software. Simple integration of OTF SD and interferometer mosaicing. Better OTF algorithms may be required, as errors in the total power imaging may limit the overall quality of mosaiced interferometer plus total power images.

Better combination of CLEAN and MEM to clean bad point sources and mem the residual visibilities.

Mosaicing system tools: easy and simple manipulation of multi-pointing data sets.

OTF Mosaicing * continuous slewing mosaicing with integration times set by the minimum allowed by the correlator at the maximum dump rate for the desired number of channels. So, if the correlator is dumping with 0.1 s rates, the MMA will be slewing at about 3-4 beams per second.

Mosaic pointing self-cal algorithms and mosaic imaging with known pointing errors

Mosaic voltage pattern self-cal algorithms and imaging with different known voltage patterns on each antenna

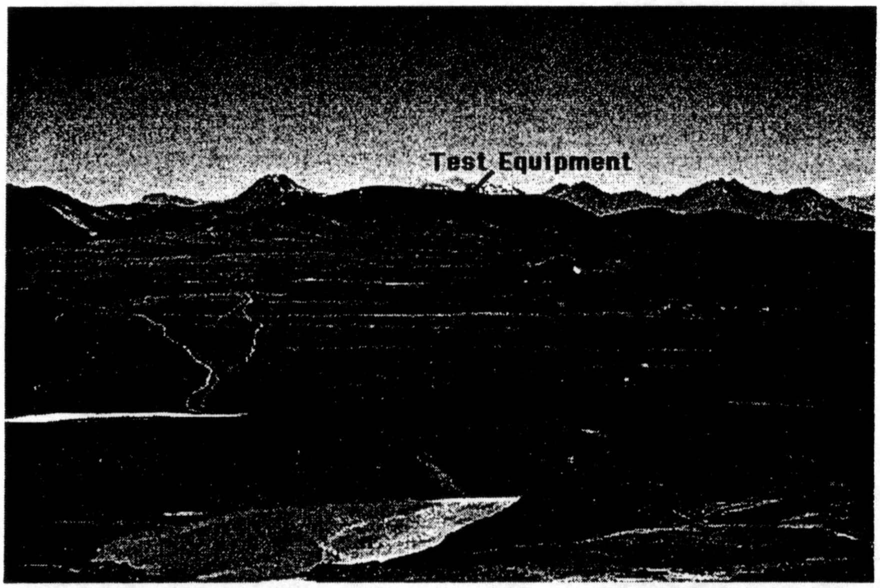
3. Image Interpretation

There are a variety of radiative transfer codes that people use to estimate things like temperature and density based on millimeter wavelength images made of multiple molecular transitions. I suppose that the quality of the images made by the MMA and the quality of the absolute flux calibration will demand much more sophisticated codes for modeling radiative transfer and the physical conditions of the emitting regions. To what extent does this belong within the jurisdiction of the MMA's imaging system?

SITE CHARACTERIZATION

Simon Radford
Last revised July 15 1998

See also other MMA site web pages: Site Studies



View south from Cerro Toco of MMA site near Cerro Chajnantor, Chile.
Photo: M. A. Gordon, 1995 April.

Overview

Goals

During the project's prehistory, NRAO conducted extensive measurements to characterize several candidate sites for the Millimeter Array. These studies have culminated in the (de facto) choice of an array site on the high (5000 m) plateau southwest of Cerro Chajnantor, Chile, about 40 km east of the village of San Pedro de Atacama. The goals of further site characterization are:

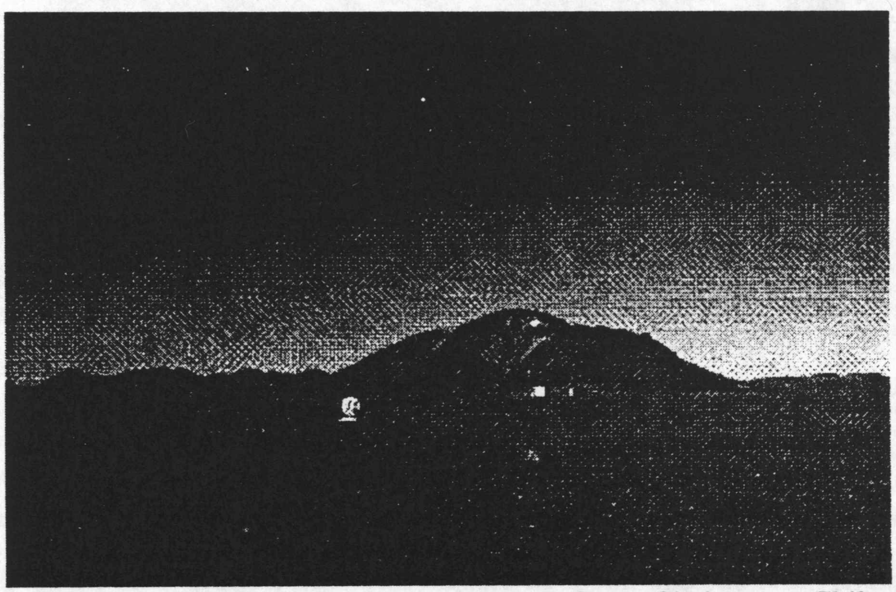
- to identify and quantify site conditions that might influence the instrument design,
- to provide a historical record of site conditions to guide priorities for instrument development and operation,
- to maintain a continuous presence on the site through construction to the start of operations, and
- to interact with other groups working on or near the site.

Areas of interest

At millimeter and submillimeter wavelengths, pressure broadened molecular spectral lines make the atmosphere a natural limitation to the sensitivity and resolution of astronomical observations. Tropospheric water vapor is the principal culprit. The translucent atmosphere both decreases the signal, by attenuating incoming radiation, and increases the noise, by radiating thermally. Furthermore, inhomogeneities in the water vapor distribution cause variations in the electrical path length through the atmosphere. These variations result in phase errors that degrade the sensitivity and resolution of images made with both interferometers and filled aperture telescopes. The site characterization effort addresses these areas:

- Radiometric properties of the atmosphere
  - opacity
  - stability
- Physical structure of the atmosphere
  - meteorology
  - stratification
  - turbulence
- Physical characteristics of the site
  - topology
  - geology

Installed equipment



Moonrise over NRAO test equipment near Cerro Chajnantor, Chile.
Photo: S. Radford, 1995 May.

Infrastructure

- Container. A 20 foot ocean shipping container provides operations base and shelter for personnel and instruments.

- Solar power system. Electrical power is supplied by a battery bank and array of solar panels.
- Communications. Voice and low-speed (<= 9600 baud) data are transmitted by satellite (INMARSAT) and cellular telephone.

225 GHz tipper

The 225 GHz tipping radiometer is the benchmark instrument for site characterization. It measures the atmospheric absorption every 10 minutes and atmospheric emission stability every fifth hour. Operation is automatic. Sporadic repair is required. The data are retrieved daily by telephone and summaries are posted monthly. The data are made available to interested parties in machine readable form.

12 GHz interferometer

The site test interferometers directly measure the tropospheric phase stability. They observe unmodulated 11.5 GHz beacons broadcast from geostationary satellites and measure the phase difference between the signals received by two 1.8 m diameter antennas 300 m apart. Because the atmosphere is non-dispersive away from line centers, the results can be scaled to millimeter and submillimeter wavelengths.

Three instruments have been constructed by NRAO's Tucson office. The first was operated near the VLBA antenna (3720 m) on Mauna Kea, Hawaii, from 1994 September to 1996 June and at the VLA since 1997 May. The second has been operating on Chajnantor (5000 m) near San Pedro de Atacama, Chile, since 1995 May. A third was built for the LSA project. ESO installed it at Pajonales in 1997 April and will move it to Chajnantor in 1998 June.

The design and operation of these instruments are described in Site Test Interferometer (Radford, Reiland, & Shillue 1996, PASP 108, 441). From the phase time series, we obtain the r. m. s. path fluctuations on a 300 m baseline, the power law exponent of the phase structure function, and the velocity at which the turbulent water vapor moves over the array. MMA Memo 129 describes the site test interferometer data reduction in detail, and MMA Memo 130 illustrates the agreement between two different methods of deriving the mean velocity of the turbulent water vapor flow in the atmosphere.

The instrument operates autonomously. Sporadic repair is required. The data are retrieved on tape about once a month and analysed in Tucson. Summaries are posted monthly.

Submm tipper

A tipping photometer has been developed in collaboration with Carnegie Mellon University to directly measure the atmospheric transparency at 350 mu m wavelength. This instrument is based on an ambient temperature, pyroelectric detector. The spectral response is defined by a resonant metal mesh. A compound parabolic (Winston) cone and offset parabolic scanning mirror together define the 6 degree beam on the sky. The detector is internally calibrated with two temperature controlled loads and views the sky through a woven Gore-tex window. Identical instruments have been deployed on Chajnantor (1997 October), at the CSO on Mauna Kea (1997 December), and at the South Pole (1998 January).

The data from this instrument are being analysed. Tasks include:

- Cross calibration between the submm tipper and other instruments, namely 225 GHz tippers on

Chajnantor and Mauna Kea, SCUBA, CSO, and AST/RO.
- Comparison of conditions at Chajnantor with Mauna Kea and South Pole.
- Continued development of the filter wheel.
- Possible deployment on Cerro Toco (5400-5500m) to investigate dependence of opacity with altitude in the area of Chajnantor.

FTS

To measure the atmospheric emission spectrum at Chajnantor, the Smithsonian Observatory is deploying a Fourier transform (polarizing Martin-Pupplet) spectrometer. This cryogenic instrument covers 200 - 5000 GHz with 3 GHz resolution and a 3 degree beam. Initial results are expected in the first half of 1998. NRAO is supplying support for field operations.

Seismometer

Installed in 1995. Data analysed by Chilean group (K. Bataille). Status of GPS rollover unknown.

Camera and temperature probe

- A surveillance camera was installed on 1997 June 15 facing southwest. Data are retrieved on tape and the images posted.
- A subsurface temperature probe was operated 1997 June - October and 1998 March - May. Data analysis is ongoing.

LSA equipment

In 1997 April, the LSA project deployed site test instruments at Pampa Pajonales, about 200 km south of Chajnantor. In 1998 June, this equipment was relocated on Chajnantor adjacent to (15 m north of) the MMA equipment. The LSA equipment includes:

- 12 GHz interferometer. Purchased from NRAO in 1996 September. Now set up parallel to the MMA test interferometer, but observing another satellite. Lag correlation of the data from the two interferometers will indicate the height of the turbulent layer (see MMA Memo 196).
- 3 channel, 183 GHz radiometer. Measures H_2O line shape. Designed and constructed by MRAO, OSO and ESO. Installed in LSA container, looking at zenith.
- Several weather stations. These are currently deployed adjacent to the containers, but will be deployed across the site in the last quarter of 1998.
- Ethernet link to MMA container.

A second 183 GHz radiometer is expected in 1998 September. Then these radiometers will be relocated to the ends of the LSA interferometer and look in the same direciton as the interferometer. Variations in the line shape will then be compared to the phase fluctuations measured with the interferometers.

Ongoing projects

Physiology studies

- John West, UCSD

Radiosondes

A surplus radiotheodolite was returned to the manufacturer for upgrade in 1998 June. We expect delivery toward the end of July. After some test and training flights in Tucson, we plan deployment in August or September. Flight protocols, cost sharing, and logistics will be negotiated between Cornell, ESO, SAO and NRAO.

Sodar

Engineering tests of ESO sodar unit are planned for 1998 July. How much interest we have in pursuing these measurements will be evaluated after these tests.

Cornell

DIMM measurements in 1998 May, July? etc.

NRO

1998 June campaign:

- dual 220 GHz tippers
- FTS at Pampa la Bola, Chajnantor, Cerro Toco?

Geology

- soil characterization for civil works

Last modified 1998 July 14
Simon Radford

Antenna Configurations for the MMA

Tamara T. Helfer & M.A. Holdaway
Last modified July 15 1980

1. Number and Size of Antenna Elements

We assume that the MMA will comprise N = 36 antennas of 10 m diameter. The geometric collecting area is then 2830 sq. m; the ''collecting length'' nD, the appropriate measure of the mosaicing sensitivity and for the fraction of occupied cells, is 360 m. Both the collecting area and collecting length are superior to the old MMA plan with N = 40 and D = 8 m. We note that the array configuration development plan will need to react to the changes and refinements in the array's concept, particularly with regard to possible collaboration with European and/or Japanese partners.

2. Limiting Configurations, Number of Configurations, and Resolution Scale Factor

Size of the Most Compact Array

The choice of a compact configuration for the MMA is driven by the desire to maximize surface brightness sensitivity, which is achieved by placing the antennas as close together as is practical. If we assume a filling factor f_{min} of 40%, which is a reasonable compromise between the competing requirements of close packing and the resultant maximum acceptable sidelobes, then the maximum baseline for the compact array is $b_{compact} = D_a\sqrt{N_a/f_{min}}$ = 95 +/- 5 m. This array would have the same resolution as a ring array that is about 66 +/- 4 m in diameter.

Size of the Largest Array

The largest configuration is assumed to have a maximum baseline of 3 km. If the sensitivity of the MMA is significantly expanded through a collaboration with the European and/or Japanese groups, then an array of 10 km diameter will be an attractive possibility.

Number of Configurations and Resolution Scale Factor

Given the assumed sizes of the minimum and maximum arrays, Holdaway (1998a) has performed a cost-benefit analysis for the number of MMA configurations, which showed that the observing efficiency of the MMA would be close to optimal with 4 configurations. We assign these four arrays the letters A (for the largest) through D (for the most compact). Given the described sizes of the D and A arrays, the resolution scale factor between adjacent configurations is about 3.6, and the configuration diameters are 95 m, 240 m, 840 m, and 3000 m. The minimum and maximum baselines for each array are listed in Table 1, along with the size of a sample beam and the time required for the half of the (u,v) cells to be sampled (FOC = 0.5). We note that the shortest baseline does not correspond to the largest angular structure to which the array will be sensitive, as mosaicing with total power data will permit arbitrarily large sources to be imaged.

Array	Minimum Baseline [m]	Maximum Baseline [m]	Array Style	Time for FOC = 0.5 [hours]	Natural Beam at 1 mm [arcs]
A	20	3000	?	10	0.06
B	18	840	?	2	0.20
C	16	240	?	0.1	0.75
D	12.8	95	filled	0	2.7

Table 1: Approximate specifications for the MMA's four main configurations.

3. Fourier Plane Coverage

The D and C arrays will provide essentially complete sampling of the (u,v) plane in a snapshot observation. The B and A arrays will require longer tracks for good imaging and sensitivity.

D Array

The driving considerations for the D array are maximum surface brightness sensitivity and excellent mosaicing capability. Surface brightness sensitivity is optimized by designing an array with the largest synthesized beam possible, which is achieved by having the shortest baselines possible. The MMA will be a *homogeneous array*, with total power and interferometric data being collected by the same antennas (Cornwell, Holdaway, and Uson, 1994). Homogeneous array mosaicing image quality is optimized by having a high density of the shortest interferometric baselines and by minimizing the sidelobes in the synthesized beam. Optimizing the short baseline coverage is best achieved with a filled array, which produces a Fourier plane coverage that to first order is a linearly decreasing function of (u,v) distance. The shortest baselines are limited strictly by the minimum safe distance which avoids mechanical collision of the antennas when pointing in arbitrary directions, which depends upon the antenna design, and less strictly by shadowing requirements. Configurations with the highest density of the shortest baselines will be a hexagonal close pack distribution of antennas, which results in a very large grating response in the synthesized beam and is therefore not acceptable. With a minimum distance between antennas of 1.28 D, we can achieve a reasonable sidelobe level of a few percent rms with an array filling factor of 40%. Such a filled compact array will result in complete instantaneous (u,v) coverage, even with only 32 antennas. Some degree of optimization is required for the zenith D array, trading off between good short baseline coverage and a large beam on the one hand and minimum synthesized beam sidelobes on the other.

D1, D2, and D3 Arrays For Observing Sources at Different Declinations

Since the beam shape will change with declination, it is always nice to have multiple configurations which are optimized to give circular beam shapes for both low elevation and zenith observations. However, the short spacing requirement for homogeneous array mosaicing, together with the physical inevitability of shadowing, absolutely require multiple compact configurations for observations of sources at various declinations. Current plans call for three compact arrays. The D1 array, with a North-South elongation of 1.2, will cover zenith observations down to somewhat below the onset of shadowing at 50 deg; the D2 and D3 will be progressively more elongated, with elongations of about 1.6 and 3. These three arrays will cover most of the range of declinations available from the Chajnantor site, -90 deg $< \delta <$ 53 deg. Observations of the small fraction of the sky which is further north and still visible from Chajnantor will need to be conducted in the C array or in a hybrid D+C configuration.

The D1 and D2 configurations will utilize a mechanical elevation stop which will limit the elevation to be above 20 deg. This limitation will permit a closer packing of the antennas in the D1 and D2 configurations. While it does remove flexibility from the compact arrays, the D1 and D2 arrays would be largely shadowed below this elevation anyway. The elevation stop will be removed from all or most antennas when they are reconfigured into the D3 array, where the antennas will be sufficiently separated that collisions are no longer a possibility. The general specifications for the D1, D2, and D3 configurations are shown in Table 2.

Array	Minimum N-S Distance	Elevation of first Shadowing	Minimum Observing Elevation	Maximum Observing Elevation	N-S Array Elongation
D1	1.3 D	50 deg	40-45	90	1.2
D2	1.9 D	31 deg	30	50+	1.6
D3	3.0 D	19 deg	14	33+	2.9

Table 2: Specifications for the three D configurations. Configurations D1 and D2 will have a mechanical minimum elevation limit of 20 deg.

We plan to optimize the three D configurations to allow for overlapping stations. The minimization of antenna stations may be important in keeping down the cost of building the pads, roads, and cables, though we still need to investigate what these costs are. Overlapping the stations will also keep the time involved in reconfiguring the antennas to a minimum. Also, we will need to design each D configuration with antenna access in mind: we should maximize the number of antennas that can be moved by a transporter without moving any other antennas.

C, B, and A Arrays

The C array will often be used for mosaicing, as the C array will have almost complete instantaneous (u,v) coverage, will have good brightness sensitivity, and will have an abundance of short baselines. Since long tracks are never needed to fill out the (u,v) coverage in C array, but rather to increase sensitivity, it will usually be better to observe a source within a few hours of transit to minimize the decrease in sensitivity caused by the atmospheric opacity at low elevations, and to observe over several days if more sensitivity is required (Holdaway, 1998b). Hence, the C array should be optimized for short tracks observed within a few hours of transit, over a range of declinations.

The B array requires about 2 hours to achieve essentially complete (u,v) coverage (FOC = 0.5, see Table 2), so it should be optimized for somewhat longer tracks, but still within a few hours of transit (Holdaway, 1998b).

The A array requires about 10 hours to achieve essentially complete (u,v) coverage. At +/- 5 hours off transit, the sensitivity loss due to the atmosphere will be severe at most frequencies; also, some sources are not above the minimum elevation limit for such long tracks. Nonetheless, the A array should be optimized for long integrations, keeping in mind that it must also have respectable snapshot coverage for those sources strong enough to be observed in this mode.

As a general requirement, we will want to have some of the shortest baselines (i.e., 15-20 m) present in even the largest arrays to permit single configuration mapping of many wide field objects (Braun, 1993). However, if there is a lot of large structure in the object, multiple configuration imaging may be required. At this point, we do not have a coherent strategy for when to combine data from multiple configurations, nor have we considered the impact of multiple configuration observations on the set of configuration designs.

Two philosophies are currently under consideration for the Fourier plane coverage for the C, B, and A array configurations. One philosophy is to achieve as uniform coverage in the Fourier plane as is practical; this approach leads to ringlike arrays, as characterized by Keto (1997) for snapshot observations and by Holdaway, Foster & Morita (1996) for longer tracks. The other philosophy is to minimize the sidelobes in the synthesized beam, as implemented by Kogan (1997, 1998a).

The Keto Approach

Keto's Releaux triangle configurations, and ring-like configurations in general, yield fairly uniform (u,v) coverage plus a narrow peak at small spatial frequencies. They also offer the advantage of achieving the maximal sensitivity for the longest baselines, resulting in smaller naturally weighted resolution than other types of arrays with the same maximum baseline, which is an attractive characteristic especially for the A array.

However, true uniform coverage in the Fourier plane has disadvantages as well:

- the sharp cutoff in (u,v) sampling at large spatial frequencies results in large (10-15%) sidelobes close to the central lobe of the synthesized beam (Holdaway 1997), which may complicate an image deconvolution and thereby lower its dynamic range (Holdaway 1996).
- optimization techniques like the elastic net method used by Keto have so far tended to produce large diameters for the central hole in the Fourier plane coverage. It is probable that this problem can be alleviated to some extent, either by using nested rings or Releaux triangles, or by changing the optimization conditions to include some number of short baselines. The nested triangle approach destroys the uniform Fourier plane coverage.
- unpublished simulations by Morita and by Holdaway show that the excess short spacing coverage which a ring array provides is actually more responsible for high dynamic range in wide-field

reconstructions than the uniform Fourier plane coverage. In other words, even if it were possible to get perfectly uniform Fourier plane coverage with a 36 element array, we probably would not want it.

The Kogan Approach

Kogan's algorithm produces antenna configurations which minimize the maximum sidelobe levels of the point spread function in some region of the image plane. This approach has the advantage of producing PSFs which should introduce fewer problems in image deconvolution. Kogan has also pointed out that in general, sidelobes that are close to the peak of the PSF can be alleviated using a taper (at the expense of image sensitivity), but that this is not true for sidelobes further out in the image plane. Another attractive feature of Kogan's approach is that it naturally shrinks the hole in the center of the (u,v) plane as the optimization extends over larger and larger regions in the image plane. This produces good coverage at short baselines in the (u,v) plane, which is one of the main shortcomings of the uniform (u,v) coverage optimization described above.

Kogan's code is flexible and can accept a variety of topographical constraints as inputs. At this writing, Kogan is investigating the arrays produced when the antennas are distributed within an annulus with a fixed outer radius and with varying inner radii. With a ''donut'' array like this, this configuration can deviate enough from a ring to taper the beam naturally, thereby reducing its sidelobes. We are currently trying to optimize the width of the donut constraint for the various different arrays. Our best guess is that it will be desirable for the C array to be a rather filled configuration (i.e. with a thick annulus), and that the A array will be much more ringlike, to meet the differing requirements of these arrays as outlined above.

The Kogan arrays are optimized for a snapshot in the zenith direction only; however, changing the declination should change only the positions and not the amplitudes of the sidelobes for a snapshot observation (Kogan 1998a). It should be noted that an array which has optimal sidelobes for a snapshot at transit may not be optimal for long tracks.

One disadvantage to Kogan's approach of minimizing the maximum sidelobe within some region of the points spread function is that rather large sidelobes can lurk just outside the region of optimization. It might be better to extend the region of optimization to the full width of the primary beam, and apply a weighting function which emphasizes the minimization of the close in sidelobes and gradually relaxes for the very far out sidelobes.

We plan to study the ramifications of these competing philosophies and ultimately to select a design based on imaging simulations of sources of different size and structure.

4. Sample Configurations

Since the configuration optimization is still in progress, we do not attempt to present optimized configurations in this document. However, to give the reader a feel for the arrays that the Keto and Kogan algorithms produce, we present sample configurations in Figures 1 through 4.

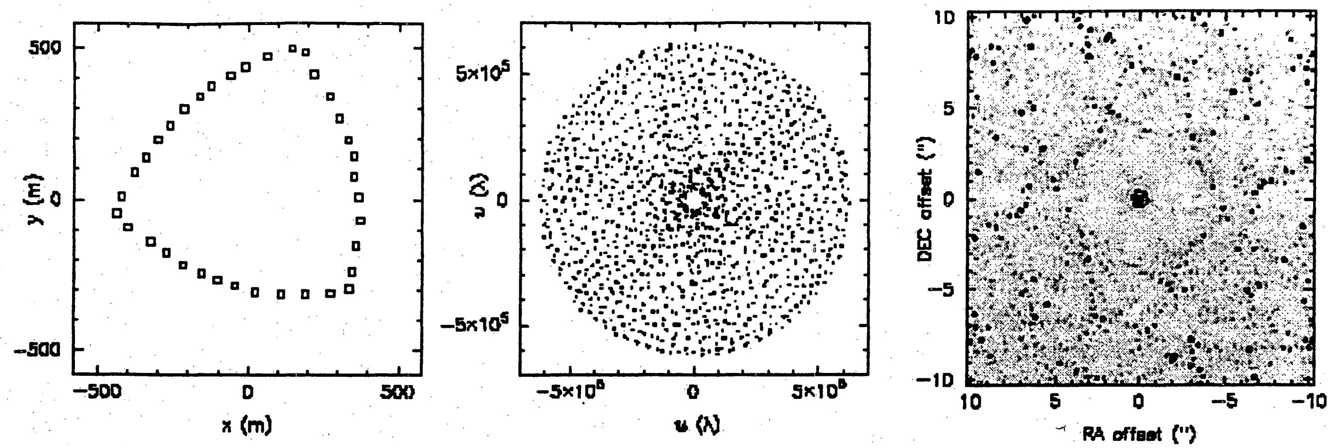


Figure 1: Sample Keto B Array snapshot at 230 GHz. (*left*) Antenna locations in meters, (*middle*) snapshot (u,v) converage, and (*right*) the resulting synthesized beam, with contours are at 10, 20, 40, 60, 80, 100%. Note the large inner sidelobes.

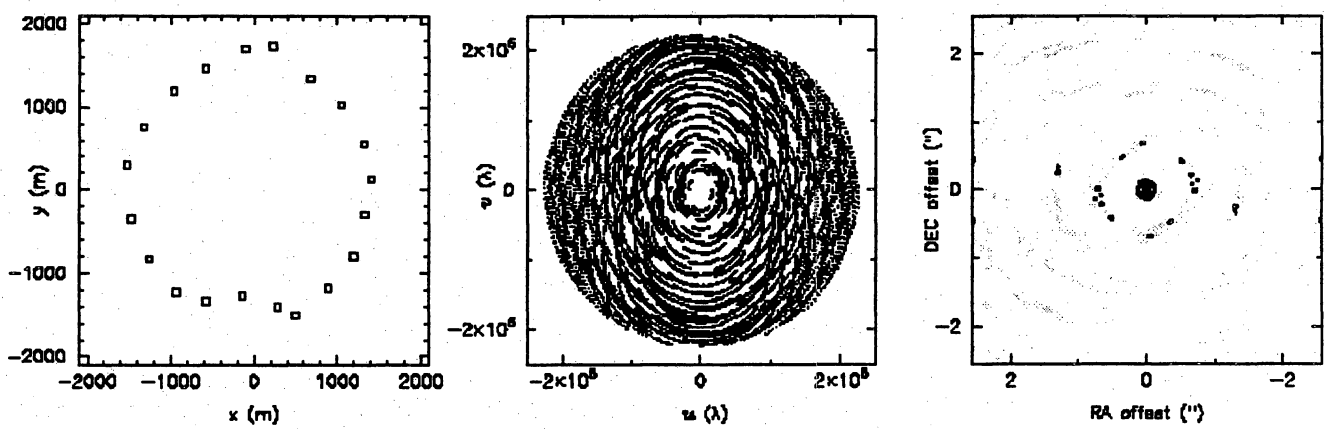


Figure 2: Sample A Array track for a Keto 20-element array optimized for 4-hour tracks (Holdaway, Foster, & Morita 1996), at 230 GHz. The contours are 0.05, 0.10, 0.15, 0.20, 0.40, 0.60, 0.80, 1.0. The outer sidelobes are reduced for long tracks, but the inner sidelobes remain high.

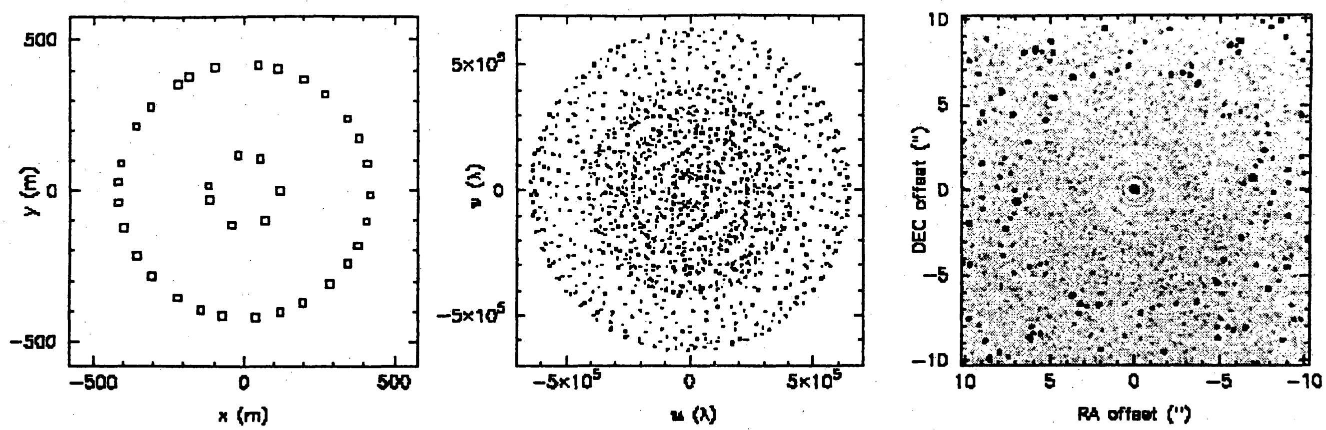


Figure 3: Sample Kogan B Array at 230 GHz. The contours are 10, 20, 40, 60, 80, 100%. The region of the image plane which was specified for the sidelobe minimization is a circular region at about 6" in radius.

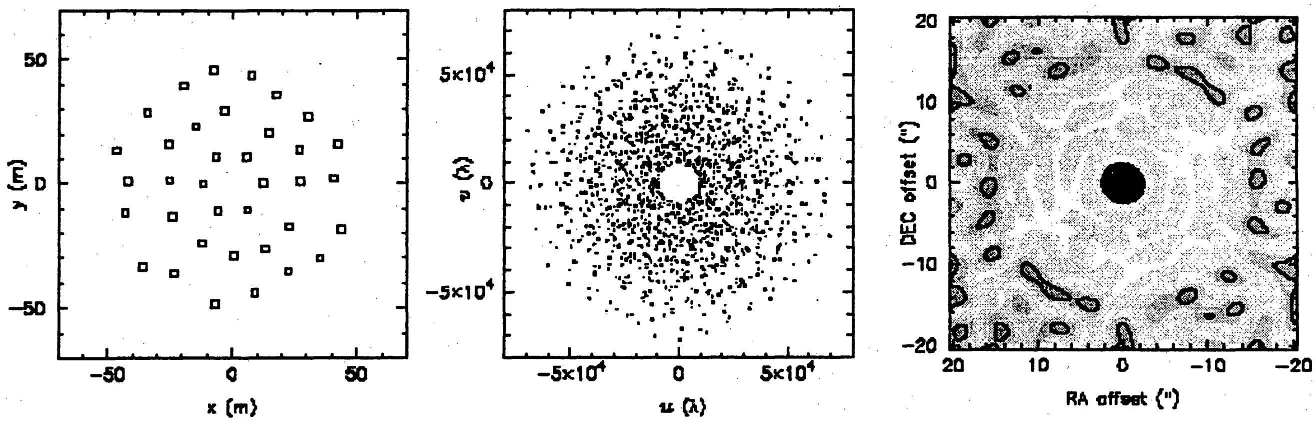


Figure 4: Sample zenith D array at 230 GHz. Except for the central hole, the Fourier plane density decreases nicely with (u,v) radius. This results in a synthesized beam with low sidelobes, shown here as 5, 10, 20, 40, 60, 80, 100% contours. Note that the primary beam of a 10 m antenna is about 30 FWHM at 230 GHz, so that the sidelobes are below 5% over essentially the entire primary beam.

5. Hybrid Arrays and Optimal Elongation for C, B, and A Arrays

From a study of the deviation of synthesized beams from circular as a function of source declination, Foster (1994) concluded that the optimal North-South elongation for tracks of varying length was in the range 1.1-1.3. In order to optimize the elongations of all of the arrays, it is important to know the expected source distribution with declination. Holdaway et al. (1996) assumed a model source distribution in order to estimate the pointing errors for the MMA antenna design. We are now in the process of looking at IRAS source distribution with declination in order to get a better estimate of this function.

Hybrid arrays, made by using stations in adjacent configurations, can be used to help minimize the shadowing and to achieve more circular beams for low-elevation sources. As stated above, a set of hybrid arrays is absolutely required for the compact configurations, but not so crucial for the larger arrays. Hybrid arrays will be studied more in the future when the basic arrays are better determined.

6. Interfaces With Other Parts of the MMA Project

- Antenna: minimum distance for close packing, hard elevation stops
- Antenna: transporter issues, such as intervening antenna clearance, road grade, etc.
- Site Development and Antenna: Antenna Pad Design.
- Site Development: Road Design.
- Local Oscillator/System: underground cables.

7. Other issues to be addressed

There are certainly other issues which have not been examined in this document which deserve closer attention. For example, all of the arrays will need to be optimized with respect to the Chajnantor site; the basic constraints of the site topography have already been implemented in both the Keto and Kogan algorithms. The arrays will also need to be optimized for different source declinations, or simultaneously for multiple declinations.

The configuration design process has been evolving rapidly in the past few months. In the month since the original draft of this chapter was written, four new MMA Memos have been submitted. Kogan (1998b) has described his optimization using ''donut'' constraints described briefly here in Section 3.2.2. Webster (1998) has investigated the idea of using nested rings or nested Reuleaux triangles to achieve a compromise between uniform (u,v) coverage and sensitivity to extended structure. Conway (1998) has explored a novel approach of having configurations laid out using a logarithmic spiral geometry. Finally, Kogan (1998c) has explored the possibility of constraining antennas to lie on two circles such that the inner circle of the A array is the same as the outer circle of the B configuration, the inner circle of the B array is the same as the outer circle of the C configuration, and so forth. We defer further discussion and analysis of these works to a future version of this document.

References

Braun, R., 1993, ''Telescope Placement at the VLA for Better Single Configuration Imaging'', VLA Scientific Memo 165.

Conway, J., 1998, ''Self-Similar Spiral Geometries for the LSA/MMA'', MMA Memo #216.

Cornwell, Holdaway, and Uson, 1994, ''Radio-interferometric imaging of very large objects: implications for array design'', A&A 271, 697-713.

Foster, S.M. 1994, ''The Optimum Elongation of the MMA A Configuration'', MMA Memo #119

Holdaway, M.A., 1998a, ''Cost-Benefit Analysis for the Number of MMA Configurations'', MMA Memo #199

Holdaway, M.A., 1998b, ''Hour Angle Ranges for Configuration Optimization'', MMA Memo #201

Holdaway, M.A., 1997, ''Comments on Minimum Sidelobe Configurations'', MMA Memo #172

Holdaway, M.A., 1996, ''What Fourier Plane Coverage is Right for the MMA?'', MMA Memo #156

Holdaway, M.A., Foster, S.M., Emerson, D., Cheng, J., & Schwab, F. 1996, ''Wind Velocities at the Chajnantor and Mauna Kea Sites and the Effect on MMA Pointing'', MMA Memo # 159

Holdaway, M.A., Foster, S.M., & Morita, K.-I. 1996, ''Fitting a 12 km Configuration on the Chajnantor Site'', MMA Memo #153

Keto, E. 1997, ''The Shape of Cross-correlation Interferometers'', ApJ, 475, 843

Kogan, L. 1997, ''Optimization of an Array Configuration Minimizing Side Lobes'', MMA Memo # 171

Kogan, L. 1998a, ''Opimization of an Array Configuration with a Topography Constraint'', MMA Memo #202

Kogan, L. 1998b, ''Opimization of an Array Configuration with a Donut Constraint'', MMA Memo #212

Kogan, L. 1998c, ''A, B, C, and D Configurations in the Shape of Concentric Circles'', MMA Memo #217.

Webster, A. 1998, ''Hybrid Arrays: The Design of Reconfigurable Aperture-Synthesis Interferometers'', MMA Memo #214.

SITE DEVELOPMENT

M. A. Gordon
Last revised July 13 1998

I. INTRODUCTION

Site development and operations (see Chapter 18) are closely linked. The physical development of an observatory affects how it operates. Yet, we need to know how an observatory will operate to build an effective physical plant. The plan for site development described below is a consequence of our present ideas for operations in Chile. This plan will surely evolve as we gain experience operating in Chile. Moreover, the Chilean custom of using high-quality prefabricated buildings in remote areas will allow us to easily make substantive changes in the physical plant as we discover what works and what does not.

II. CONCEPT

Large, sophisticated optical observatories have operated in remote areas of Chile for decades. Indeed, perhaps the most sophisticated observatory of all, the Very Large Telescope (VLT) of the European Southern Observatory (ESO) is now under construction at a remote site approximately 96 km (60 miles) south of Antofagasta. Our development plan results from inspections of these observatories, extensive discussions with their directors past and present, and our own experience operating radio telescopes in the United States for forty years. To succeed with the MMA, we will need to be flexible and creative.

A principal feature of the MMA operation will be remote or "service" observing. The local operations staff will make observations for astronomers. During these observations, these astronomers will be able to oversee the observing via an Internet connection and, when necessary, modify the instructions. This mode is now common for the NRAO's VLA and 12-m telescopes. This situation is similar at many other astronomical observatories. There will be exceptions, of course, especially during the early years of operations. Data will be transported to the astronomers probably in the form of magnetic tapes.

The MMA "site" in Chile will involve several locations. The observing site is the location of the instrument itself, the Llano de Chajnantor in the Andes mountains. The nearby village of San Pedro de Atacama will serve initially as the construction office and, with a scope that may change with time, the local operations center. During construction and operations, the MMA will need a business office near the port of Antofagasta. It will receive ocean, air, and overland shipments, export material to and from the telescope and local operating site, make local purchases, and interact with the regional authorities. The MMA will require an office in Santiago because Chile's head Foreign Office is the only place to process shipping documents associated with duty-free imports, no matter which Chilean port is used. Finally, the MMA will rely heavily upon the sophisticated resources of the NRAO sites in the United States. Each of these Chilean sites will need to be developed. Each of these sites will evolve differently because of changing requirements as the MMA moves from construction through interim operations into normal operations.

A. Llano de Chajnantor

The telescope site lies at an elevation of 5,000 m (16,500 ft) in Region II of Chile, at a latitude of 23S. Geologically, the site is a "bench" on the western side of the Andes range, with excellent drainage and a line-of-sight to a nearby community.

Logistically, this site has three important advantages: easy access, proximity to developed communities, and a gas pipeline. It lies near an international highway (Camino de Paso de Jama or International Route 27) connecting Chile with Argentina and now being paved. It lies within a 1-hour drive (approximately 55 km or 34 miles) east of the tourist village of San Pedro de Atacama (population 1,000); within a 2-hour drive (approximately 180 km or 110 miles) southeast of the mining support city of Calama (population 120,000) serviced daily by major Chilean airlines; and within a 5-hour drive (approximately 390 km or 240 miles) east of the port city of Antofagasta (population 300,000) also serviced daily by Chilean airlines. Finally, a new high-pressure gas line will pass by the periphery of the site, providing reliable and inexpensive energy to power gas-turbine electric generators.

Additionally, Chilean telephone companies are now installing broadband, fiber-optic cables and modern switching systems to link Chilean cities to accommodate the rapidly expanding economy. We can easily connect the MMA at Llano de Chajnantor into this network either by fiber-optic cable or state-of-the-art microwave links.

To make this site viable for the MMA, we will need to improve an 18-km (11-mile) existing mining road (with switchbacks) connecting the Paso de Jama highway with the site. It is likely that we might also want to improve an existing, straighter, 32-km (20-mile) mining road connecting the site to the Paso de Jama highway via the eastern side of the nearby hills Cerro Toco, Cerro Chascon, and Cerro Chajnantor.

Gas Atacama, an international consortium planning the gas pipeline between gas-rich northern Argentina and energy-thirsty northern Chile, has agreed to route the pipeline near but not across the MMA site. They will provide a gas tap at a place of our choosing, to allow us to site a gas turbine electric generating plant to power the MMA itself and the site support facilities. Energy provided to the site in this form should be reliable, inexpensive, and visually unobtrusive compared with the option of a high-voltage transmission line between Calama and the Llano de Chajnantor.

Potable water is a more difficult but solvable problem for the site. The Atacama desert receives little moisture, although the Andes mountains receive more because of the cooler temperatures associated with their altitude and their proximity to the wetter, eastern side of the range. The small, known accumulations of water on the western slopes of the Andes have already been committed to providing water for desert communities. For example, Antofagasta's water is piped about 320 km (200 miles) from the Andes to the coast. However, Andes water also collects into underground aquifers in the valley (Salar de Atacama) where San Pedro de Atacama lies. The town manager (and, independently, a well driller familiar with the area) has told us that good water adequate for MMA operations lies approximately 30 m(98 ft) below the surface, accessible through a well. The simplest way to supply the MMA site would be to process this water through a treatment plant near San Pedro de Atacama and truck the water to the site as needed. Alternatively, it may be possible to drill a well at the MMA site itself. Environmentally, it would be prudent to use this water sparingly at the site by installing low-water toilets and reprocessing grey water, if any. In effect, European Southern Observatory now does this to supply their VLT site at Cerro Paranal, approximately 96 km (60 miles) south of Antofagasta.

B. Operations Support

The remaining locations in Chile will serve to support operation of the MMA. Their functional aspects will change as the MMA moves from construction into operations and, correspondingly, so will the site characteristics. Largely, these changes will depend upon the individual preferences of the first few Chilean hires, on the Chilean economy at the time of their hiring, on the cash flow from the funding agencies, and what we learn is necessary to support the array.

1. San Pedro de Atacama

Initially, this village will surely be the center of construction operations. It is the closest community to the MMA site. Built to support its tourist industry, its few modern hotels could house temporary visitors to the construction operations. The highway connecting it with Calama is excellent. The runway of its small airport has just been paved and, possibly, commercial feeder flights may begin to use it in support of the tourist trade.

Land needed for the operations center and for off-center private housing may have limited availability. San Pedro de Atacama is a village of privately-owned land surrounded by government or "fiscal" land. Chilean law prevents foreign nationals from owning land within 100 km of an international border. The MMA should plan to lease fiscal land directly from the government or through an intermediary like the University of Chile or the Catholic University. Discussions so far indicate there will be no insurmountable problems in this regard.

The extent of the MMA development in San Pedro de Atacama will depend upon how construction proceeds. We are planning to send antennas to Antofagasta, Chile, by ship and to San Pedro de Atacama by road in the largest practical modules, so that a minimum of assembly will be required. If possible, we will assemble the antennas at the 2,425 m (7,960 ft) altitude of San Pedro, equip the antennas with cabling and electronics, and truck the completed antennas to the much higher-altitude MMA site. Such an operation would require a 2-storey assembly building, a machine shop, an electronics workshop, offices, a library, dormitories, a dining facility, a water treatment plant, and some recreational facilities such as tennis courts and swimming pool. The extent of these facilities will depend upon how quickly we procure the antennas, that is, ultimately upon the cash flow from the funding agencies.

After three antennas have been delivered to the site, limited observing can and should begin to test the array and to produce astronomical data. Such operations will require high-level support staff like computer programmers, electronics engineers, and support scientists generally unavailable locally. These employees, although temporary, may insist upon bringing their families to Chile and require the MMA project to provide family housing as well, some of which will surely be in San Pedro de Atacama.

As operations increases and construction wanes, the character of the work and of the support staff in San Pedro will change. Those remaining will tend to be "permanent" MMA employees, largely Chilean nationals, who will operate and maintain the array. Buildings suitable for construction will become unnecessary, and they should be removed. On the other hand, additional office and laboratory space will be required. Given the widespread use of high-quality prefabricated buildings in the Chilean mining industry, the MMA should also use such buildings in San Pedro de Atacama. Prefabricated buildings make it easy to change the physical plant to adapt to its changing function.

Eventually, MMA operations should become routine, very much like those of NRAO's VLA. Only a small staff may be required in San Pedro de Atacama, and most of the support staff could be moved to a Chilean city with more amenities such as Antofagasta where employee recruitment and retention will be easier. At this stage, the physical plant in San Pedro could be further reduced if prefabricated buildings are used.

Whatever the evolution of MMA operations based in San Pedro de Atacama, the MMA management needs to be sensitive to the character of the village. It is an international tourist destination because of its 16th century architecture, its geothermal areas, its pre-Columbian archeological sites, its indigenous Atacamañan residents, and its unique charisma. The village itself has strict architectural codes. Our location with respect to the village and the architecture of our buildings will affect our being accepted as desirable members of the community.

2. **Antofagasta**

In Region II, Antofagasta is *the* important city. It is the economic and administrative capital of the region. It is an international port. It has several commercial flights each day. It has at least one English-language international school. It has two small universities, and it is the largest city in Region II. The MMA organization will need to have a small office there to receive and ship goods, to buy supplies unavailable in Calama or San Pedro de Atacama, and on occasion to represent the interests of the MMA to the regional government.

As the MMA moves from construction into full operations, the role of the Antofagasta office could expand substantially. Because of the limited amenities in San Pedro de Atacama itself, most long-term MMA families will choose to live elsewhere. To support this expanded presence, the Antofagasta office would need to grow considerably.

3. **Santiago**

Chile is a country of about 14 million inhabitants. About 50% live in the

Santiago-Valparaiso area, and five million live in Santiago itself. Santiago's environs while smog-impaired during part of the year are exceedingly pleasant. It is an international city with lots of amenities. It has foreign-language schools that can prepare students adequately for admission to foreign universities, such as passing the US SAT and achievement examinations with high scores, or the German Arbitur exam, or the French Baccalauréat exam. Three of its universities, the University of Chile, the Catholic University, and the University of Santiago are among the best in South America. Most substantive Chilean companies maintain offices there. It is the entry point for most international flights. It is where you have to be to make and maintain important political and economic connections. Chileans and foreigners enjoy living there. Simply put, it is the capital of Chile.

The MMA operations will need a representative in Santiago. Shipping documents for duty-free imports can be processed only by the Foreign Office in Santiago. Specific goods are more available in Santiago than elsewhere in Chile. Visitors to the MMA will arrive first in Santiago.

What is in question is the extent to which the MMA will need facilities in Santiago. Unlike the Cerro Tololo International Observatory (CTIO), Carnegie Southern Observatory (CARSO), and the European Southern Observatory (ESO), the MMA is expected to operate primarily as a service telescope. Like the VLA, astronomers need not travel to the telescope to make excellent observations. Accordingly, no reception center will be required, no guest house will be required. When MMA staff and other visitors do come to Santiago, the large range of commercial hotels and restaurants in the city will suffice.

I believe that the MMA will need a small business office in the Vitacura or Providencia districts, staffed with one or two people. This staff will process customs documents, purchase and ship items unavailable in Region II, represent the MMA in governmental matters, and coordinate their activities with the CTIO and ESO offices now located there.

4. United States

The sophisticated support resources of the NRAO in the USA will be impossible to duplicate in Chile, owing principally to the diversity of the instrumentation maintained in the USA. While the MMA management will maintain its equipment as much as possible in Chile, the ultimate support will be the NRAO facilities in the United States. I would expect technical development of new sophisticated equipment and software to occur in the United States, as well as the identification and correction of subtle flaws in hardware and software, and support of MMA users located in the US.

III. DEVELOPMENT DETAILS

A. Llano de Chajnantor

1. Peripheral development

a. Access road from Paso de Jama to the site, 18 km (11 miles) with a double-asphalt surface, 1.6 km (1 mile) of guard rails on the switchback turns.

b. Gas tap on GasAtacama high pressure gas line, connecting pipe to gas turbine generator

c. Fiber-optics link from site to San Pedro de Atacama, approximately 56 km (35 miles) , Or, broadband microwave link (E-1 links) from site to San Pedro de Atacama

d. Water well, if economically feasible

2. **Actual site development**

a. Gas Turbine generator, 2 MW minimum (rated to produce 4 MW at sea level).

b. Diesel or Gas emergency generators, 1 MW minimum at altitude, to power cryogenics

c. Transformer station to switch between generators and alter voltages as required.

d. On-site roads, approximately 20 km, 7 m wide, compacted but unpaved, to connect pads with service buildings. (We need a specific layout to estimate this number accurately.)

e. Approximately 145 antenna pads, reinforced concrete, with signal and power connections.

f. Intra-pad signal (fiber-optic and coax) and power connections, approximately 20 km (12 miles). (We need a specific layout to estimate this number accurately.)

g. Water storage and distribution system to accommodate up to 20 workers.

h. Sewage disposal system.

i. Internal telephone system, data compatible

j. Internal power distribution system

k. Antenna barn with 3 bays, each 40 ft x 50 ft (6,000 ft^2 or 557 m^2), includes transporter repair station, with elevated partial pressure of oxygen.

l. Warehouse, 1,000 ft^2 (93 m^2), prefab

m. Control building & first-aid station, 15,000 ft^2 (1,393 m^2), with elevated partial pressure of oxygen, prefab

n. Emergency dormitory, 2,000 ft^2 (186 m^2), with elevated partial pressure of oxygen, prefab

o. Generator building, 2,000 ft^2 (186 m^2), prefab (?)

B. San Pedro de Atacama

1. Peripheral development

a. Electrical generation plant, 1 MW maximum, fired by natural gas if San Pedro is not on the Chilean power grid

b. Well, probably 30 m (98 ft) deep

c. Water treatment/ recovery plant

d. Sewage treatment

2. Actual site development

a. Laboratory, auditorium, & library building(s), 12,000 ft^2 (1,115 m^2), prefab
b. Antenna barn for assembling antennas, 2,000 ft^2 (186 m^2), prefab. (This 2-storey structure may conflict with local zoning laws.)

c. Warehouse, 4,000 ft^2 (372 m^2), prefab

d. Control & first aid, building(s), 8,000 ft^2 (743 m^2), prefab

e. Welding, carpentry, mechanical shop, 3,000 ft^2 (279 m^2), prefab

f. Dormitory, 8,000 ft^2 (743 m^2), *masonry* for acoustic isolation

g. Recreational building (s), 8,000 ft^2 (743 m^2), prefab

h. Fiber-optics or microwave link terminal

i. Electric power distribution system

j. Telephone system, internal, data compatible

k. Sophisticated communications facilities such as LAN, facsimile, connection to Internet

l. Houses, 5 @ 2,000 ft^2 (186 m^2), prefab, perhaps scattered through community

m. Outdoor recreational facilities

n. Parking lot for 40 vehicles

o. Security wall, adobe

C. Antofagasta

1. Actual Site Development

a. Offices, 2,000 ft^2, with garage space for two vehicles, rented

b. Transit warehouse, 2,000 ft^2

c. Sophisticated communications facilities such as LAN, telephone, facsimile, and connection to Internet.

D. Santiago

1. Actual Site Development

a. Offices, 2,000 ft^2 with meeting room, with garage space for two vehicles, rented

b. Sophisticated communications facilities such as LAN, telephone, facsimile, and connection to Internet

CONSTRUCTION, INTEGRATION AND INTERIM OPERATIONS

Robert Brown
Last revised July 13 1998

I. GENERAL PRINCIPLES

Organization of the construction phase of the MMA project will be structured around the interests and capabilities of the countries or international organizations that become partners with the NRAO in the MMA or in a conceptually larger array that subsumes the MMA. Because pending partnership arrangements are not yet finalized it is not possible to specify in detail how the array hardware will be constructed and by which organization it will be constructed. Nevertheless, there are clearly identifiable goals that will guide the construction phase of the MMA or an enlarged array irrespective of the partners that may become involved. These goals, or principles, include the following:

- The antennas, the single most costly piece of the array, will be fabricated and assembled by an antenna contractor. The antenna design will be the responsibility of the contractor and it will be the contractor's responsibility to validate the antenna performance to the specifications set for the MMA;

- As many instrumentation sub-assemblies as possible will be fabricated under commercial contracts. This would include such items as machining of the receiver dewars, and mounting of the correlator chips on the boards;

- There will be no in-house development and fabrication of hardware that can be purchased commercially. Items such as cryogenic refrigerators, cryogenic compressors, optical fiber transmission lines and connectors, and lasers for metrology or calibration all will be purchased;

- Fabrication of specialized RF devices including the superconducting millimeter and sub-millimeter wavelength mixer chips, discrete transistors or MMIC chips, and varactor diodes used for LO multiplication will be done under contract to commercial, university or government laboratories currently specializing in such work. No device fabrication facilities will be set up at the NRAO for the MMA. In addition, some part of the assembly of discrete devices into finished components such as mixer blocks, amplifiers, and frequency multipliers may be performed where appropriate and cost effective by commercial or other organizations. Evaluation and testing of components is likely to remain a function of the NRAO;

- Final assembly and test of the instrumentation systems--the integration of subassemblies--will be done by MMA personnel at the NRAO or at those observatories who are partners in the MMA or that larger array that subsumes the MMA. This includes assembly of the receiver inserts for each of the MMA frequency bands, assembly of those receiver inserts in the cryogenic dewars, assembly of the local oscillator system and assembly and wiring of the correlator boards into the correlator racks.

In all of the above the guiding principle is this: Fabrication of production quantity MMA hardware will be done under contract whenever the experience in fabricating such hardware exists within industry, university or government laboratories; when such experience exists only at the NRAO (or within participating partner observatories) the fabrication will be done at the NRAO with MMA personnel hired and trained for that purpose. The NRAO will not attempt to *teach* industry skills needed for the MMA that it does not already have, nor will the NRAO attempt to duplicate industrial skills in-house when needed services are available commercially.

II. CONSTRUCTION

The construction plan for the MMA is organized around two realities: First, the MMA will be built on a presently undeveloped site in the Altiplano of northern Chile; and second as an interferometric array the MMA is an assembly of multiple copies of functionally similar hardware and for this reason it can be *brought to life* incrementally. The former criterion, or reality, establishes a priority for development of the site infrastructure early in the construction phase of the project. The latter criterion provides an opportunity to make productive scientific use of early subsets of the array capabilities with those capabilities growing as hardware is added incrementally. But to exploit this opportunity demands that a high priority be placed in the construction phase on all *one-off* pieces of the array instrumentation; this would include things such as the correlator (or a subset of it), monitor and control hardware and software (or a functional subset of it), and the IF and LO signal transmission system (or a subset of it). These two realities are the cornerstones of the MMA construction plan.

Development of the array site on the Llano de Chajnantor at 5000 meters elevation and development of the array support facility near the village of San Pedro de Atacama that is located at 2425 meters elevation is the emphasis of the initial construction phase. The site development will involve construction of all the concrete antenna foundations, optical fiber and power communications to all the antenna stations, roads and site buildings. It will also include drilling for a source of water and plans for sewage disposal. Electrical power on site and communications between the site and the support facility near San Pedro are major construction enterprises. The plan for electrical power involves local generation from natural gas provided by a convenient tap to a gas pipeline that runs near the MMA site and is operated by the multi-national firm GasAtacama. There are no commercial facilities for communication between the site and the support facility near San Pedro; we will excavate a trench adjacent to the GasAtacama pipeline right of way and lay the cable for the 50 kilometers that separates these two operational sites. Development of the support facility will proceed apace with construction on the array site. The San Pedro base will incorporate laboratory, office and residential facilities suitable for the immediate needs of employees who are integrating the MMA instrumentation and suitable also for the longer term needs of the MMA operational staff.

As each antenna is completed at the contractors facility it will be assembled and validated. Afterward it will be shipped to the MMA Operations Support Facility (OSF) near San Pedro de Atacama (SPdA) with as little disassembly as possible consistent with the needs prescribed by ocean shipment. Once at the OSF it will be re-assembled and checked for optical and mechanical performance. It will be fully cabled and mated with its cryogenic equipment and receiver dewar. The initial dewar will contain a small subset of the eventual complement of receiver inserts. The RF performance of the antenna will be checked and, once verified, the fully-functional antenna will be transported from the OSF to the array site. This process will be repeated for each antenna and the MMA on site will grow incrementally in number of antennas on-site. As receiver inserts for new frequency bands become complete they will become part of the initial complement of instrumentation on each new antenna and they will be retrofitted to the antennas on site at that time. Thus the array will also grow incrementally in scientific capability; at each stage it will be capable of doing productive science.

III. SYSTEM INTEGRATION: ROLE OF THE ENGINEERING TEST INTERFEROMETER

Initially, integration of the MMA instrumentation will be impossible in Chile because neither the equipment nor the staff needed to verify compatibility and integration of new pieces of hardware will be in Chile. Everything and everyone will be in the U.S. or at partner observatories elsewhere. Because the initial integration and validation tasks are near-term in the construction phase and likely to be very demanding on the time of those individuals responsible for major instrumentation systems, we plan to carry them out on an engineering test interferometer made up of the first two MMA antennas; these two antennas initially will be located at the VLA site on the Plains of San Agustin in New Mexico.

As each new piece of equipment is completed, each new receiver insert for example, it will be installed initially on the test interferometer. The purpose of doing so is not to make scientific observations, but rather to verify that new instruments fit mechanically on the structure, that they receive the power and communications they need and that they do not interfere in any way with anything else. Once validated, subsequent quantities of those pieces of instrumentation can be sent directly to the OSF in Chile for installation on the present and future antennas there with the confidence that integration will go smoothly. The test interferometer is to streamline the engineering integration of the MMA in a resource-rich environment.

The engineering test interferometer will need support from software written not only to control the instruments and drive the antennas but also to handle the data flow. This provides an excellent opportunity to experiment on a small scale with the approaches and techniques being adopted for the MMA software and to give astronomers and array operators the chance to criticize the appearance and functionality of the software. Furthermore, an important role for the engineering test interferometer is to provide a platform on which software to support more sophisticated observing techniques or calibration modes can be developed and refined. As successive versions of the software are completed they will be ported to enhance continually the capabilities of the expanding array in Chile.

The operations staff in Chile will receive their initial training on the engineering test interferometer. By working in close consultation with the instrumentation and software developers the operations personnel will not only learn from the experts but they will establish the personal relationships that will bind the long term array operation in Chile to MMA research and development efforts based at the NRAO in the U.S. that will remain an important and continuing part of MMA operations.

When the last of the new instrumentation modules developed for the MMA has been successfully integrated into the engineering test interferometer and verified, and when development of new software is better done on the array in Chile, the engineering test interferometer will be disassembled and shipped to the site in Chile to be incorporated there as part of the final array.

IV. INTERIM OPERATIONS

Owing to its superb site, the fact that it will be in the southern hemisphere, and the quality of its instrumentation, the MMA will be capable of doing unique scientific observations in a very early stage of completion. The plan for interim operations of the MMA is to encourage exactly this. It requires provision for early completion of the things such as the signal transmission system on site and a correlator that may be a throw-away device or as the current planning projects, a subset of the final correlator. Making the developing MMA available to astronomer users as early as possible has the following advantages:

- It forces an early assessment, and refinement, of operational issues. Of special concern are issues involving staff recruitment and retention. An early understanding of what issues are important to employees and potential employees will permit problem solving to occur before issues become a crisis in full MMA operation;
- It provides a tangible motivator to people working on the project by illustrating clearly the purpose toward which everyone is working;
- It allows retrofits to be made to the hardware, software or communications systems such that the astronomer receives the product he or she would like to see; in doing so it gives the MMA staff

informed criticism early;

- It develops an educated user community early that can debug the array and speed the commissioning process;

- It brings the Chilean community of scientists, educators, administrators, government and labor officials on board the MMA observatory early enough in the operational phase that community voices can be heard at a time when procedural refinements can be facilitated. In this way the MMA can become an asset not just to astronomy world-wide but also an asset to Chilean institutions and the citizens of the Republic of Chile.

Interim operations can be controlled so that at all times it is a positive force for development of the MMA. Initially we can expect that interim operations will be restricted to the nighttime hours when construction work cannot be done on site. But even this will be a welcome asset to those astronomers who have long anticipated the chance to use the MMA.

Post-construction Operations

M. A. Gordon
Last revised July 13 1998

I. INTRODUCTION

Operating a complex radio telescope in Chile will be a new experience for the NRAO. While large astronomical observatories have successfully operated in Chile for decades, rural northern Chile lacks sophisticated technical resources and amenities. Our plan results from discussions with the director, the administrator, and two former directors of Cerro Tololo Interamerican Observatory (CTIO), with several administrators of the European Southern Observatory (ESO), with the director of Carnegie Southern Observatory (CARSO), with the project manager of the new Magellan Telescope being built at Las Campanas, and from our own experience operating radio telescopes in the United States. We have also discussed our plans with long-term employees of CTIO and ESO.

The plan described below is necessarily tentative. It presumes an operating mode that may take time to perfect. It presumes that we shall be able to find employees willing to live and work in northern Chile, which depends upon the Chilean economy there at the time the MMA is hiring. In reality, we know that conditions can differ from what we have recently ascertained. To succeed, the MMA operations management must be analytical, flexible, creative, and willing to build on the experience of the CTIO, CARSO, and ESO observatories. Most importantly, the staff in Chile must control the MMA, with the NRAO remaining in a primarily advisory and support role.

Chapter 16 above, Site Development, describes the physical plant we believe is necessary to operate the MMA. These chapters are linked.

II. CONCEPT

The MMA will operate somewhat like the Very Long Baseline Array (VLBA) headquartered in Socorro, New Mexico. It will be a "service" instrument, observing without astronomers present at the operations centers. Astronomers need not travel to Chile to observe, although they may choose to do so. Rather, this observing mode will free them from *having* to travel to the MMA to observe. In addition, service observing will give the local staff the freedom to juggle observing programs to match the current receiver status and atmospheric transparency. Such a mode requires the MMA to provide astronomers with the capability to monitor the observing over the Internet, so as to make program changes when necessary.

A. Operating Centers

Support of MMA operations will require four locales in Chile, and several in the United States. The instrument itself will be situated on the Llano de Chajnantor, a geologic "bench" at an altitude of 5,000 m (16,500 ft) in the Andes mountains east of the village of San Pedro de Atacama. The operations center will be located near this village because of its proximity and its lower altitude of 2,450 m (8,040 ft). The local business office probably will be in Antofagasta, a seaport as well as the capital of Region II of Chile. Finally, a small business office must be located in the capitol of Chile, Santiago, to process duty-free imports and to maintain contacts with the national government. The NRAO sites in the United States will oversee long-term technical development as well as offer high-level technical support when necessary.

Similar to the Very Large Array (VLA) in New Mexico, the principal operating center of the MMA may change with time. San Pedro de Atacama is a small village (population 1,000) with few amenities other than those required to support its tourist industry. Few employees families will want to live there for a long term, especially those with school-age children. As the MMA evolves into stable operations, we believe it likely that some aspects of its operations will move to a larger community with more amenities. Such changes could make long-term employment attractive to skilled professionals. The modern fiber-optic telephone network now being installed in Chile should easily facilitate this relocation. In this case, the San Pedro de Atacama facilities will become principally a maintenance facility.

B. Character of Chilean Operations

1. Management

Management decisions should be local ones. The MMA director should make all decisions involving operations in Chile. All employees in Chile should report to the MMA director, regardless of whether they are "permanent" Chilean hires or ones "borrowed" from related organizations such as the NRAO itself. The sponsoring organization, the NRAO, should confine its involvement with the MMA to financial, logistic, and technical support as well as general, observatory-wide policies such as access to and scheduling of the MMA.

2. Salaries and benefits

As far as possible, employee salaries and benefits should be uniform among all MMA employees. By the time the MMA moves into full operation, we expect that Chilean professional salaries will be competitive with a world market. These salaries would include job classifications and the salary steps within them. Exceptions would be temporary employees "borrowed" from other organizations. These employees probably would have continuing financial commitments at home. "Benefits" would include medical insurance, pension contributions, educational allowances, housing, and travel allowances where appropriate. Such benefits as well as work rules should be in strict accordance with Chilean law regardless of the eventual diplomatic or international status of the MMA organization.

3. Contracting support services

As is customary in the Chilean mining industry, the MMA should contract for commercial services when they are available. For example, Chile has several large companies that provide food service to remote locations. The employer need only supply a kitchen and dining room, and specify the variety and quality of the food to be served. This situation also applies to medical services, housekeeping services, payroll, and vehicle leasing and maintenance. The MMA should actually hire only those employees unavailable or inappropriate to obtain from commercial service companies, such as management and administrative personnel, support scientists, engineers, programmers, and mechanics. Not only is this system flexible and cost-effective through competition, but it also frees the MMA management to focus on topics and workers essential to the MMA operations.

Contracting for services is already common practice for the NRAO. Much of the Charlottesville, and all of the Socorro and Tucson facilities are situated on university campuses. They contract with the host universities for housekeeping, building maintenance, telephone, and other services where possible or appropriate. Kitt Peak National Observation (KPNO) provide Arizona Operations with food and telephone services on Kitt Peak. As a unit, the NRAO contracts for all of its payroll services.

III. STAFFING

A. Sistema de Turno employment for the MMA and its Operations Center

To operate the MMA in Chile, *all* consultants recommend a rotating shift system known in Chile as the "Sistema de Turno" for staffing the operations center and the maintenance of the MMA itself. In Chile the Turno system is used by all international observatories and most mining operations. It complies with Chilean labor laws. Typically, it consists of one week "on" and one week "off" in a manner to provide 80 to 88 work hours over a two week period. Variations are common. A construction project in a remote area east of Iquique operates on a two week "on" and a 10 days "off" system. Customarily, the employer provides room, board and transportation to and from an urban assembly point.

A Turno-like system is not new to the NRAO. Telescope operators on Kitt Peak, Arizona, work according to a similar system, called the "Fixed Salary, Fluctuating Work Week" or Regulation 778.114 of the US Labor Department. Despite the extra work required to calculate payroll and vacations, the Kitt Peak employees are enthusiastic about their work schedule. It affords them continuity for projects, family, and occasional second jobs during their "off" time.

An effective Turno system must be appropriate to the operation of the MMA. This system is not appropriate for management people who need to be continually available. It is also inappropriate for employees responsible for creating new systems or equipment. However, it works well for "interchangeable" personnel like telescope operators and maintenance people who must be available seven days a week, 24 hours a day. There is extra cost involved. Statistics show that it requires about 2.4 employees for every Turno position to ensure overlap and continuity.

Given the difficulties of staffing a location like San Pedro de Atacama, the Turno system may prove to be the only practical solution.

To accommodate a Turno system, the MMA would need to provide dormitories at its operations center near San Pedro de Atacama. Our advisors recommend that the dormitories be sized so that each Turno-employee could have the same room and the same bed each visit. In this way, that employee could leave personal effects in the room and could decorate the room to suit his or her preferences.

The MMA should establish pickup points for Turno employees only in Calama, at first. Region II has a network of modern, commercial buses linking its cities. Some of these buses serve San Pedro de Atacama more than once daily and, of these, a few continue on the Paso de Jama road into Argentina. The principle would be that commuting employees need to get themselves to the collection point by the most appropriate means and at their own expense.

Professional employees would either live in San Pedro and take substantial holidays as compensation for long hours on the job, or commute from elsewhere in Chile with some of the commuting time being considered working time. The VLA used a similar system for many years to transport employees from Socorro to the Plains of San Augustin.

B. Support offices in Antofagasta and Santiago

The Antofagasta and Santiago offices would not require a Turno system, nor would one be appropriate to their function. The Antofagasta office would process shipments in and out of the seaport and the airport, representing the MMA to the regional government when necessary, and purchasing supplies and services available in Antofagasta. The Santiago office would provide a similar role, with connections to the central rather than the regional government. For continuity , the same personnel should be available in these offices Monday through Friday. Each office might require only a small staff.

The role of these offices may change with experience. After 30 years of operating in Chile from La Serena, the CTIO has chosen to use only ports of entry near Santiago even though the city of La Serena is contiguous with the port of Coquimbo. The CTIO has found that the high traffic levels at the Valparaiso seaport and the Santiago International airport give the widest opportunities for shipping. Equally important, they have found that, in most cases, these ports are less expensive to use than the port of Coquimbo even though the Santiago goods must be trucked to and from La Serena. Because the Foreign Office in Santiago must process all papers for duty-free shipping regardless of which Chilean port is used, CTIO requires an Santiago representative.

An alternative scenario for the NRAO might be to withdraw all but essential operations support from San Pedro de Atacama to somewhere else in Chile just as the VLA operations has moved from the Plains of San Augustin to Socorro, use the Santiago ports of entry to ship and receive all international goods, and truck these goods overland from Santiago to Antofagasta and San Pedro de Atacama. In this case, the Santiago business office of the MMA would function similarly to CTIO's Santiago office.

IV. OPERATIONS COSTS

This section discusses the cost of operating in Chile. Section A describes estimates these costs in US dollars from data from different epochs; Section B, historical variations in the purchasing power of US dollars in Chile; and finally, Section C, the estimate for operations in Chile in terms of 1998 dollars.

A. **Operating Expenses in Chile**

There are several ways to estimate the operating costs of the MMA in Chile. One is to itemize and total all expected expenses; another, to total personnel salaries and benefits and divide by whatever fraction that is typical for similar operations in Chile; a third, to find a similar institution operating in Chile and adopt its operating budget adjusted by the number of employees. While all of these will be in error because of the MMA's uniqueness, the latter two methods may be least in error because they include operating expenses which the NRAO can not foresee.

For similar institutions, the ratio of salaries and benefits to total budget tends to be the same. In 1995 the NRAO value was approximately 0.73. For CTIO, the value is 0.70. Both values reflect an exceptionally tight funding climate. Operating at lower ratios makes more funds available for repairs and improvements. Our consultants recommend a value of 0.6 for Chile, especially for the early years of operations.

Table 1: MMA Personnel in the United States and Chile

Categories	Employees	Percent of Total
Scientists	9	8.8
Engineers and programmers	18	17.6
Supervisors	3	2.9
Clerical personnel	13	12.7
Technicians	44	43.1
Maintenance personnel	15	14.7
Totals:	102	100.0
Exempt employees	30	29.4
Non-exempt employees	72	70.6

In 1995 the NRAO estimated the personnel needed to support and operate the MMA in Chile. This estimate includes all MMA employees, that is, both in the United States and in Chile. Table 1 lists the categories. The estimate resulted from experience operating the VLA in New Mexico and from an analysis of the MMA requirements in Chile. The total includes additional employees to compensate for efficiency losses for the few employees working at the high-altitude MMA site on a daily basis. Salaries and benefits (31.5%) for these 102 employees total US$4.3M in 1994 dollars. Dividing the NRAO estimate for salaries and benefits by 0.6 gives an estimate of US$7.2M to operate the MMA in 1994 dollars.

The CTIO provides a good comparison for the MMA. In early 1995, the CTIO had 138 employees, of whom 20 were US hires and 118, Chilean hires. Salaried ''exempt'' positions were 29% of CTIO positions, which is what the NRAO model projects. The distribution of CTIO employee classifications corresponded well to the NRAO projections, subject to differences in job titles between the two observatories. CTIO salaries and benefits (33%) totaled US$5M. Scaling to the 102 projected MMA employees gives US$3.7M, and dividing by 0.6 predicts an operating budget for the MMA of US$6.2M in approximately 1995 dollars.

An accurate projection for the MMA in Chile involves additional factors. First, the MMA plans to contract for food and house-keeping services now provided in-house by CTIO employees. Second, the mix between temporary world-market and permanent Chilean hires may be different for the MMA than for CTIO, especially during its early years. Third, the remoteness of the MMA operations center will require significantly higher salaries than observatories operating from cities like Santiago and La Serena. Geographic adjustments of 20 to 30% above Santiago salaries are common for Region II. Fourth, its remote location may require larger transportation budgets than CTIO's La Serena location. Finally, scaling the unusually tight CTIO budget may provide an unrealistically low estimate for operating costs. Considering the projections described above and these factors, we believe that U$8M in 1995 dollars is a reasonable estimate for routine operations of the MMA in Chile.

In addition to the cost of routine operations in Chile, the MMA will have a continuing need to develop new instrumentation and software. These developments will provide additional scientific capabilities for the telescope as new technology becomes available. An annual investment of even 2% of the capital cost of the MMA would provide a budget of US$4.5M for the development of equipment and software that should be considered as part of the annual operating costs. Furthermore, the power of the MMA will depend in part upon its real-time accessibility to astronomers worldwide. We estimate the cost of supporting this access--through the Internet and via satellite links-- to be about US$1M. Finally, the the need for a shared rotation of the highest level staffing of the MMA between the operations center and NRAO facilities in the US to be about US$0.5M per year.

B. Purchasing Power of US Dollars in Chile

The purchasing power of the US dollar in Chile is driven by market forces. These include how much Chilean banks are willing to pay for US dollars on the international market, how much Chileans are willing to pay for foreign-made goods, and the cost of living in Chile. At this writing, Chile has no national debt. Its economy is expanding. It has a favorable trade balance with the US and, consequently, more dollars than it needs.

Variations in the purchasing power of dollars in Chile also involves the relative price inflation of both countries. Consistent with economic practice, both countries track inflation through a variety of consumer and wholesale price indices. A principal US index --there are many -- is the seasonally adjusted Consumer Price Index or " CPI", published by the US Federal Reserve Data Bank (FRED) and available on the World Wide Web. For Chile, the approximate equivalent is the Indice de Precio de Consumador or "IPC". Both indices are model dependent, they are calculated on the basis of a hypothetical "market basket" of a typical family that may or may not apply to the MMA situation.

Until recently, Chile has experienced high rates of inflation. It has become common practice for Chilean companies to write contracts, and in some cases pay salaries, in units of the Unidad de Fomento (promoted unit) or UF. The Chilean government adjusts the UF to compensate for variations in the internal buying power of the Chilean peso.

Specific exchange rates between the US dollar and the Chilean peso are the Dolar Acuerdo (agreed rate), the Dolar Informal (informal rate), and the Dolar Interbancario (interbank rate). The Dolar Acuerdo reflects the number of pesos per dollar in contracted transactions. The Dolar Interbancario applies to mercantile and financial transactions within the banking industry. The Dolar Informal is the rate that results from tourists and foreign companies exchanging pesos for dollars as needed.

Combining the IPC and CPI with the Dolar Informal rates experienced by AURA over the last several years, I believe that the buying power of the US dollar in Chile has fallen at an annual rate of about 6%. Over a longer period, the rate is closer to 8%.

Scaling budgets to future years is difficult. Despite more than thirty years of experience, CTIO has found it impossible to predict the variation in the buying power of the US dollar in Chile. Market forces affecting the dollar/peso exchange rate and the US and Chilean inflation rates are too complex to predict. The CTIO management recommends that projections be limited to only one year in the future. Last year, the dollar's purchasing power in Chile fell by approximately 8%.

C. Projecting MMA Operations costs in US dollars

Summing the cost estimates described in Section A above and adjusting these estimates to 1998 US dollars through considerations discussed in Section B, we estimate the annual cost of operating the MMA to be approximately US$15M 1998 dollars.

Considering the expected impact of the MMA upon our understanding of the universe in which we live as well as the costs of the powerful and successful Hubble Space Telescope, we believe the cost/benefit ratio of operating the MMA will be remarkably reasonable.

I am grateful for the suggestions of Glen Blevins, Hernán Bustos, Robert Brown, Enrique Figueroa, Jeffrey Kingsley, Amanda Muñoz, Peter Napier, Angel Otárola, Frazer Owen, Monroe Petty, Víctor Realini, and Dale Webb in preparing this specific material and that of Chapter 16.

SCHEDULE AND TIMELINE

Richard Simon
Last Changed July 20 1998

Introduction

This chapter will outline the schedule and project planning for the Millimeter Array Project. There are two key aspects of planning for the MMA:

- Tasks and milestones which must be accomplished
- Associated target dates.

The logical structure for the project is built around the concept of a "Work Breakdown Structure", or WBS. The WBS is simply an outline plan of all the work to be accomplished, and provides an ideal framework for scheduling, costing, and tracking progress. Once a baseline WBS has been created, the inevitable changes and unexpected developments any real world project experiences are easily incorporated into the WBS, and the impact of problems or unexpected difficulties can be allowed for.

There will be two principle tables maintained in this chapter, and updated as necessary:

1. The Project WBS, expanded to level 3.
2. A summary table of major milestones and target dates.

At the current time, both these tables are still in draft format and will certainly be revised significantly. The format and numbering of the schedule will be tied directly to the WBS, but is not in the current draft. The schedule dates and format are a very early draft with known inconsistencies. At this early point in the development of the WBS, there are also disparities in the depth of planning; some sections have progressed only to the planning needed for the Design and Development phase, while other sections include details covering the entire MMA project.

Table 1

Millimeter Array Project Work Breakdown Structure to Level 3

1			**Administration**
	1.1		**Management**
		1.1.1	Engineering
		1.1.2	Science
		1.1.3	Business
		1.1.4	Schedule
		1.1.5	Budget

1.1.6 Personnel

1.2 Facilities

1.2.1 Additions to Tucson Laboratories
1.2.2 Modifications to AOC Space
1.2.3 CDL Rental Space and Remodel
1.2.4 Common Infrastructure

1.3 Agreements in Chile

1.3.1 Site Use Agreements
1.3.2 OSF Agreement
1.3.3 Scientific Partnerships

1.4 Partnerships

1.4.1 LSA/MMA with ESO
1.4.2 International
1.4.3 Domestic

2 Site Development

2.1 Environmental Impact Survey (EIS)

2.1.1 Survey potential contractors
2.1.2 Establish contract requirements
2.1.3 Select vendor
2.1.4 Monitor contract
2.1.5 Assess Impact Survey

2.2 Soil sampling

2.2.1 Survey potential contractors
2.2.2 Establish contract requirements
2.2.3 Select vendor
2.2.4 Monitor contract
2.2.5 Assess Soil report

3 Antenna

3.1 In-house designs

3.1.1 NRAO Design
3.1.2 OVRO Design

3.2 Specifications

3.2.1 Project Book
3.2.2 PDR
3.2.3 CDR

3.3 Procurement of Prototype Antenna

3.3.1 RFP
3.3.2 Vendor Information Meeting
3.3.3 Evaluate Bid
3.3.4 Sign Contract
3.3.5 Prototype Antenna Contract Supervision
3.3.6 Prototype Antenna Acceptance

3.4 Apex

4.2.1 Specifications
4.2.2 Circuit design
4.2.3 Mechanical design
4.2.4 Order parts
4.2.5 Assemble
4.2.6 Test

4.3 Conceptual Design

4.3.1 Specifications
4.3.2 Optics Decision
4.3.3 Multiple Dewar Decision
4.3.4 Provision for 30 GHz Rx
4.3.5 Antenna Interface
4.3.6 Front-End Layout
4.3.7 Rx Configuration
4.3.8 Rx maintenance procedures
4.3.9 Rough Cost Estimate
4.3.10 Detailed project planning

4.4 Dewar

4.4.1 Feed through selection
4.4.2 Design
4.4.3 Engineering
4.4.4 Final Design
4.4.5 Manufacturing
4.4.6 System integration support

4.5 Cryogenics

4.5.1 Cryogenics selection
4.5.2 Radiation shield development
4.5.3 Verification of heat load calculations
4.5.4 Design
4.5.5 Engineering
4.5.6 Final Design
4.5.7 Manufacturing
4.5.8 System integration support

4.6 Optics

4.6.1 Window
4.6.2 IR Filter
4.6.3 Lens
4.6.4 Feed

4.7 LO Injection

4.7.1 LO Leveling decision
4.7.2 Design
4.7.3 Engineering
4.7.4 Final Design
4.7.5 Manufacturing

4.7.6 System integration support

4.8 Polarization Diplexor

4.8.1 Specification
4.8.2 Development
4.8.3 Design
4.8.4 Engineering
4.8.5 Final Design
4.8.6 Manufacturing
4.8.7 System integration support

4.9 Electronics

4.9.1 Bias Circuitry

4.10 Receiver Integration

4.10.1 Components Integration/Receiver Package
4.10.2 Test & Evaluation
4.10.3 Documentation

4.11 Test & Calibration System

5 LO System

5.1 Specifications

5.2 LO Reference

5.2.1 LO Subsystem Specifications
5.2.2 Test Equipment
5.2.3 375 MHz Synthesizer
5.2.4 1.5.3.4 1.5 GHz Synthesizer
5.2.5 8 GHz Synthesizer
5.2.6 10 GHz Synthesizer
5.2.7 12 GHz Synthesizer
5.2.8 14 GHz Synthesizer
5.2.9 10 GHz Round-trip Phase Meter
5.2.10 3.2-5.2 GHz Synthesizer/PLO/Fringe Generator
5.2.11 Frequency Standard: 1 Hz, 1 MHz, 1 GHz, 10 GHz"
5.2.12 Fringe & Phase Switch Synthesizer. 100 MHz (Fig.5&6)
5.2.13 VCXO Cleanup Loop (Fig.5)
5.2.14 16 GHz Synthesizer
5.2.15 26 GHz Synthesizer
5.2.16 10-15 GHz Synthesizer f1 (Fig.5&6)
5.2.17 Bins & Interconnects
5.2.18 Integration & Test

5.3 Multiplier Chain LO

5.3.1 Specifications
5.3.2 General R&D
5.3.3 Multiplier R&D
5.3.4 Production for Prototype Receiver

5.4 Photonic LO

5.4.1 Photonic Phase Cal System

5.4.2 Photonic LO Development

5.5 Selection Chain/Photonic

5.6 Production

5.7 Continued Support

6 IF System

6.1 IF System Specifications

6.2 Test Equipment

6.3 4-12 GHz Antenna Converter (Fig.2)

6.4 De-Multiplexer with Matrix Switch (Fig.3)

6.5 2-4 GHz Base-Band Converter (Fig.4)

6.6 Bins & Interconnect

6.7 Integration & Test

6.8 Decision: Base-Band Converter vs DSP

7 FO System

7.1 OF System Specification

7.2 Fiber Specification

7.3 Cable Specification

7.4 Tx/Rx for IF

7.5 Tx/Rx for LO

7.6 Tx/Rx for Round-Trip Phase Meter

7.7 Tx/Rx for M/C

7.8 Bins & Interconnect

7.9 Integration & Test

8 Correlator

8.1 Specifications

8.1.1 Capabilities

8.1.2 Configurations

8.2 Test Correlator

8.2.1 Specifications

8.2.2 Adapt GBT design

8.2.3 Order parts

8.2.4 Assemble

8.2.5 Test

8.2.6 Deliver to VLA site

8.3 Preliminary Design

8.3.1 Evaluate overall architecture

8.3.2 Define board functionality

8.3.3 Straw man system layout

8.3.4 Evaluate thermal and power characteristics

8.4 Sampler

8.4.1 Specifications

8.4.2 Circuit design
8.4.3 Chip design
8.4.4 PC board layout
8.4.5 Mechanical design
8.4.6 Chip fabrication
8.4.7 PC board fabrication
8.4.8 Test fixture
8.4.9 Prototype assembly
8.4.10 Prototype test
8.4.11 Design modifications
8.4.12 Fabricate, assemble and test with design mods"

8.5 FIR Filter

8.5.1 Target specifications
8.5.2 Computer simulation
8.5.3 Implementation decision
8.5.4 Circuit design
8.5.5 Chip design
8.5.6 PC board layout
8.5.7 Mechanical design
8.5.8 Chip fabrication
8.5.9 PC board fabrication
8.5.10 Test fixture
8.5.11 Prototype assembly
8.5.12 Prototype test
8.5.13 Design modifications
8.5.14 Fabricate, assemble and test with design mods"

8.6 Custom Boards

8.6.1 Memory Board
8.6.2 Correlator Board
8.6.3 Long-Term Accumulator Board
8.6.4 System Control Boards

8.7 Correlator Chip

8.7.1 Specifications
8.7.2 Select vendor
8.7.3 Chip design
8.7.4 Prototype chip fabrication
8.7.5 Prototype chip test
8.7.6 Design modifications
8.7.7 Fabricate and test design mods
8.7.8 Fabricate production run

8.8 Racks

8.8.1 Design mechanical layout
8.8.2 Design power distribution
8.8.3 Select control computer

8.8.4 Design control wiring
8.8.5 Design signal wiring
8.8.6 Order parts
8.8.7 Assemble prototypes

8.9 Post-Correlation Processor

8.9.1 Specifications
8.9.2 Select data processing hardware
8.9.3 Design rack
8.9.4 Order parts
8.9.5 Assemble
8.9.6 Test

8.10 Software

8.11 Prototype Correlator Production

8.11.1 Determine configuration
8.11.2 Order parts
8.11.3 Assemble
8.11.4 Test
8.11.5 Deliver to VLA site

8.12 Site Correlator Production

8.12.1 Determine configuration
8.12.2 Order parts
8.12.3 Assemble
8.12.4 Test
8.12.5 Deliver to MMA site

8.13 Continued Support

9 Computing

9.1 Requirements Specification

9.2 Software Standards

9.3 Monitor & Control

9.4 Data Production

9.5 Near Real-time imaging

9.6 Scheduling

9.7 Archiving

9.8 Off-Line Data Processing

10 System Integration

10.1 Overall Specifications for all systems

10.2 Project Book

10.3 Specification of interfaces and standards

10.3.1 Mechanical
10.3.2 Electrical/electronic
10.3.3 Logical

10.4 Monitor and Control

10.4.1 MC Specification

	10.4.2	MC Board Specification
	10.4.3	MC Board available
10.5		**Configuration control**
10.6		**Test Interferometer**
10.7		**Holography System**
	10.7.1	Holography Rx/Tx Design
	10.7.2	Specifications
	10.7.3	Fabrication & Assembly
	10.7.4	Test and Calibration
	10.7.5	Integration Holography System/Antenna
10.8		**Testing**
	10.8.1	Central LO/Transmission/Antenna LO
	10.8.2	Antenna IF/Antenna LO
	10.8.3	Receivers/Antenna IF/Transmission
	10.8.4	Central IF (including baseband)/Central LO
	10.8.5	Transmission/Central IF
	10.8.6	Central IF/Sampler
	10.8.7	Central IF/Sampler/Delay and Correlator
	10.8.8	Monitor and Control/All Subsystems
	10.8.9	Round-trip phase with fiber on drum
	10.8.10	Phase stability and phase noise, injected coherent cw"
	10.8.11	Dewar mounting in Antenna
	10.8.12	IF, LO, Transmission components mounted on antenna"
11		**Calibration & Imaging**
11.1		**Site Characterization**
	11.1.1	Atmospheric characterization and monitoring
11.2		**Configuration Studies**
11.3		**Radiometric Phase Correction**

Table 2

Millimeter Array Project Key Milestones and Dates

Facilities construction in Chile (San Pedro & High Site)

	Start Date	Stop Date
High site permit process	1 Jun 98	31 Dec 01
San Pedro Permit process	1 Oct 00	31 Dec 01
San Pedro Facilities A&E design work	1 Mar 01	1 Oct 01
High Site A&E design work	1 Mar 01	1 Oct 01
High Site initial civil works (roads, power, utilities)	1 Jan 02	30 Nov 02
High site inner Roads complete		1 Dec 02
San Pedro Lab, Residence and Office facilities available		1 Dec 03
High Site antenna erection facilities available		1 Dec 03
High Site control room facilities available		1 Mar 04
Antenna Pads 1-10		1 Mar 04
Antenna Pads 10 - 100	2 Mar 04	1 Mar 05
Remaining antenna pads, including utilities, cabling	2 Mar 05	1 Mar 06

Antenna procurement

	Start Date	Stop Date
Antenna PDR		28 Jul 98
Antenna specs writing, CDR	30 Jul 98	1 Nov 98
Final bid package for prototype antenna(s)	2 Nov 98	31 Dec 98
Prototype antenna bids due		1 Apr 99
Prototype antenna contract award		1 Jun 99
1st Prototype antenna delivery w/ some tests		31 May 01
Outfitting 1st prototype	1 Jun 01	31 Aug 01
Single Dish pointing and gain tests, prelim holography	1 Sep 01	30 Nov 01
Final Holography tests	1 Dec 01	31 Dec 01
2nd Prototype antenna delivery		1 Jan 02
Outfitting 2nd prototype antenna	1 Jan 02	30 Mar 02
Interferometric testing (first fringes)	1 Apr 02	31 Dec 02
Final antenna acceptance		1 Jan 03
Antenna contract negotiation	2 Jan03	1 Jun 03
Award Antenna Contract		2 Jun 03
Antenna 3-11 delivered	1 Mar 04	1 Mar 05
Antenna 12-20 delivered	1 Mar 05	1 Mar 06
Antenna 21-29 delivered	1 Mar 06	1 Mar 07
Antenna 30-37 delivered	1 Mar 07	1 Oct 07

Antenna outfitting efforts on the high site

Assume single erection crew transitions to operations => ~8 antennas/year

Multiple passes for some or many antennas, as receiver inserts become available

	Start Date	Stop Date
Erect antenna 3 on high site	1 Mar 04	1 Jun 04
Outfit antennas 4 to 11 on high site	1 Jun 04	1 May 05
Initial observations with 8 elements	1 May 05	
Outfit antennas 12-20 on high site	1 May 05	1 May 06
Outfit antennas 21-29 on high site	1 May 06	1 May 07
Outfit antennas 30-37 on high site	1 May 07	1 Oct 07

Amplifier and receiver fabrication

	Start Date	Stop Date

Receiver fabrication

	Start Date	Stop Date
Mass production of receivers begins	1 Mar 03	
Assembly & integration of first 3 receiver inserts	1 Jun 04	
Delivery of Receivers to high site	1 Sep 04	

Antenna electronics and fiber optics

	Start Date	Stop Date
Mass production of back-end electronics begins	1 Mar 03	
Delivery of first production racks to high site	1 Sep 04	

LO System

	Start Date	Stop Date

Correlator

	Start Date	Stop Date