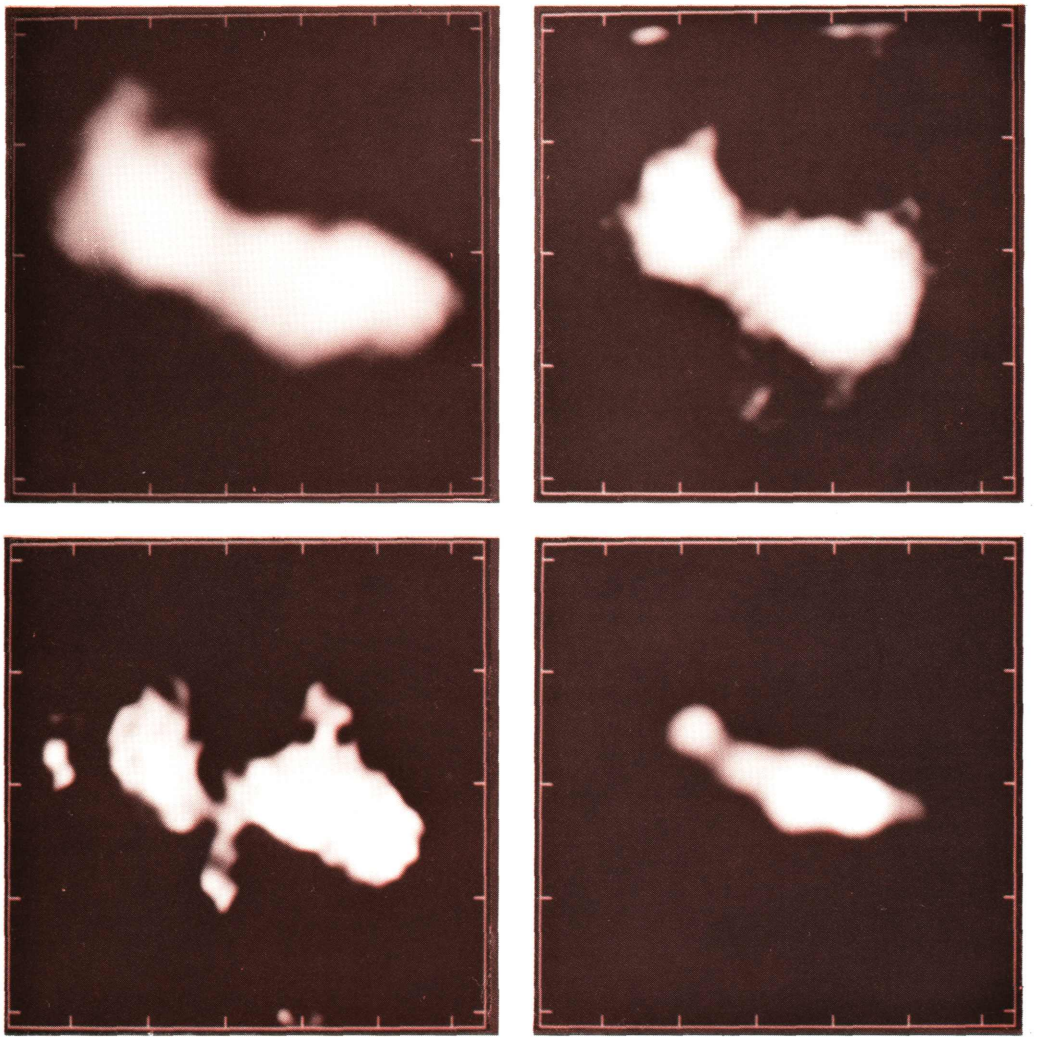
MILLIMETER ARRAY DESIGN CONCEPT

MMA DESIGN STUDY VOLUME II

MILLIMETER ARRAY



January 1988

The National Radio Astronomy Observatory is operated by Associated Universities Inc., under contract with the National Science Foundation.

Cover Illustration: Hat Creek Interferometer maps of the nuclear region of the starburst galaxy M82. (1) an integrated intensity map of CO ($J = 1$–0) at 8″ resolution; (2) integrated intensity map of HCO^+ ($J = 1$–0) at 10″ resolution; (3) integrated intensity map of HCN ($J = 1$–0) at 10″ resolution; and (4) the 3.3 mm continuum at 6″ resolution. HCN and HCO^+ require much higher densities than CO to be excited, and thus the maps reveal the densest regions of the molecular clouds. HCN and HCO^+ have very similar excitation requirements, so the differences in the maps are probably due to an enhanced abundance of HCO^+ in the most active part of the starburst. The 3.3 mm continuum is predominantly free–free emission from gas ionized by young stars. [J. E. Carlstrom (1988), "Aperture synthesis maps of HCN, HCO^+, CO and 3 mm continuum", in *Galactic and Extragalactic Star Formation*, R. E. Pudritz and M. Fich, Eds., (D. Reidel: Dordrecht), in press.]

NATIONAL RADIO ASTRONOMY OBSERVATORY

EDGEMONT ROAD CHARLOTTESVILLE, VIRGINIA 22903-2475
TELEPHONE 804 296-0211 TWX 910 997-0174

April, 1989

Dear Colleague,

The enclosed materials are intended to provide you with an overview of the current state of our planning at the NRAO for a U.S. millimeter-wave synthesis telescope (the **Millimeter Array**, or **MMA**, as it has come to be called). Included are the following:

- A document describing the status of recent MMA planning, with an emphasis on those scientific arguments that are the most driving, in the sense of best serving to define the instrument: *The Millimeter Array Design Concept*, MMA Design Study Volume II, NRAO, January 1988.
- A summary of the progress made, and decisions reached, in the last twelve months. Separate summaries are provided, in the categories listed below, based on the efforts of the various MMA Working Groups:

  Site and Configuration

  Antennas and Telescope Optics

  Mosaicing, and the need (if any) for a Central Element

  Receivers

  Correlator.

- An abstract of the issues which were identified by the MMA Advisory Committee, at its meeting in February 1988, as most in need of further attention. These comments provided the stimulus for the work summarized above.

During 1989 we will continue to refine the MMA design. Many new features in the design will reflect new requirements which have resulted from the experience gained by scientific investigations on existing millimeter-wave interferometers. At the end of this year we will circulate a draft MMA proposal for review by the entire astronomical community.

I hope that you will find this material to be of interest. We welcome your suggestions and criticisms, as well as your active participation in any aspect of the MMA project.

Sincerely,

Robert L. Brown

OPERATED BY ASSOCIATED UNIVERSITIES, INC., UNDER COOPERATIVE AGREEMENT WITH THE NATIONAL SCIENCE FOUNDATION

A BRIEF OVERVIEW OF THE MILLIMETER ARRAY

The Millimeter Array design concept, summarized in Volume II of the MMA Design Study, describes a versatile imaging instrument which emphasizes the following capabilities:

- Sub-arcsecond imaging at 115 GHz and higher frequencies;
- Wide-field imaging, mosaicing;
- Rapid imaging, "snapshots" of high fidelity;
- Sensitive imaging at high frequency ($\geq$ 350 GHz);
- Simultaneous multi-band operation.

Together these capabilities define a unique instrument; astronomers using the MMA will explore scientific areas new to millimeter-wave research. Two examples are particularly illustrative: The combination of high sensitivity and sub-arcsecond angular resolution provided by the MMA at frequencies of 230 and 350 GHz will permit the photospheric emission from hundreds of nearby stars to be detected and imaged, the stellar radii to be determined and the positions established to astrometric precision. The same combination of instrumental parameters will provide images, at a resolution superior to that of the Hubble Space Telescope, of the redshifted dust emission from galaxies at the epoch of formation (z = 5–10). Indeed, these "protogalaxies" will be the dominant source of background confusion at 1 mm at levels approaching 1 mJy.

High sensitivity implies that the total collecting area of all the individual elements in the array should be made as large as possible, while fast imaging is achieved by distributing the collecting area over many elements. The precise requirements of how many elements and the size of the individual antennas are then decided by minimizing the total array cost. Sub-arcsecond imaging places a constraint on the array dimension: $0\farcs1$ at 230 GHz, for example, requires an array of maximum extent 3 km. Finally, sensitive imaging at high frequency demands that the MMA be located on a high-altitude site with excellent atmospheric transparency. Considerations such as these drive the design of the MMA; they are described in some detail in Volume II of the MMA Design Study. A summary of the MMA design parameters is given below:

Array —	
Number of Antennas:	30–40
Total Collecting Area:	1750 m^2
Angular Resolution:	$0\farcs07\lambda_{(\mathrm{mm})}$
Antennas —	
Diameter:	7.5–8.5 m
Precision:	$\lambda/40$ at 1 mm
Pointing:	1/20 beamwidth
Transportable	
Configurations —	
Compact:	$<$ 100 m
Intermediate:	300–1000 m
High Resolution:	3 km
Frequencies —	
Emphasis on:	200–350 GHz
Capability at:	70–115 GHz; 30–50 GHz
Desirable:	Simultaneous multi-band
Site —	
High-Altitude—suitable for precision imaging at 1 mm	

MMA Advisory Committee Meeting, February 1988

Abstract of the Issues Raised

R. L. Brown

Site Selection and Testing

- Investigations into the correlation between measured 225 GHz opacity and radiosonde data need to be pursued. If this correlation is sound, then a large number of potential MMA sites can be evaluated from historical data having a timebase of decades.
- High-altitude sites which provide the best high-frequency atmospheric transparency are favored. What constitutes an 'acceptable' site at, say, 460 GHz?
- Very high-altitude, Southern Hemisphere sites should be 'looked at' using extant radiosonde data.
- Total-power fluctuations of the sky brightness temperature need to be measured and compared with interferometer phases to investigate Smoot's suggestion that one can infer radio 'seeing' from the fluctuations.

Array Configuration

- What is the minimum number of array configurations that will permit us to realize most of our scientific objectives? Can we make this minimum number equal to one (i.e., fixed antennas)?
- Is there a role for fixed 'outrigger' antennas, allowing us to achieve the highest resolution, but without the u-v coverage that would be necessary for full imaging capability?
- We need to pay specific attention to the interaction between possible configurations and possible sites, owing to topographic limitations. This may also affect the number of array elements.

Antennas/Telescope Optics

- What is the antenna cost and how does that cost scale with diameter when proper account is taken of all the factors involving precision operation (especially pointing)?
- What are the cost/performance trade-offs associated with an unblocked aperture? Can the mirror supports (particularly the subreflector legs) be made sufficiently rigid? Can we sky-chop? Are the polarization characteristics of an unblocked aperture a limitation in the resulting images?
- What effect does a Q-band (9-mm) capability have on the telescope design and on the telescope optics?

- How do we obtain total-power measurements? How do we underilluminate? How do we obtain simultaneous observations in two bands that are separated by a factor of two or more in frequency?
- The Central Element question needs further investigation along the lines outlined in the Working Group Report.

MMA Frequencies

- The widest possible frequency coverage is a primary design emphasis. The 2-mm window should not be ignored; indeed at a high site the atmospheric transparency may be lower at 2 mm than at 3 mm, because O_2 remains a problem at 3 mm.
- The potential of the enormously broadband quasi-optical SIS mixers is such as to eliminate the need to identify observing 'bands'. Moreover, if successful, the number of MMA receivers is reduced markedly and the maintenance load eased commensurately. Is this technology realizable? What is likely in 5–10 years time? What are the performance trade-offs relative to narrow-band waveguide mixers?
- The receivers should incorporate DSB mixers, with the sidebands separable in the correlator.
- Simultaneous observations in two (perhaps) widely-spaced frequency bands are needed if we are to obtain a self-cal solution from a strong line in one band that we can use to correct the phases in the other band.
- What is the cost of extending array capabilities to higher frequencies (460 GHz)?

Correlator

- Flexibility is the central emphasis.
- Observers would like to split and analyze both sidebands in the correlator. Within the IF passband of each sideband one would like to place several spectral windows and to have independent control over the central (IF) frequency and the frequency resolution of each of these windows. The BIMA correlator is a model that should be scrutinized.
- How does the correlator cost/complexity combination scale with such flexibility? Does this cost then have implications on the number of array elements?

Mosaicing

- The need to mosaic places severe constraints on the performance of the array.
- How reliable must telescope pointing be in order to accommodate mosaicing? How do pointing errors degrade the mosaiced image?
- If the sidelobe levels of the primary response patterns of the individual antennas are high, what deleterious effects on the images will result?

- Does the need for high dynamic-range mosaiced images drive us necessarily to unblocked antenna apertures?

The Advisory Committee

- The Advisory Committee is taking on too many roles. It would be better to have one, or more, technical advisory committees that would focus on specific design questions and, in addition, broaden the Science Workshop so that the MMA can be evaluated as a scientific instrument. Meetings of the Science Workshop should include optical and infrared astronomers, theoreticians, and others who might benefit from the MMA, as observers or otherwise.
- Greater emphasis should be placed on making the community aware of the project and, of course, on listening to their ideas.

MMA Progress Reports

March, 1989

— CONTENTS —

MMA SITE AND CONFIGURATION

The scientific requirements for the MMA include provision for imaging at an angular resolution comparable to that of the Space Telescope, $0''.1$, at the principal frequency bands of the MMA. This requirement leads us to an array of 3 km maximum extent. Not all MMA observations need this angular resolution, and, hence, not always will the array be so extended. Indeed, the mosaiced images of extended molecular regions, in the disks of external galaxies and in Galactic molecular clouds, will usually be made with the MMA in a densely-packed configuration of diameter no greater than 100 m. Thus the MMA must be reconfigurable, the antennas transportable, and the site big enough, and flat enough, to permit such reconfiguration. The site must also be at high altitude, to realize the sky transparency needed for precision imaging at 200–350 GHz.

Over the past year we have completed a process of sifting through geographical data on potential MMA sites in the U.S. Southwest that satisfy the above criteria. From surveys of topographical maps we established a list, which we believe to be complete, of over fifty sites that are above 9000 feet altitude and South of latitude 36° N. With this starting list we ranked sites in groups depending on elevation, size, access, and environment (density of trees, distance from the nearest town, etc.). We currently have the following four sites in the 'top' group (listed here in no particular order):

- *Springerville, AZ:* This is a large (10 km), high (9200 feet) plateau in the Apache National Forest about 15 miles from Springerville. Access is by a paved state highway.
- *South Park, CO:* Located in a shallow valley 60 miles West of Colorado Springs at 9300 feet altitude, this site is mostly private land. The prime, flat area is approximately 3 km East–West by 6 km North–South; it is bisected by U.S. Highway 24.
- *Alpine, AZ:* As with the Springerville site, this is also located in the Apache National Forest, adjacent to U.S. Highway 666. We have little information on the site, but what we've seen is interesting: altitude 9900 feet, with a flat open area at least 3 km East–West by 5 km North–South.
- *Magdalena Mountains, NM:* This site is on a high (10,500 feet) mountaintop on the grounds of the Langmuir Laboratory Scientific Preserve in Cibola National Forest. The site is irregularly shaped but would accommodate a 3 km MMA configuration.

We have not yet sought permission to use any of these sites; discussions on this issue with the agencies and/or individuals involved have just begun.

We are also looking carefully at an even higher altitude site on Mauna Kea, HI, recognizing that there will be, in this case, a compromise to be reached between the potential gain of greater atmospheric transparency and the loss of

the longer MMA baselines, owing to the unfavorable topography of that mountain. Finally, an even more extreme possibility for the MMA site exists in the Southern Hemisphere, in Chile, where it appears possible to find a flat site at very high altitude. The logistics would be formidable and the incremental cost increase of the project could be substantial. We will follow with interest ESO's analysis of these same considerations as they assess southern sites for the VLT.

In Situ Measurements

For two years we have operated a 225 GHz tipping radiometer on the potential MMA site in the Magdalena Mountains. As part of its cycle of operation, this device not only does sky tips to measure the zenith 225 GHz opacity, but once every ten hours it takes an hour of rapidly sampled total-power measurements at the zenith. Tests at the VLA confirm the correlation between the power spectrum of these fluctuations and the variations seen in the VLA interferometer phase. Thus, the 225 GHz fluctuation spectrum can be used to provide an estimate of the interferometer phase expected on a remote site.

Ultimately, the plan is to test each of the most promising sites with one of the radiometers. For each site we hope to obtain a measure of the fraction of time that the atmospheric transparency at the site will permit operation at each of the four principal MMA frequency bands, and from the fluctuation spectrum, the fraction of time at each site during which coherence can be maintained with an interferometer over timescales of a few minutes.

Bob Martin at Steward Observatory has demonstrated the effectiveness of a related technique which permits one to estimate the atmospheric transparency at high frequency, given NOAA radiosonde observations and a good atmospheric model. The advantages of this technique are that one has access to a very extensive radiosonde database, compiled over two decades, which provides a measure of the long-term trends and that one can use to 'site test' in a computer. The disadvantage is that radiosondes are not launched from the site, itself, that one wishes to test. One must assume, *post facto*, that the atmosphere measured some distance from the radiosonde launch site is, in fact, representative also of the sky over the site of interest. In addition, the radiosondes provide no measure of the prospective interferometer coherence time.

We have accumulated the NOAA radiosonde measurements near each of the prospective MMA sites, including those near Mauna Kea (Hilo) and northern Chile (Antofagasta). A preliminary analysis of these data has been made and is available for distribution. The radiosonde data from Albuquerque, NM, are of particular interest to us because we can compare directly the sky transparency over the Magdalena mountains, as inferred from the radiosonde, with our direct *in situ* radiometric measurements. Exceptionally good mean seasonal agreement is seen for the two techniques, but there are noticeable differences in some of the data taken simultaneously. Presumably, this latter disagreement is attributable to local effects. A careful analysis of the radiosonde data for all the prospective MMA sites is in progress. It will be used in conjunction with the radiometer

measurements, not instead of the radiometer work, in helping us to characterize each of the potential MMA sites.

MMA ANTENNA AND TELESCOPE OPTICS

Since the 350 GHz atmospheric window is one of the two frequency bands given particular emphasis by the MMA (the other being the 1-mm band, 200–300 GHz) the antennas must permit precision operation at this frequency. The MMA Design Study Volume II had specified a surface accuracy of $\lambda/16$ at the shortest wavelength, 850 microns. However, this implies a 46% reduction in gain, relative to the antenna efficiency at lower frequencies—i.e., a somewhat marginal performance. Tightening the specifications to a 25 micron r.m.s. surface brings the gain reduction down to a little over 10%, which gives the array high performance in this atmospheric window. As a by-product, the array would also have usable performance up to the 500 GHz atmospheric window, but the primary objective in setting this surface specification is to achieve good performance at 350 GHz.

Another critical parameter is the telescope pointing accuracy. In order for mosaicing to be successful, the pointing of each element needs to be good to approximately 1/20 of a beamwidth, or about 1 arcsecond at 1 mm. This aspect is being studied in detail within the Central Element Study Group. The Design Study Volume II had specified 3 arcseconds, but this now appears to be inadequate.

The Working Group on MMA antennas reviewed the existing telescope designs of those few operational antennas that achieve performance similar to the revised MMA specifications. In doing so we discovered that if these tighter specifications are to be realized there are several implications. First, owing to thermal effects, an all-steel antenna design is unlikely to be able to meet the specifications. Incorporating carbon-fiber- reinforced plastic (CFRP) into the design may be appropriate—a route chosen for other successful high-precision antennas (IRAM, SMT). At present, there is little experience within the U.S. in the use of CFRP in antenna construction. The revised specifications would probably make the antennas more costly—and the dominant component in the overall construction cost of the project. During the lifetime of the array, receivers and computers are likely to be replaced as newer technology becomes available, but the antennas themselves are unlikely to be replaced. This suggests that the basic antenna design may provide the ultimate limitation in performance of the array: it would be unfortunate to compromise this at the start with too marginal a design. Experience with the SMT project suggests that the cost equations used to optimize the size and number of antennas may, in any case, need slight revision. However, we do not expect a very dramatic deviation from the current proposal for 30–40 antennas, each 7.5–8.5 m in diameter.

There are various options which could be chosen for the antenna design, e.g., an off-axis feed to reduce sidelobes, and a polar mount so that sidelobes do not rotate on the sky. However, as a starting point a trial antenna concept has been developed for a symmetric on-axis, Az–El antenna using a Coudé focus

cabin arrangement. The philosophy behind the design it to achieve a means for bringing the antenna beam to a convenient focus for coupling into the receiver. For reasons of convenience and reliability, the receivers should be mounted in fixed positions within the plinth of the antenna.

Figure 1 illustrates the proposed design. It is a conventional Cassegrain except that the secondary focus is brought right down into the plinth (Coudé focus). The receivers are located in fixed positions in the base of the antenna, a design feature which offers the following advantages:

(1) The receivers are easily accessible;
(2) Telescope balance is not affected by removing or installing receivers;
(3) Effects of cable and compressor line drag on the pointing are eliminated.

Primary Mirror

It is assumed that the diameter D of the primary reflector will be 8.0 m. The primary will be a symmetric paraboloid with a focal length f of 3.2 m, yielding a primary f-ratio (f/D) of 0.4. This value gives a compact antenna. It could be reduced further (to $f/D = 0.35$, say), but the tolerances in locating the secondary are more stringent (both for pointing and for beam quality). The primary mirror requires a central hole to pass the beam from the secondary to the receivers. Its size is determined by the field-of-view and the required secondary focal length. A diameter d_h of 1.0 m is close to optimum.

Secondary Mirror

The secondary mirror will be a symmetric hyperboloid. Its size is set principally by the field-of-view (approximately 10 arcseconds) and the requirement to have the secondary focus in the receiver room. It should be as small as possible, within these constraints, in order to minimize blockage and allow for the possibility of nutation for beam switching at some stage.

A diameter of 800 mm and a distance between foci of 9.79 m gives an effective f-ratio at the secondary focus of 12, which is convenient for most quasi-optical systems. With some more optimization the secondary diameter could be reduced but probably could not be made much less than 750 mm. Similarly, the focal ratio could be changed, but it is fairly tightly constrained and will probably be in the range 11 to 13.

Mirrors M1–M4

Four mirrors are used to remove the rotations produced by azimuth and elevation movements and produce a constant beam in the receiver cabin. M1 is fixed to the primary mirror structure, and M2–M4 to the rotating part of the azimuth bearing. As now envisaged, these are all plane mirrors, for the following reasons:

(1) They are cheaper to fabricate than shaped reflectors;
(2) They are simple to align;

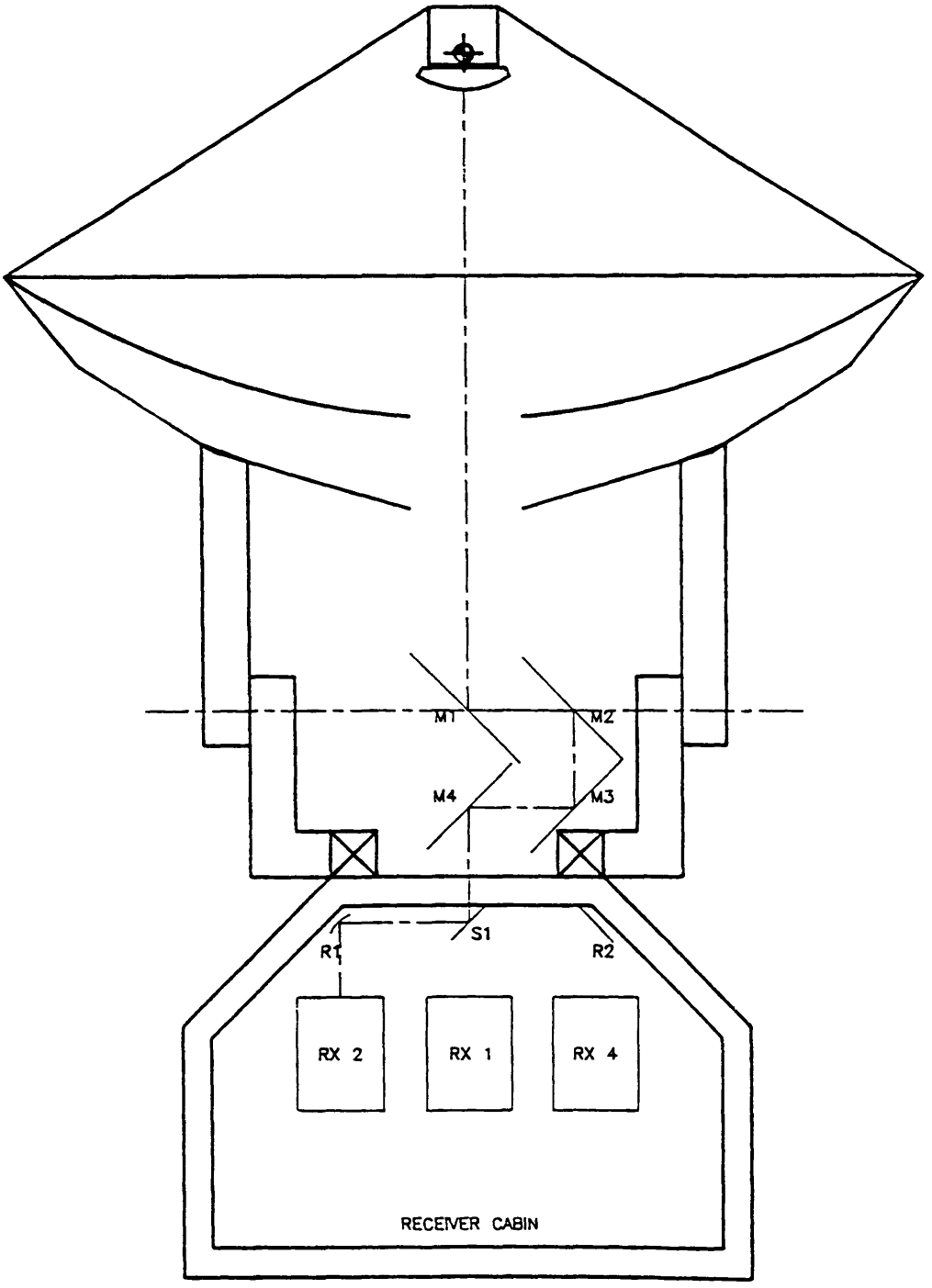


Figure 1. Proposed optics for the MMA antennas.

(3) The imaging properties of shaped mirrors for off-axis beams are unknown.

The sizes of these mirrors vary from about 1.4 x 0.95 m to 1.0 x 0.74 m.

Receiver Room

The receiver room is located in the base of the antenna mount. Five receiver bays are planned (Figures 1 and 2). One is in the center of the room and is intended for a focal plane array at frequencies down to about 200 GHz. It receives the signal directly from M4, through the azimuth bearing. Note that the image of the sky rotates relative to this receiver position, so that either the receiver has to be rotated or software correction has to be applied.

For the other receiver positions the beam is deflected by selection mirror S1 to one of the receiver focusing mirrors R1–R4. These receivers may be single beam at any frequency from 43–500 GHz, or possibly an array at the higher frequencies (300 GHz and higher).

The receiver front-ends should be about 0.75 x 0.75 x 1.0 m. Although this is smaller than the receivers on the 12-m telescope, some of the associated electronics can be removed to racks at the side of the receiver room.

Discussion

Table 1 summarizes the main parameters of the design. The specification that the secondary focus be brought to the receiver imposes quite stringent conditions on the geometry and dimensions. The values given here therefore cannot be changed by very much. As far as the telescope performance is concerned, the main consequence is the relatively large central blockage, leading to increased sidelobes and reduced aperture efficiency. The reduction in gain due to the 1-m diameter hole in the primary is about 5.4%. In order to reduce the size of the central blockage significantly, the secondary focus would need to be brought much closer to the vertex, and M1 and M4 (for example) would need to be focusing reflectors. Because of the disadvantages listed earlier, this should be considered only if the blockage is shown to be unacceptable.

Figure 2 gives the main dimensions set by the choice of the optics. Figures 3–5 show how the beams of different receivers couple to the telescope. A Gaussian beam giving an edge taper of 11 dB is assumed, and the 1- and 2-beam radii contours are plotted. The 2- beam radii contours are a fairly conservative estimate of the clearance required, but this should not be reduced, if at all possible.

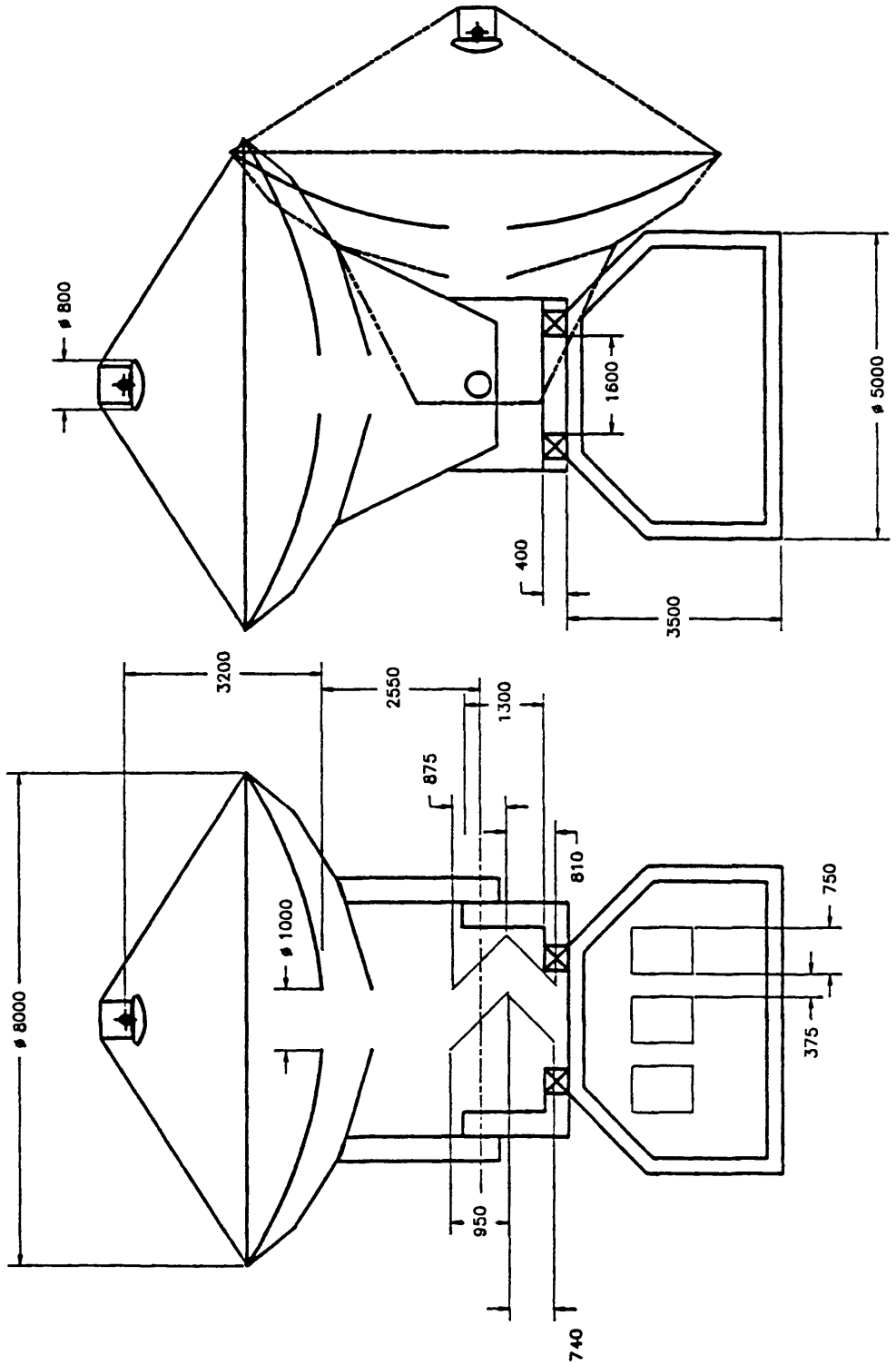


Figure 2. Preliminary choices for dimensions of antenna (in units of millimeters).

Table 1. Principal design parameters.

Primary Mirror —		
Shape		Paraboloidal
Diameter	D	8.00 m
Focal length	f	3.20 m
Focal ratio	f/D	0.4
Edge angle	θ_p	64°01
Central hole diameter	d_h	1.00 m
Secondary Mirror —		
Shape		Hyperboloidal
Diameter	d_s	0.80 m
Distance f_s between foci	$f_s = 2c$	9.7908 m
Additional path length	$2a$	9.18409 m
Eccentricity	e	1.06897
Magnification	M	30
Paraxial focal length	$f_a = a(e^2 - 1)/2$	0.655 m
Equivalent Paraboloid —		
Focal length	$F = Mf$	96.0 m
Focal ratio	F/D	12.0
Edge angle	$\theta_s = 2\cot^{-1} 4F$	2°39
Plate scale	$1/F$	2.15 arcsec mm^{-1}

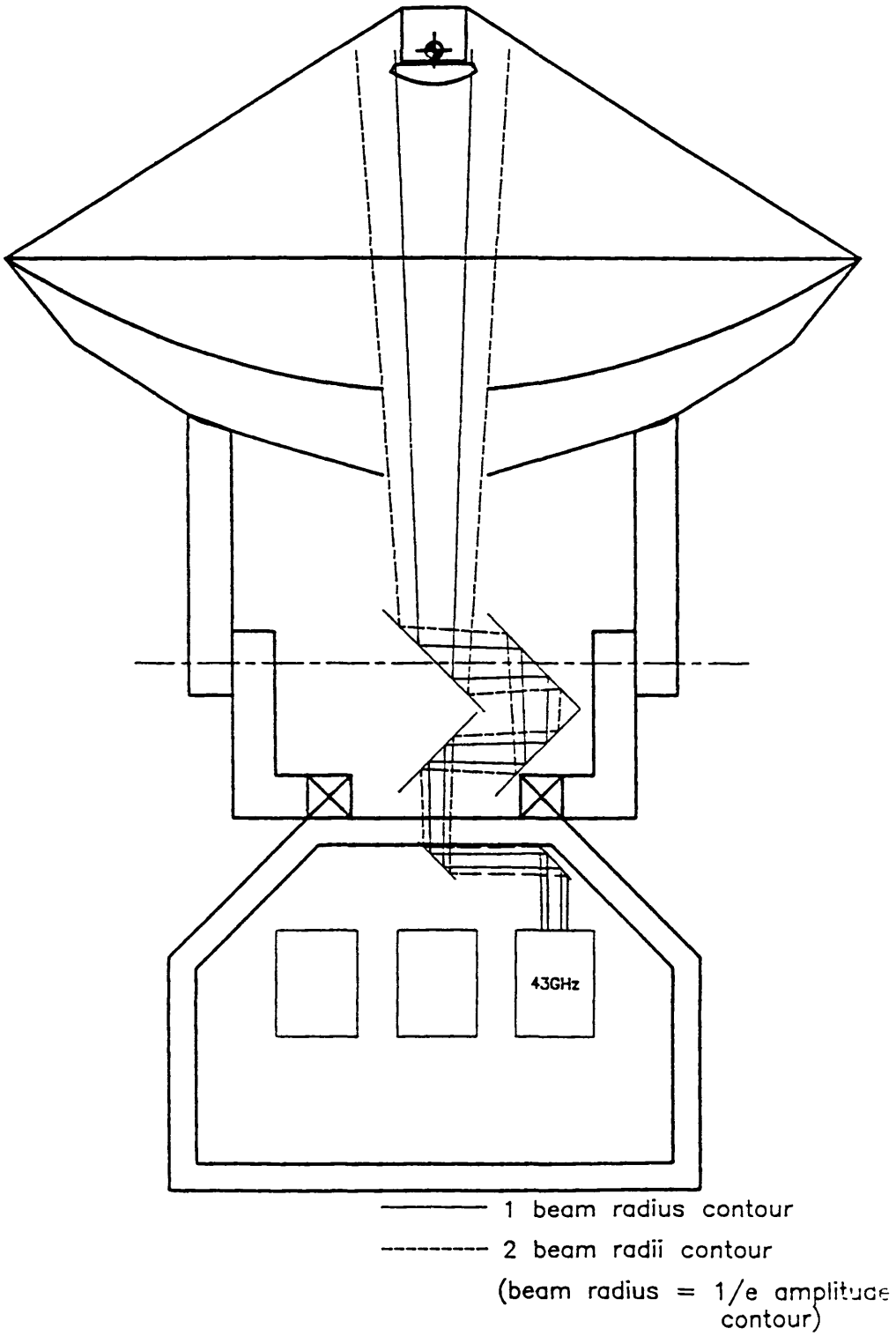


Figure 3. 43 GHz beam path.

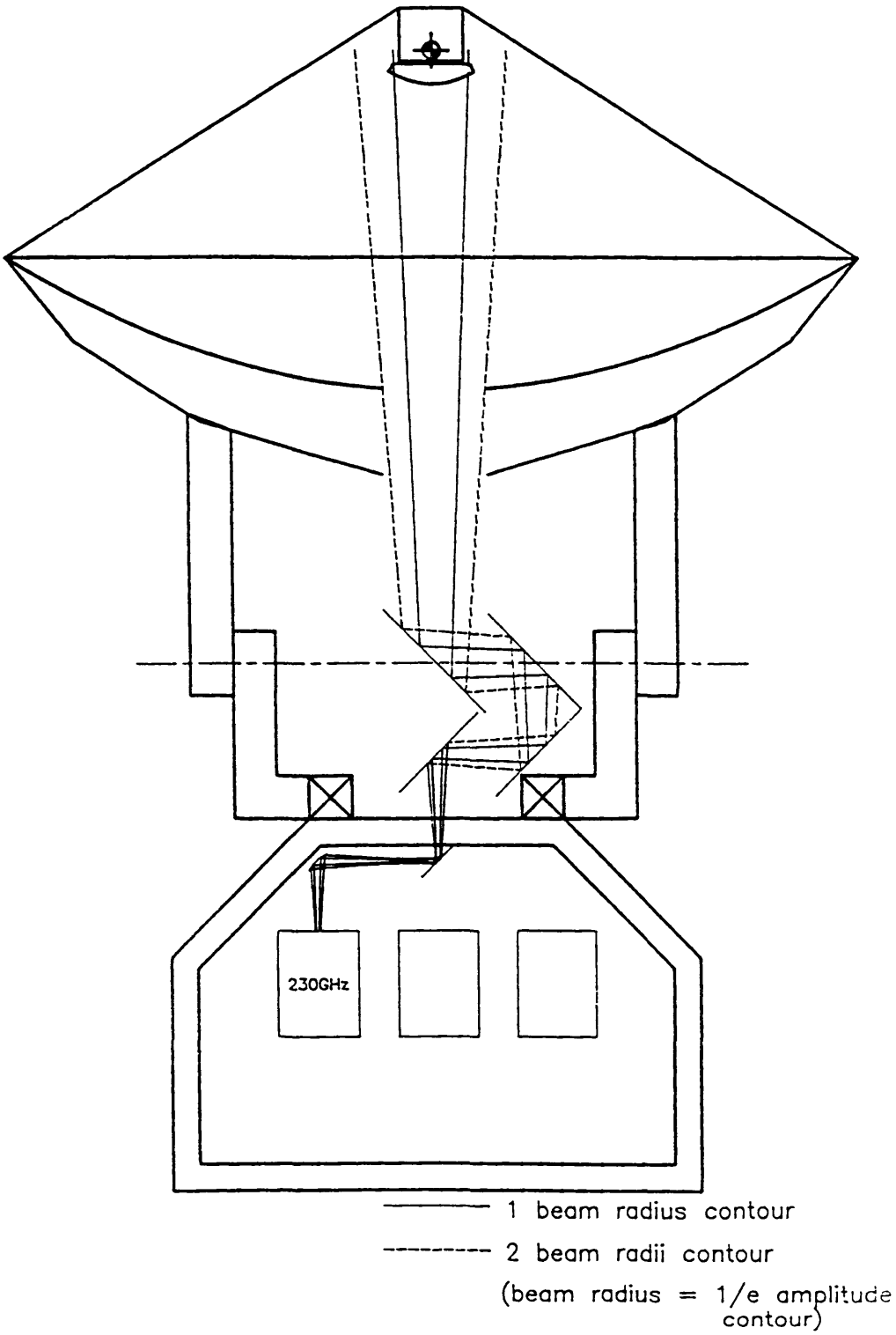


Figure 4. 230 GHz beam path.

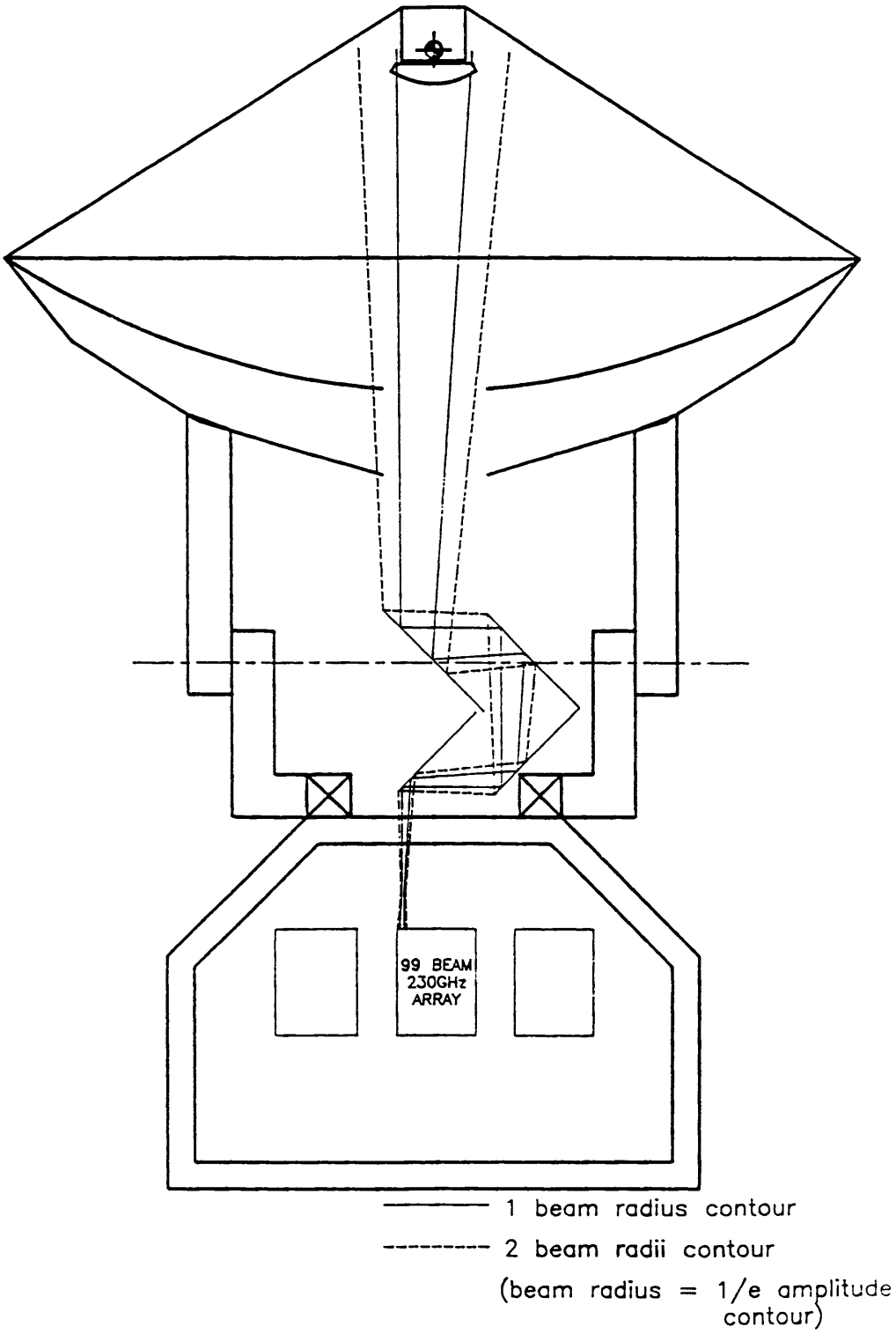


Figure 5. 230 GHz offset beam.

MOSAICING AND THE CENTRAL ELEMENT

One crucial requirement of the MMA is that it should be able to measure essentially all Fourier components relevant to a region of sky, up to the cutoff in maximum spacing of a few kilometers. Particular emphasis must be given the measurement of short-spacing information, because it is crucial for imaging large objects. By 'short', we mean spacings less than or close to the antenna diameter of 7.5 m, say 0–10 meters. In conventional interferometers, these and shorter spacings are often missing completely. However, since the MMA will regularly image objects much larger than an individual primary beam, these spacings are vital. There are three possible schemes for obtaining the measurements:

(1) *Large Single Dish:* This is the most conventional of the schemes. A large single-dish is used to image the sky in total-power mode. The required spacings could be obtained by inverse Fourier transformation of the image followed by correction of the sensitivity function of the dish (i.e., the inverse Fourier transform of the primary beam). According to the conventional wisdom, the size of the single-dish should be 2–3 times *larger* than the required spacings.

(2) *Multi-Telescope Array:* An array of smaller telescopes is used as an interferometric array, and a subset of the required spacings is obtained directly. The size of the elements should be 2–3 times *smaller* than the size of the large array elements. To avoid shadowing between the densely packed elements, the elements should be mounted on a backing structure which can point at and track the object. Structures on the largest scale-size are derived from total-power measurements using the larger interferometer dishes.

(3) *Homogeneous Array:* This is the most radical of the schemes and the one of greatest potential. There would not be a central element: each antenna in the interferometer array would be equipped to observe in a total-power mode, preferably simultaneously with the interferometer measurements. As the array is scanned for mosaicing, the total-power observations can be used to form an image which could be inverse-transformed to obtain spacings up to some fraction of the dish diameter. In principle, this can be generalized for the interferometer elements, and it provides information on scales ranging down from the shortest spacing by some fraction of the dish diameter. Therefore, if the shortest spacing is about a dish diameter, and if the extrapolated measurements are good at offsets of up to half a dish diameter, then there is no gap to be filled.

Some variant of the mosaicing technique would be used in each of the three cases. Although all three options are theoretically feasible, there may be practical difficulties with each. Our current task is to investigate these obstacles and

reassess the viability of each option. The main questions can be summarized as follows:

(1) Given perfect observations, will mosaicing allow recovery of all spacings:
   (a) down to about half the dish diameter, from interferometric measurements?
   (b) up to about half the dish diameter, from the total-power measurements?

(2) Will the various instrumental errors corrupt the mosaicing? For example, will pointing errors be a limiting factor? Or reversing the question, what are the necessary specifications on pointing, illumination stability, cross-talk, etc.?

(3) How do observational details affect the design? For example, can we design a homogeneous array which allows sufficient non-shadowed integration time? Beam-switching will be required for total-power measurements—how far a throw is necessary? Can the optics accommodate it?

(4) Are the total-power measurements significantly more difficult than the interferometric measurements? Is a single-feed array with N elements easier to maintain than the feeds on N separate telescopes?

Answers to these questions are difficult to obtain. Mosaicing is a particularly thorny procedure to analyze theoretically since in part it relies on a nonlinear algorithm, MEM, for a crucial step. Consequently, we do not believe that a full theoretical analysis of mosaicing is possible. As an alternative approach, we suggest that many of the issues concerning the central element are best addressed by computer simulations. We also believe that simulations are necessary to settle more general questions about the Millimeter Array design. Since the MMA will be the first radio interferometric array designed for mosaicing, we believe that a demonstration of the array via computer simulation is vital. For these reasons we have developed a simulation package to test mosaicing of complex objects in realistic conditions. The package produces pseudo-data that will allow us to accurately simulate mosaicing observations with the MMA in the presence of atmospheric disturbances, pointing errors, and aperture-illumination errors. It will also be useful for investigating some more general questions about the MMA design; for example: are equatorial mounts for the antennas advantageous? what is the required upper limit on pointing errors? It is planned that the package will be used over the entire design phase of the MMA, although some simple results are already available. These include studies of the effect of incomplete knowledge of the primary beam on the reconstruction of a mosaiced image and of the degradation of such an image by pointing errors in the individual pointings.

Specific results from the simulation program are in hand for the following:

- A demonstration of the imaging properties of the compact config-

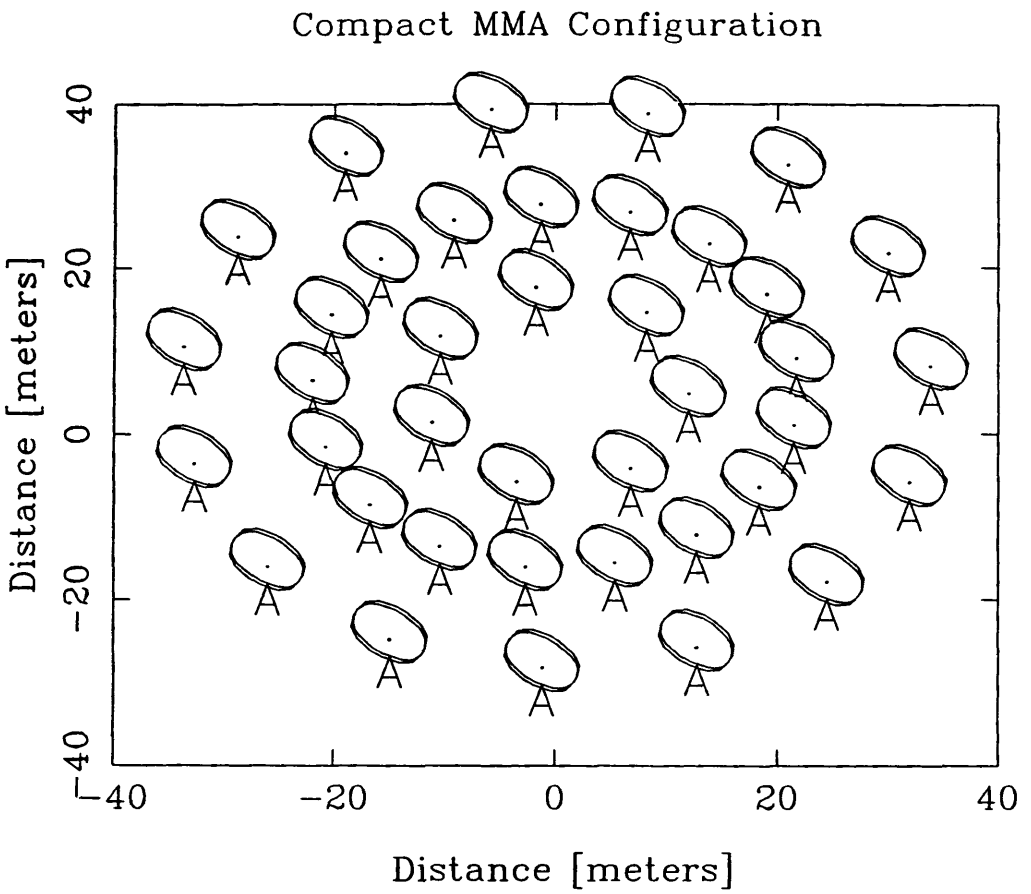


Figure 1. A possible compact configuration of array elements for the MMA.

uration for perfect data. The instantaneous u-v coverage of the compact array shown in Figure 1 is good enough that an excellent 100-pointing mosaic can be obtained in a few hours. A comparison of the MMA simulation with that obtained from the simulation of a representative 9-element array shows the imaging of the MMA to be a significant improvement.

- Investigation of atmospheric models appropriate for the simulations. This covers the relationship between fluctuations in opacity, brightness temperature, and excess path length, as well as the spectrum of the turbulence. We believe that we have a formalism suitable for generating visibility data with reasonably authentic errors.
- Investigation of the effects of truncation of our knowledge of the primary beams of the array elements. The news here is very encouraging: reasonable-quality images can be obtained even when the primary beam is known only down to the 5% contour. This has been investigated only for well-isolated objects; the next step is to

see if it still holds for a mosaic of part of a crowded field. This robustness can be attributed to the redundancy of information in the image plane sampling, and to the deconvolution algorithm used. This does bode well for the future simulations concerning pointing errors and may indicate that we have been overly concerned about the effect of errors in mosaicing.

- A large part of the required software is now in hand, and we are proceeding with further tests, most immediately concerning pointing errors.

The simulations lead us to the tentative conclusion that a large central element seems not to be required: complete short-spacing information can be obtained from a homogeneous array. However, this conclusion is subject to revision as we get even more experience with the simulations.

In addition to the simulations, observational tests are being carried out. The Hat Creek BIMA millimeter array is being used for various tests of mosaicing which should shed some light on a number of topics, including the importance of pointing errors. Tim Cornwell and Robert Braun are planning to test spectral line mosaicing using H I observations of M33 made with the VLA. Here the goal is to make use of the fact that spectral line observations provide the equivalent of the 'zero-spacing' flux from the measured autocorrelations; in principle these can be incorporated into the imaging. Now we wish to investigate the extent to which this works in practice.

The computational resource requirement of mosaicing has already been analyzed in substantial detail and is presented in Chapter V of MMA Design Study Vol. II. There, a rough cost estimate—$4M for all required calibration and imaging hardware (mosaicing requirements are predominant among total needs)—was given; this is still our best estimate. The overall needs of the MMA in off-line computing are essentially the same as those of the VLA; what differs for the MMA is simply the mix (i.e., the way these needs are proportioned) and the total. In addition, the budget for on-line computing hardware was estimated in Vol. II as $2M; on-line needs have still not been considered in detail, but our experience with VLA and VLBI real-time operation makes this analysis relatively straightforward.

RECEIVERS

During the time period since the *Millimeter Array Design Concept* was written, there have been several developments in millimeter wave receivers which may influence our approach to the MMA receivers.

1. Improvement of SIS receiver sensitivity

Through the use of integrated tuning circuits, we have been able to reduce the mixer noise temperature to within a factor of two of the photon noise $\hbar\omega/k_B$ in SIS mixers for the 3-mm band [1]. At 115 GHz, with $T_{\rm mixer} \leq 5.6$ K DSB, the receiver noise temperature referred to the mixer input flange in our laboratory test receiver is $\leq$ 9.5 K DSB. Allowing for window losses in a well-designed telescope receiver, it should be possible to achieve $T_{\rm rcvr} < 20$ K DSB, measured at the dewar input.

2. An SIS receiver with no adjustable tuners

An experimental fully-integrated 3-mm SIS mixer has given promising results over 70–115 GHz (a 49% bandwidth) without any tuning adjustments [2]. The receiver noise temperatures plotted in Figure 1 are referred to the input flange of the mixer. We have not yet investigated the saturation (dynamic range) characteristics or sideband ratio of this mixer.

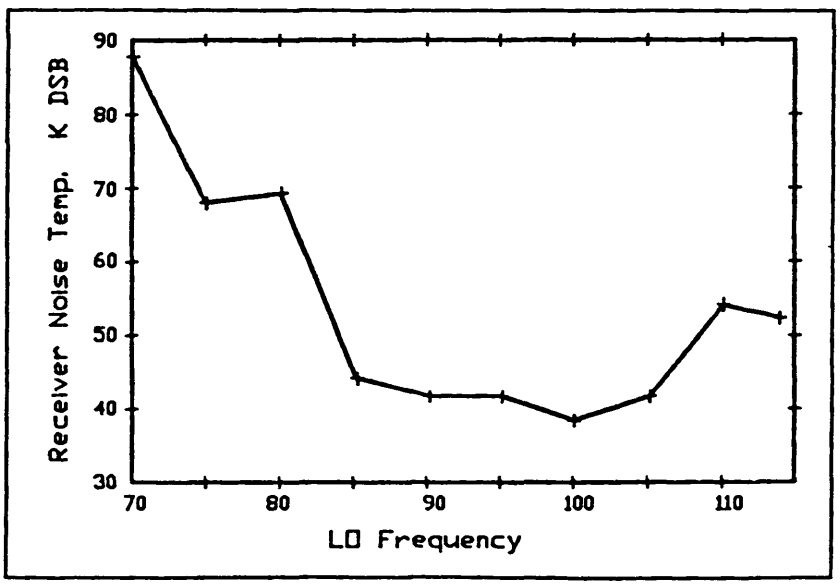


Figure 1. Noise temperature, referred to the mixer input flange, of an SIS receiver using an experimental fully-integrated mixer with no adjustable tuners [2].

3. Planar log-spiral mixer

Caltech results for an SIS junction mounted at the vertex of a planar log-spiral antenna on a quartz substrate [3] have excited considerable interest in the radio astronomy community. Reported receiver noise temperatures are plotted in Figure 2. At the lower frequencies the receiver was driven into saturation by thermal radiation from a room-temperature black-body; this is not surprising as the thermal radiation in a 500 GHz bandwidth at 300 K is about 2 nW, which is of the same order as the LO power required by a single-junction SIS mixer at 100 GHz. Series arrays of junctions and integrated tuning circuits may eventually enable this type of mixer to be used to advantage on the MMA.

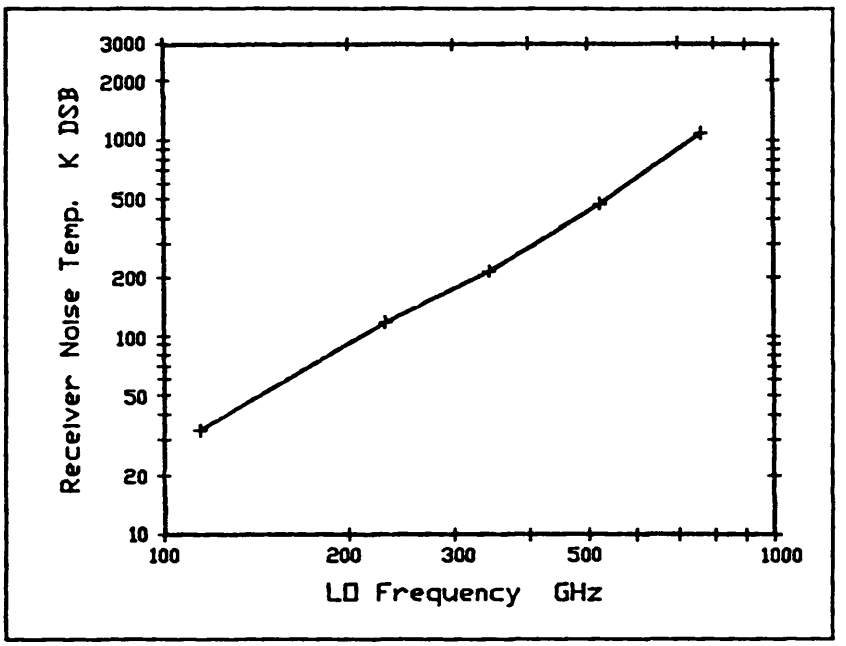


Figure 2. Noise temperature of an SIS receiver using a planar log-spiral mixer [3].

4. Improvement in SIS junction quality

Work at UVa under NRAO contract has made good progress. They have recently made Nb/Al-Al_2O_3/Nb SIS junctions with the highest quality parameter ($V_{\mathrm{mixer}} = 1400$ mV) ever reported, and NbN/Ox/PbBi edge junctions with the highest V_{mixer} and gap voltage ever reported for edge junctions. These two SIS technologies may well end up as the prime choices for low-noise millimeter and submillimeter heterodyne receivers. At present the UVa group is in a precarious financial state, largely as a result of their loss of support from NASA and JPL. Unless substantial additional support is found soon, they will be forced to terminate their work on SIS devices. Such an occurrence would represent a serious loss to this country of unique technology and expertise.

5. Local oscillators

The local oscillators for the MMA will almost certainly be varactor-diode frequency multipliers driven by phase-locked Gunn oscillators. Commercial development of InP Gunn diodes is continuing, but development of very widely tunable (25–40% for the MMA) oscillators using those diodes will almost certainly need to be undertaken as part of the MMA project. Possibly, two LO's will be needed per MMA frequency band.

Reliable, widely tunable frequency multipliers will need to be developed. Very high-quality planar (i.e., whiskerless) Schottky mixer diodes have recently been made at UVa [4]. With suitable modification these should be well-suited to multiplier operation. Unfortunately, budgetary considerations have forced NRAO to discontinue its support for all Schottky diode work at UVa.

Recent work at Caltech on planar grids of varactor diodes has demonstrated high efficiency multiplication at millimeter wavelengths [5]. An efficiency of 9.5% has been achieved experimentally in doubling to 66 GHz, and it is anticipated that 30% efficiency should eventually be possible in doubling to $\sim$ 200 GHz. These designs are for input powers of many watts, and it is not clear whether they could be suitably scaled for operation at the milliwatt level available from Gunn oscillators.

References

[1] S.-K. Pan, A. R. Kerr, M. J. Feldman, A. Kleinsasser, J. Stasiak, R. L. Sandstrom, and W. J. Gallagher, "An 85-116 GHz SIS receiver using inductively shunted edge-junctions", *IEEE Trans. Microwave Theory Tech.*, Vol. **MTT-37**, No. 3, Mar. 1989.

[2] A. R. Kerr and S.-K. Pan, NRAO. Private communication.

[3] T. H. Büttgenbach, R. E. Miller, M. J. Wengler, D. M. Watson, and T. G. Phillips, "A broadband low noise SIS receiver for submillimeter astronomy", *IEEE International Microwave Symposium, Digest of Technical Papers*, June 1988, pp. 469–472.

[4] D. G. Garfield, R. J. Mattauch, and W. L. Bishop, "Design, fabrication, and testing of a novel planar Schottky barrier diode for millimeter and submillimeter wavelengths", *Proc. IEEE Southeastcon '88*, April 1988, pp. 154–160.

[5] C. F. Jou, W. W. Lam, H. Z. Chen, K. S. Stolt, N. C. Luhmann, and D. B. Rutledge, "Millimeter-wave diode-grid frequency doubler", *IEEE Trans. Microwave Theory Tech.*, Vol. **MTT-36**, No. 11, Nov. 1988, pp. 1507–1514.

CORRELATOR

The correlator design for the Millimeter Array is still in a rudimentary state. It consists of the concept given in Design Study Volume II [1], along with some refinements to the cost equation and its parameters given in a memo by D'Addario [2]. In February 1988, the Advisory Committee recommended [3] that "flexibility" be a major design goal of the correlator, with particular emphasis on being able to separate the sidebands of a DSB front end and on being able to observe more than one line at a time. Here we attempt to address these and other flexibility issues.

1. Assumptions

As a starting point, consider a correlator that accepts 2 GHz of signal bandwidth from each of 40 antennas and computes the complex cross-power spectrum on each baseline with a frequency resolution of 2 MHz. This is the machine whose cost we attempted to estimate in the earlier documents; let it be called the "nominal" correlator. The cost is still very uncertain, being a strong function of rapidly changing technology.

Little has been said about the organization of the signals or about the various possible modes of operation. Here we assume that:

(a) The 2 GHz bandwidth is supplied in $J \geq 4$ separate channels, and these can be processed independently by the correlator. The channels may be from different front ends, from different parts of a wide IF in one front end, from different polarizations, or some combination of these.

(b) Fringe stopping has been done in the LOs. This means that the fringes have been stopped at the total LO frequency for each channel or at the center frequency of the channel.

(c) The front ends use double-sideband mixers, but accurate phase switching of the first LO in increments of $\pi/2$ is possible, and this can be synchronized with correlator integrations.

2. Options

Separation of Sidebands. Given the assumption that accurate phase-switching of the first LO is available, visibilities in each sideband can be measured separately. If the switching interval is not too short, then this requires doubling the size of only the long-term accumulator in the correlator. The long-term accumulator is usually less than 10% of the cost of the cross-correlation part of the correlator. If the short-term dump time of the correlator is 100 msec, then for 40 antennas the total switching cycle time would be 6.4 sec (assuming Walsh function switching patterns), which becomes the quantum of integrating time for sideband-separated observations.

The sideband separation does not in itself impose any penalty in signal-to-noise ratio (SNR); that is, a line appearing in only one sideband will be detected with the same SNR whether or not the sideband separation procedure is applied. But the noise of both sidebands is present in each result, so the SNR is worse than if the front ends had been single-sideband, typically by a factor of $\sqrt{2}$.

Polarization. For polarization measurements, the nominal correlator can be re-organized to produce the necessary cross-polarized correlations provided that only half the maximum bandwidth is processed. That is, if half of the input channels are unused then the corresponding sections of the correlator can be devoted to the additional correlations. This is possible in either the FX or XF architecture. The cost lies in providing the appropriate interconnections between correlator sections; in accord with our experience in designing the VLA and VLBA correlators, the incremental cost can be made negligible by careful design. On the other hand, if full polarization measurements must be made at the full bandwidth, then the cost of the *cross-correlation part* of the correlator must be doubled. With present technology, it appears that the cross-correlation part dominates the cost in any architecture, in which case the total cost also doubles.

In the XF case, it is also possible to obtain polarization measurements by degrading the spectral resolution while continuing to process the full bandwidth. This does not work with an FX correlator.

Multiple Observing Bands or Lines. This has a bigger effect on the design of the front ends and IF processing than on the correlator. The question is to what extent the correlator input channels are separately tunable. We have assumed that a minimum of 4 channels is available, and for various reasons 8 or 16 channels may be more appropriate. It is straightforward to arrange that each channel may be tuned to any frequency within the bandwidth of the first IF; in addition, any available front end can be connected to any channel.

It may be desirable to operate different channels at different resolution or bandwidth. This will be possible if the correlator is organized as J quasi-independent correlators, each handling one channel of all N antennas. Such an organization also facilitates a future increase in J by adding identical hardware, but makes difficult a future increase in N.

High Resolution. The maximum bandwidth of each channel is B/J, where B is the total bandwidth (assumed to be 2 GHz) and J is the number of channels. If the channel bandwidth is reduced by filtering, then the resolution can be decreased by recirculation. This applies to either the FX or XF architecture, but in the FX case the additional memory must be built into every stage of the FFT engine and the number of stages must be increased by $\log_R \alpha$ where α is the bandwidth reduction factor and R is the FFT radix. In any case, the resolution becomes b/α^2. The same is true if the total bandwidth is reduced by not using some channels, keeping the bandwidth of each channel the same, provided that the cross-correlation hardware of the unused channels can be re-allocated.

The FX correlator is at a disadvantage because it must be built for a partic-

ular maximum FFT length. The VLBA correlator has a degree of freedom not available here, namely variation of the tape speed to keep the correlator input bandwidth constant while varying the observing bandwidth.

Trading Baselines for Bandwidth or Resolution. If each antenna actually provides more channels than the correlator can process, then by not using some antennas and connecting the extra channels from the remaining antennas to the now-free correlator inputs, larger bandwidth can be accepted. This would be useful only for special purposes, such as observing more lines simultaneously when SNR and u-v coverage on any one line is not a limitation.

Another such trade is conceivable. While retaining all antennas at the full bandwidth, some baselines might be of little interest because of redundant u-v coverage. In principle, the cross-correlation hardware for these baselines might be re-allocated to improve the spectral resolution on the other baselines. However, this increases the complexity of correlator organization to such an extent that it should not be implemented unless a strong case is made for its importance.

References

[1] *The Millimeter Array Design Concept*, Millimeter Array Design Study Volume II, Section VI.4, NRAO, Jan. 1988.

[2] L. R. D'Addario, "Millimeter Array Correlator Cost Equation", February 1988.

[3] R. L. Brown, "Millimeter Array Advisory Committee Suggestions", April 13, 1988.

The Millimeter Array Design Concept

EDITED BY R. L. BROWN AND F. R. SCHWAB

National Radio Astronomy Observatory

January 1988

MMA Design Study, Volume II

The National Radio Astronomy Observatory (NRAO) is operated by Associated Universities, Inc., under contract with the National Science Foundation.

CONTENTS

I. Introduction

Millimeter-wave astronomy has had an important impact on virtually all areas of contemporary astronomy. It provides a testing ground for our theories of stellar evolution, galactic evolution, and the evolution of the universe itself. Recognizing the promise and potential of millimeter-wave astronomy, many nations and international consortia have built, and are building, modern telescopes and support facilities for millimeter-wave research (e.g., West Germany, France, Holland, England, Sweden, Iraq, Canada, and Japan). In the U.S., no new national instruments have been constructed since the NRAO 36-foot telescope was completed in 1967.

In 1983, the NSF formed a subcommittee, chaired by Professor Alan Barrett, to offer guidance on the priorities for the future development of millimeter and submillimeter astronomy. Among the committee's three principal recommendations[1] was a request for a design study for a national millimeter-wavelength synthesis array. This particular recommendation was an encouragement that the NSF build on the seminal interferometric millimeter work being done with the Owens Valley Radio Observatory's 3-element interferometer and the Berkeley Radio Astronomy Laboratory's Hat Creek interferometer. These instruments continue to demonstrate the wealth of scientific insight that can be gained from observations at high angular resolution. The Barrett Committee saw the opportunity, and pointed to the need, for an imaging synthesis array for millimeter-wavelength astronomy. The array, they noted, should have the following properties:

(1) 1 arcsecond resolution at 115 GHz,
(2) 1000–2000 m^2 total collecting area,
(3) good imaging capability at 1 mm.

Design of an array intended to fulfill these needs was begun at the NRAO in 1984. The resultant conceptual design was presented to a scientific workshop of approximately fifty astronomers held in Green Bank in September 1985. The workshop participants were asked to assess the scientific justification for such an array and then refine the array design in light of their scientific demands. Volume I of this report summarizes the scientific discussion. This volume presents the refined design concept.

The strong scientific endorsement given for the millimeter-wave array (MMA) at this workshop placed particular emphasis on the array's imaging capability. The MMA should be a true imaging array, it was concluded. Frequently the objects to be observed are many times larger than the telescope beamwidth. This means that many overlapping regions must be observed and the individual images merged into a single mosaic image. Such mosaics can

[1] A. H. Barrett, C. J. Lada, P. E. Palmer, L. E. Snyder, and W. J. Welch (1983), *Report of the Subcommittee on Millimeter- and Submillimeter-Wavelength Astronomy*, National Science Foundation Astronomy Advisory Committee.

be constructed only if the array is fast and has good "snapshot" imaging capability (i.e., good instantaneous u-v coverage). Many antennas are required, to maximize this capability. In addition, the working groups noted that new science, not possible with any existing instruments, would follow if the MMA provided a sensitive high frequency capability (to as high as 350 GHz) and sub-arcsecond resolution. The combination would provide a unique opportunity for astronomers to study thermal emission from the photospheres of nearby stars, to image the dust continuum emission in galaxies throughout the local universe, unconfused by Galactic "cirrus," and to detect the redshifted dust IR-luminosity from primeval galaxies.

The realization of these scientific opportunities requires a refinement of the design parameters suggested by the Barrett Committee for the millimeter-wave array. Specifically:

(1) Sub-arcsecond imaging at 115 GHz (and higher) implies array baselines ≈ 3 km.
(2) Rapid imaging means that the 1000–2000 m^2 of collecting area be divided among many elements. We consider 40 antennas of 7.5 m diameter (total collecting area 1767 m^2; 780 simultaneous baselines).
(3) Sensitive imaging at high frequency (to 350 GHz) requires that the array be located on a high site with low atmospheric opacity.

In the sections that follow we describe a conceptual design of an array that is tailored to these requirements.

II. An Imaging Millimeter-Wave Array

1. Overview of the Scientific Considerations

At the Green Bank scientific workshop the eight working groups, arranged by scientific discipline, met separately and considered the millimeter-wave array design in the context of their specific scientific needs and goals. A comprehensive review of these discussions is included in Volume I of the MMA concept document. As one might expect, the fact that observations at millimeter wavelengths are of benefit to so many areas of astronomy means that the demands on a single observational facility are disparate, and in a few cases conflicting. Nevertheless, the corpus of scientific considerations presents several recurrent themes that form an outline for the design emphasis of the millimeter-wave array.

Spectroscopic Observations: One uses spectroscopy

(1) to establish the character and distribution of material in an astronomical object;
(2) to investigate kinematics; and
(3) to establish relationships between one species and another that may provide conceptual or theoretical insight (e.g., in astrochemistry, the extent of isotopic enrichment through stellar processing).

In order to discuss the content and distribution of molecular material, we need to work with an image of a strong line—one often chooses CO or its isotopomers—in an object which is spatially extended, often more extended than the telescope beam. Mosaic images are needed. For heterogeneous or structured objects, we need (mosaic) images of more than one transition and isotopomer of CO or of another molecular species in order to disentangle molecular abundance from excitation and opacity effects. This is true for all observations that have as their goal determination of molecular abundances. The need this imposes on the design of the MMA is for multi-band imaging that can be accomplished rapidly in many overlapping fields, to allow construction of a mosaic image. An abstract of the spectroscopic design emphasis as outlined by the Green Bank working groups is given in Table 1.

Continuum Observations: Each of the scientific working groups emphasized the need for the MMA to provide good images of the thermal continuum emission from heated dust. At millimeter wavelengths one has the opportunity to measure the dust content and spatial distribution unencumbered by optical depth corrections or contamination from Galactic infrared "cirrus." Thus the millimeter observations complement the IR surveys, such as that from IRAS (NASA's Infrared Astronomy Satellite, active for nine months during 1984), and they extend this work to very much higher angular resolution. Wide-field mosaic images of dust emission at high frequencies, in bands centered near 250 and 350 GHz, are needed.

Table 1. Spectroscopic Observations

Subject Area	Field of View	Angular Resolution	Frequencies	Species	Instrumental Emphasis
High z	$\leq 10''$	$< 1''$	35–360	Redshifted lines	Frequency coverage
Low z	$\lesssim 10'$	$< 1''$	115, 230, 345	CO, HCO^+, other mols.	Good CO imaging
Mol. clouds	$\lesssim 10'$	$< 1''$	3, 1.3, 0.87 mm bands	CO + isotopes, many mols.	Good images, frequency coverage
Sun and stars	N/A	N/A	N/A	N/A	N/A
Circumstellar shells	$< 6'$	$0''.1$	3, 1.3, 0.87 mm bands	CO, CS, HCN, SiO	Good images, frequency coverage
Solar system	$< 30''$	$< 1''$	115, 230, 345	CO + isotopes	Fast imaging
Astrochemistry	$\lesssim 10'$	$1''$	35–360	Many	Frequency coverage

The emphasis on continuum sensitivity at high frequencies, made possible by wide bandwidths, also facilitates observations of the thermal emission from solid surfaces (such as asteroids) and stellar photospheres, which have not been possible heretofore.

The criteria which affect the MMA design and are most critical to continuum observations are summarized in Table 2.

The capabilities of the MMA given greatest emphasis by the Green Bank scientific workshop can be assembled within the following three categories which are amplified below:

(1) Provision for rapid imaging,
(2) Sensitive imaging at 230 and 345 GHz,
(3) Wideband/multi-band facility.

2. Rapid Imaging Capability

An imaging millimeter array (MMA) synthesis telescope will be of especial value for observations which must be completed quickly. Here "quickly" can mean within a few seconds, or perhaps a few minutes, or it may mean within a day or two. But, in any case, many vital projects must necessarily be finished

Table 2. Continuum Observations

Subject Area	Field of View	Angular Resolution	Frequencies	Physical Process
High z	$\geq 10'$	$< 1''$	35, 90 (250 & 350?)	3 K fluctuations
	$1'$–$10'$	$< 1''$	35, 90, 250, 350	Synchrotron jets
	$5''$–$10''$	$< 1''$	90, 250, 350	Thermal dust
Low z	$\leq 10'$	$< 1''$	250, 350	Thermal dust
Mol. clouds	$\lesssim 10'$	$< 1''$	250, 350	Thermal dust
Circumstellar shell	$< 6'$	$0''.01$–$1''$	250, 350	Thermal dust
Sun and stars	$< 30'$	$\leq 1''$	35, 90, 250, 350	Nonthermal flares, Thermal BB
Solar system	$30''$	$\leq 1''$	35, 90, 250, 350	Thermal BB
Astrochemistry	N/A	N/A	N/A	N/A

in a time short compared with the time needed to physically move antennas. For this class of research it is not feasible to improve the u-v coverage by repositioning antennas—the phenomenon to be studied, the requisite atmospheric transparency, or the scale of the investigation requires short integrations.

1. Transient phenomena.

Solar and stellar flares: Transient phenomena, bursts, are familiar features in the radio continuum emission from the sun and nearby active stars. Lacking millimeter-wave observations, the physical processes which give rise to several important types of solar and stellar bursts cannot be firmly established. The most enigmatic are the apparently correlated gamma-ray–millimeter-wavelength solar flares. Recent evidence (mostly from the Solar Maximum Mission spacecraft and millimeter-wavelength radiometers) has demonstrated that electrons and protons are accelerated almost simultaneously to very high energies. In particular, electrons attain energies of 10 to 100 MeV within one or two seconds of the onset of the flare, and emit both millimeter waves and continuum γ-rays of high intensity. This continuum radiation is accompanied by nuclear γ-ray lines at energies less than about 10 MeV due to protons, and neutrons are sometimes detected at Earth.[1]

At the present time there is no widely accepted explanation for this very rapid acceleration. Some argue that a "first phase" process must be the cause because of the very short time scale, possibly involving electric fields in double

[1] *Cf.* Chupp, E. L. (1984), *Ann. Rev. Astron. Astrophys.* **22**, 359.

layers. Others argue that stochastic acceleration can act on short enough time scales.[2,3] The discriminating observations are unavailable. At radio frequencies the distinguishing spectral characteristic of γ-ray–millimeter-wavelength flares is that the flux density increases with frequency into and perhaps beyond the millimeter regime. Observations at millimeter wavelengths are obviously of great interest, but there have been no studies, either at millimeter wavelengths or at γ-ray energies, with reasonable spatial resolution. Moreover, since the angular scale of the active regions, as derived from the time scale, can be no larger than a few arcseconds, a measurement of the actual angular size must come from the millimeter-wave emission—not even NASA's Gamma-Ray Observatory (GRO, to be launched around 1990) has anything approaching arcsecond resolution at photon energies greater than a few MeV.

With the firm evidence that some physical process is capable of producing relativistic electrons on the sun so rapidly, we may expect to witness the same phenomenon in other objects where magnetic energy is available, *viz.*, on flare stars, interacting binary stars, and X-ray binary stars. At centimeter and decimeter frequencies, of course, this expectation is fulfilled.[4] Less clear, however, is the situation at millimeter wavelengths. Certainly, millimeter-wavelength observations of flare stars and close binaries offer the important advantage of allowing detection of fast electrons much closer to the site of their acceleration, and for this reason we may expect a close correspondence between optical/UV flare emission and millimeter-wavelength emission. Any such correspondence observed would reveal a great deal about the conversion of magnetic to thermal energy in stellar active regions.

For all such nonthermal, continuum flares, proper imaging with a millimeter-wave array requires not only high sensitivity (i.e., low-noise receivers and wide bandwidths) but also excellent instantaneous u-v coverage since the time scales of the phenomena are but a few seconds to a few minutes. The MMA must provide good "snapshot" capability.

Planetary Atmospheres: Somewhat longer-term variations, especially diurnal variations, are the key to understanding the photochemistry and circulation of planetary atmospheres. Spectroscopic observations of important constituents in planetary atmospheres allow one to derive abundances of gases as a function of altitude, together with the temperature vs. pressure profile of the atmospheres. The line intensity provides a measure of the convolved abundance and temperature profile in the atmosphere, whereas the shape of the line depends on the presence and abundance as a function of altitude. Variations in these parameters are not only expected, but are pivotal, to our understanding of planetary atmospheres. Here potential MMA observations of CO on Venus and Mars are illustrative.

Imaging of CO in the atmospheres of Venus and Mars offers the possibility

[2]Melrose, D. B. (1983), *Solar Phys.* **89**, 149.

[3]Bai, T., Hudson, H. S., Pelling, R. M., Lin, R. P., Schwartz, R. A., and von Rosenvinge, T. T. (1983), *Astrophys. J.* **267**, 433.

[4]Hjellming, R. M. and Gibson, D. M., Eds. (1985), *Radio Stars*, (D. Reidel: Dordrecht).

to study the diurnal, latitudinal, and seasonal variations of the atmospheric temperature and CO abundance in a direct way. In the upper atmosphere of Venus, strong diurnal CO abundance gradients exist. Images of the $J = 1$–0, 2–1 and 3–2 lines may be inverted to yield the longitudinal and latitudinal CO abundance and temperature variation in the atmosphere in the 80–120 km altitude region. Such studies will provide fundamental input to models of the general wind circulation and photo-chemistry of this region in the atmosphere. In addition, direct measurement of winds by observing the Doppler shift of lines near the limb of the planet will substantially improve wind-circulation theories.

The CO abundance in the atmosphere of Mars is not strongly altitude dependent and, therefore, one can use the CO lines at millimeter wavelengths to sound the atmosphere for its temperature profile. Images of the planet Mars could then be used to produce global maps of the temperature profile on Mars to study latitudinal and seasonal variations, as well as the effect of large scale meteorological phenomena (e.g., global dust storms) on these profiles.

The burden that such a study of the diurnal variation of the photochemistry of planetary atmospheres places on the MMA is that it provide an excellent spectroscopic image, in more than one transition, in a single day's observation.

The same constraint, with the same time scale, is germane to cometary observations. Repeated spectroscopic observations are needed in order to investigate the effects of the varying intensity of the solar radiation field on the apparent line strengths. The MMA needs to provide a reliable spectroscopic image in several transitions in a single day's observation.

2. Wide-field images: mosaics. Questions concerning the evolution of a class of object from one state to another require observational data for many objects spread over a large area of sky. When the angular scale of the phenomenon is larger than the primary telescope beamwidth the ability to create a single image from many overlapping images is indispensable. This places a burden on the observing technique and on the capabilities of the instrument. Specifically, one must be able to observe rapidly a raster of adjacent fields with sufficient instrumental stability and reliable enough calibration to enable one not only to image each field but also to combine all separate images into a coherent whole. Since the u-v coverage in each separate field must be good enough to produce a well sampled image even though the data extend over one or a few hour angle ranges, the "snapshot" capability of the array must be nearly complete. Excellent instantaneous u-v coverage is a prerequisite for an array to image wide fields.

Mosaics of wide fields are needed to address many questions regarding

(1) the kinematic and chemical evolution of GMCs through the star formation stage;
(2) the gas content and distribution in galaxies; and
(3) the pressure of the hot (10^7–10^8 K) gas within clusters of galaxies.

Star Forming Regions: Understanding star formation begins with under-

standing the structure of a molecular cloud before stars form. The simplest scenario, in which a cloud simply collapses to form a star, does not seem realistic in general. Instead, the velocity fields of molecular clouds are dominated by supersonic turbulence, the source of which is not clear. A powerful probe of turbulence is the study of velocity fields as a function of scale size. For example, do they follow a Kolmogorov spectrum? Existing studies cover only about one order of magnitude in size. The MMA will offer a unique opportunity to obtain information over a much larger range of scales; 0.008 pc to 9 pc at 500 pc seems achievable. While the upper end of this range could be done on existing instruments, the MMA, using the mosaicing capability, allows coverage of this range of three orders of magnitude in a single observation. The speed provided by a synthesis array would allow an experiment in turbulent scale sizes to be done on several clouds with different properties (e.g., those forming high mass stars, those forming low mass stars, quiescent clouds, Bok globules). More importantly, the present data indicate that line widths decrease with decreasing scale size, so that the transition to subsonic turbulence can be explored only with an array's capability for high spatial resolution.

In pursuing the turbulent structure to small scales, one hopes to begin isolating individual turbulent elements (fragments) which could eventually become individual stars. In particular, the distribution of angular momentum on small scales within a much larger environment can be studied. By protostellar we mean that no luminous condensed object (a main-sequence star or pre- main-sequence star) has formed. Since the gravitational energy released by contraction is effectively radiated away during the phases we wish to study here, we assume that the temperature should remain low. These objects are in the natural domain of an MMA. We find that the best way to identify such fragments is through their dust emission at short wavelengths. As long as the dust temperature does not drop too low, the brightness temperature will be proportional to the dust temperature and the column density. Then, in the absence of temperature gradients induced by internal sources, the interferometer will respond primarily to enhancements in the column density which occur as a fragment contracts. The identification of such collapsing protostars within large GMCs is made possible by the mosaicing capability of the MMA.

Determination of the physical and chemical properties—temperature, density, ionization state, and relative abundances—is also necessary for testing models of protostellar collapse. Here the questions are not only the value of these parameters within protostellar clumps but, perhaps more importantly, how they differ from one clump to another and from the GMC as a whole. Already we have evidence that individual star forming regions are dissimilar; the differences are most pronounced in high resolution interferometer maps. Figure 1 compares maps of OMC1/IRc2 in different molecular transitions. The dissimilarities can partially be explained by differences in excitation and partially by differences in chemistry. The questions are: Why the differences and what do they mean? Large mosaic images of many clouds are needed to answer such questions.

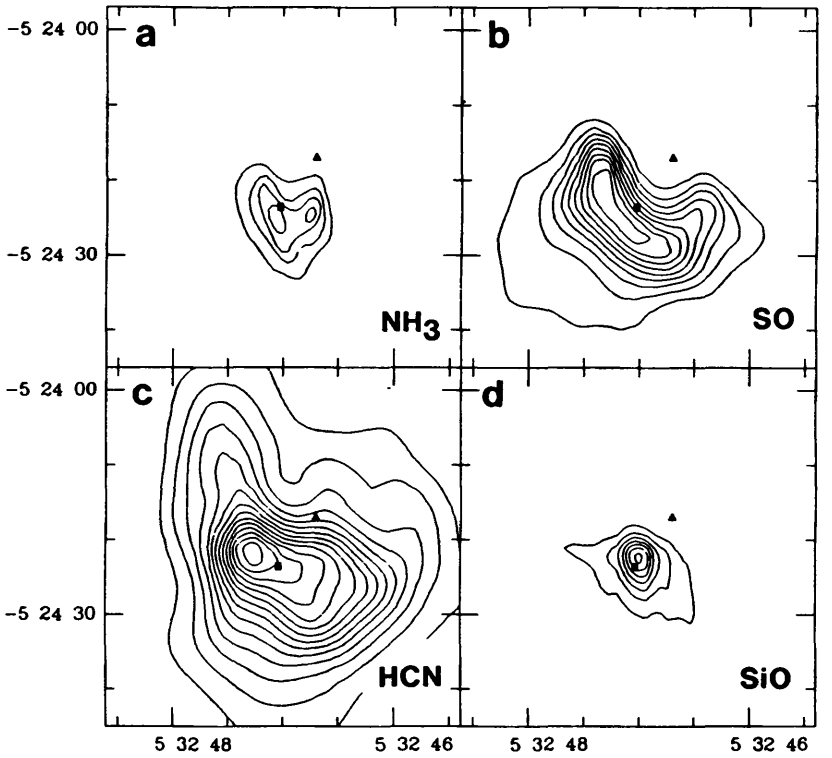


Figure 1. Millimeter-wave radio images of OMC1/IRc2 in different molecular transitions. [Plambeck, R. L., Vogel, S. N., Wright, M. C. H., Bieging, J. H., and Welch, W. J. (1985), in *International Symposium on Millimeter and Submillimeter Wave Radio Astronomy*, (Granada: International Union of Radio Science), p. 240.]

Spectroscopy of Galactic Disks: Puzzles of a similar nature are presented by the observed differences between the distribution of atomic and molecular gas in nearby galaxies. With a capability to produce mosaic images, quickly, of CO and other tracers, the molecular star-forming material in all the galaxies included in the Shapley–Ames catalog can be studied with a resolution of 1″ and a sensitivity sufficient to detect most giant molecular clouds in galaxies within 40 Mpc of the sun (considerably more distant than the Virgo cluster). Scientifically, the mosaic MMA images, together with VLA 21 cm images and optical data, will allow us to follow in detail the flow of interstellar matter from the atomic state to the molecular state, and subsequently into stars, for a variety of galaxy types in a wide range of galaxian environments.

In the Milky Way a large fraction of the high surface-brightness CO is from GMCs. By extension, with the MMA it will be possible to determine in detail the relationship of molecular clouds to spiral structure, how molecular gas is distributed in detail, and what relationship the molecular gas has to star formation and the general interstellar medium in galaxies of all morphological types in a large variety of environments. Because the integrated emission from galaxies can be detected to such large distances, it will be possible to

investigate how the molecular content affects the evolution of galaxies. For example, CO emission should be detectable from the several hundred ultra-IR-luminous galaxies thought to exist in the IRAS dataset, at distances to $z \approx 0.3$. Furthermore, within individual galaxies, the global gas kinematics will be observable with unprecedented resolution. New distance estimators, such as using tidally limited GMC diameters to calibrate physical size, may be possible with the MMA, leading to more precise measurements of the Hubble constant.

Wide-field MMA images of CO taken together with other ISM tracers will allow us to obtain a fairly complete picture of the global properties of the dust and gas in the interstellar medium, and the effects of spiral structure, HII regions and supernovae in large numbers of galaxies. The kinematic information obtained from the MMA will be superior to all of the other dense ISM tracers, however. Hα is subject to severe extinction effects, and the spatial resolution at CO and its isotopes will be more than an order of magnitude greater than what is possible in HI at the VLA. The high spatial resolution is of critical importance in at least two areas: determining the variation of the scale height of the gas with radius in edge-on spiral galaxies, and the mapping of central bars. The former is an important datum in determining the structure of possible non-luminous halos, and the latter in testing models of barred galaxies. High resolution studies of non-spiral systems will allow a detailed comparison to be made of properties as a function of morphological type. Mosaiced MMA images are crucial to these studies.

Temperature Decrements in the Microwave Background: Finally, perhaps the most demanding observations for which mosaic images are needed are those studies of the decrement in the microwave background expected, and seen, toward clusters of galaxies. Here the physical process (the *Sunyaev–Zel'dovich effect*) is well known. When the cosmic background photons encounter hot plasma, for instance intergalactic plasma in clusters of galaxies, they are inverse Compton scattered by the electrons in the plasma. This process increases the energy of the photons but conserves photon number. The effect is to shift an initially blackbody spectrum to higher frequencies, resulting, as shown in Figure 2, in a small temperature decrement, or cool spot, in the radiation if it is observed in the Rayleigh–Jeans regime through a cluster. The magnitude of the effect is directly proportional to the integral of the pressure through the intergalactic plasma, and thus provides a useful diagnostic for plasma properties. Of more general interest is the fact that observations of the Sunyaev–Zel'dovich effect provide a means to measure distances to clusters, distances which are independent of any intermediate rungs on the cosmic distance ladder. This possibility arises because the amplitude of the Sunyaev–Zel'dovich effect depends on the parameters of the plasma, but is independent of the distance. On the other hand, the measured X-ray flux from the hot intergalactic plasma does depend on the distance. Observations to date are not sufficiently sensitive to provide good distance estimates even to nearby clusters and, therefore, the power of this method in determining the basic cosmological parameter H_0 has

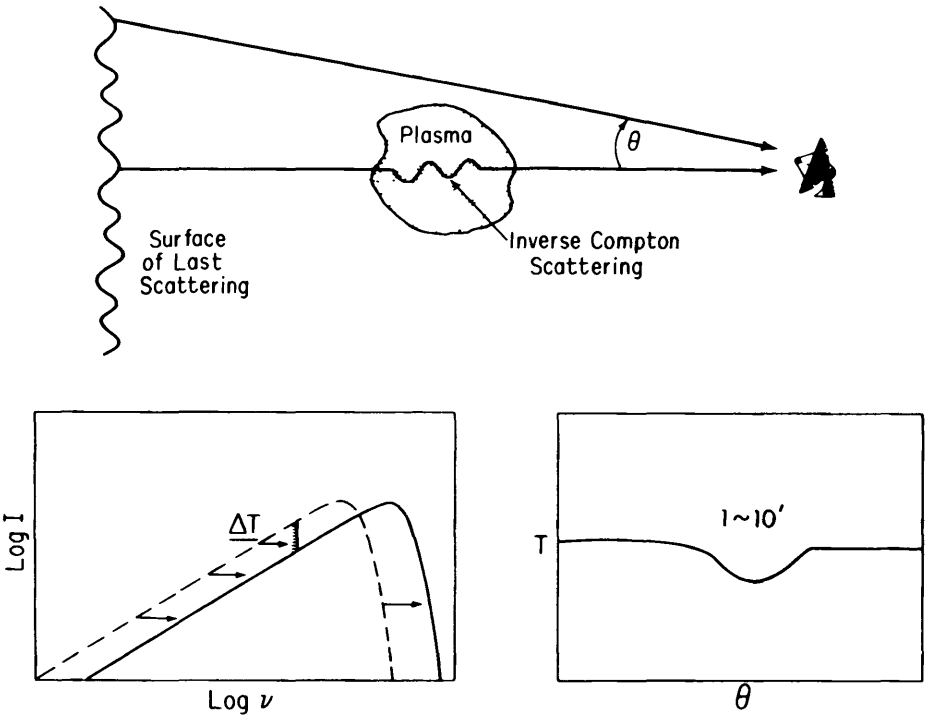


Figure 2. A portrayal of the mechanism by which small temperature decrements, or cool spots, are produced in the microwave background radiation via the Sunyaev–Zel'dovich effect.

not yet been utilized. The X-ray measurements necessary to refine our values of H_0 should be available within the next decade, thanks to planned space missions. What about the microwave observations? Two problems face us: the contribution of radio sources within the cluster or background sources, and the need to image the cluster so as to obtain information about the distribution of the cluster plasma. The latter requirement is important because the X-ray flux and the Sunyaev–Zel'dovich effect depend on different powers of the plasma density, so that the distribution of hot plasma in the clusters must either be mapped or modeled. The high frequencies at which the MMA will operate will help substantially with the first of these two problems. The ability of the MMA to make accurate images will also help with both problems mentioned above. Sources can be located and excised from the data, and the amplitude of the temperature decrement can be mapped with adequate accuracy. Since the amplitude of the Sunyaev–Zel'dovich effect falls off slowly with distance from the cluster center, the area to be mapped is several times the cluster core radius R_c, or since R_c is typically > 1 arcminute, regions as large as $10'$ in extent will need to be mapped at a few tens of arcsecond resolution. This will require very large mosaics.

3. Stable atmospheric transparency. Since the relative transparency of the atmosphere is continuously variable, those observations which depend on superior transparency must be attempted when suitable conditions occur and

must be completed as quickly as possible, before atmospheric conditions deteriorate. Observations of the Sunyaev–Zel'dovich effect toward clusters of galaxies, mentioned above, are illustrative in this context. Relatively long integrations are needed to reach the expected amplitude of the signal, and this sensitivity must be reached in each of the 10–100 overlapping fields which ultimately contribute to the mosaic of a cluster several arcminutes in diameter. The need to accomplish this sort of observation rapidly, while the best atmospheric conditions obtain, is evident.

All the highest frequency observations, near 350 GHz, must also be done in those intervals where the atmospheric transparency is most favorable. Here the precise duration (hours to days) of the "best" conditions depends on the site, but it is certainly bounded, and some crucial observations must be done in these periods. An example is provided by observations of the CO $J = 3$–2 line in distant galaxies. Mosaic images at this frequency could easily take many hours of integration; the faster the array the more such research will benefit.

3. Sensitive Imaging Capability at 230 and 345 GHz

A cornerstone for the science proposed for the MMA is its capacity to provide well-sampled, high resolution images at frequencies corresponding to the $J = 3$–2 and $J = 2$–1 transitions of the common molecular tracer CO and its isotopomers at the appropriate Doppler shifts. Spectroscopically, many important questions involve the spatial distribution and kinematics of molecular material, and such questions are best addressed by appeal to observations of strong lines of an abundant, and widely distributed, constituent. Carbon monoxide fulfills these requirements. In the continuum, observations at 250 and 350 GHz provide a high resolution, uncontaminated, view of the emission from thermal dust grains and solid surfaces. Owing to the steep spectral dependence of thermal emission, one can obtain this particular insight only by observing at wavelengths of 1 mm and shorter. A wealth of scientific opportunities is opened to astronomers given access to an imaging instrument at these wavelengths.

1. CO spectroscopy. The importance of observations of carbon monoxide with the MMA cannot be overemphasized. Not only are the rotational transitions of CO sensitive indicators of the temperature of molecular matter, but isotopic observations can provide information on the abundance of trace elements throughout the ISM in the Milky Way and in other galaxies. Further, the integrated flux density in the rotational transition $J \rightarrow J - 1$,

$$S_J = \lambda^{-4}(J+1)e^{-hc(J+1)/(2kT_{ex}\lambda)},$$

increases with increasing J in the Rayleigh–Jeans regime. If one can produce images in more than one rotational transition, the spatial variation in excitation temperature, T_{ex}, then can be established. In any event, the higher rotational transitions (higher J) sample more highly excited gas, higher kinetic temperature, and/or higher density. A complete picture of the thermodynamics and kinematics of a molecular region requires CO images in several rotational transitions.

Galactic Disks and Nuclei: Observations of external galaxies, where sensitivity is at a premium, will greatly benefit by an imaging capability at CO $J = 2$–1. Observations in the $J = 2$–1 line will provide four times the sensitivity that observations in the $J = 1$–0 line provide, assuming equal signal strength and comparable system temperatures for both transitions. With this sensitivity, it will be possible to produce images of isotopic tracers such as ^{13}CO and $C^{18}O$ throughout the disks of dozens of nearby spiral galaxies, and thereby address such questions as the extent to which astration and enrichment of processed material affects the chemical evolution of galaxies. These questions can even be addressed as a function of Hubble type. The structures of GMCs in the disks of galaxies can be investigated for galaxies throughout the local group. Thus, the possible variation of the structure of molecular clouds can be assessed as a function of galactic environment.

The kinematics of the central regions of galaxies may be studied with a resolution corresponding to linear sizes as small as 5 pc for galaxies in the Virgo cluster to perhaps 50 pc in galaxies 100 Mpc distant. These highly obscured and mysterious regions are often inaccessible at optical wavelengths, yet they seem to hold an important key to the understanding of the energetics, morphology, and evolution of their host galaxy. Here the pivotal questions regarding the circulation of gas in the nuclear region can be answered from the CO observations in a far more unambiguous way than is true for the center of our Galaxy.

Protostars and Circumstellar Shells: The capability to observe the higher order CO rotational transitions is needed to understand the kinematics and excitation of both protostars and evolved stars. The unambiguous identification that a "cold-core" fragment seen in a molecular cloud is indeed a protostar requires the kinematic signature of collapse, probably combined with rotation. For this purpose, high spectral and spatial resolution is required in order to map the line centroid and shape across the object. The ordered motions of collapse and rotation are quite small at early stages and will only emerge from the turbulent velocity field when the object is quite small. The free fall velocity reaches 3 km s^{-1} when a $1M_{\odot}$ protostar reaches a radius of 140 A.U. ($1''$ at the distance of the nearest clouds). Consequently, only the high spatial resolution of an array will be capable of mapping the velocity fields in these objects.

Determination of the physical properties (temperature, density, ionization state, and chemical abundance) is also necessary for testing models of protostellar collapse. In particular, one wants to map these properties across the object. The sample probes that have been used in extended clouds (CO for temperature, rarer CO isotopes for column density) can be used on the protostellar fragments until the densities become so great that these species become optically thick or thermalized. For determining the density, multiple rotational transitions are required.[5]

[5]Snell, R. L., Mundy, L. G., Goldsmith, P. F., Evans, N. J., II, and Erickson, N. R. (1984), *Astrophys. J.* **276**, 625.

Observations of the gaseous circumstellar shells surrounding evolving stars are also of great importance. Presently we know that mass loss accompanies stellar evolution along the giant branch and that subsequent stellar evolution is modified according to the rate and duration of the mass-loss phase, but we do not know the physical process by which the extended circumstellar envelopes are produced.

CO observations yield well-defined values of the velocity of mass outflow (typically 4–60 km s^{-1}) and the stellar systematic velocity. CO emission from the circumstellar shells of more than 120 stars has been detected,[6] and more than 300 are now detectable with current technology. The inferred mass-loss rates of 3×10^{-8} to 3×10^{-4} solar masses per year indicate that late-type stars return significant amounts of processed material (as much as 0.3 solar masses per year) to the ISM in the Galaxy. Virtually all the details of the physics of circumstellar shells have yet to be confirmed observationally, owing to the limited angular resolution of the instruments used. Further, given the steep gradient in excitation temperature in an outflowing molecular shell, the needed observations are those in several rotational transitions which allow the full spectrum of excitation conditions to be sampled. The CO $J = 3$–2 and CO $J = 2$–1 capability of the MMA is crucial in this regard.

Planetary Spectroscopy: The unique opportunity, provided by the MMA, to image the $J = 2$–1 and $J = 3$–2 lines of CO in Venus and Mars offers the possibility to study diurnal, latitudinal, and seasonal variations of the atmospheric temperature and CO abundance in a direct way. Here use of several rotational transitions is needed because different transitions preferentially sample different altitudes in the atmosphere—higher excitation lines are formed lower in the atmosphere. Hence one can derive an altitude distribution for the abundance of CO.

Images of the $J = 1$–0, 2–1, and 3–2 lines provide the fundamental input to models of the general wind circulation and photochemistry of the atmosphere. We can look forward to surprises. The apparently non-terrestrial $CO/^{13}CO$ ratio in the Venusian atmosphere is a puzzle that apparently reflects poorly understood fractionation processes during the planet's evolution or within the early solar nebula.[7,8]

2. Thermal continuum radiation from dust and solid surfaces. High resolution imaging of dust and solid surfaces will be one class of research for which the MMA will provide an unparalleled capability. As a true imaging instrument at 230 and 345 GHz, the MMA will be able to take advantage of the very steep spectral dependence of the thermal dust emissivity (which varies from ν^2 to ν^4 at millimeter wavelengths) to enable astronomers to study thermal emission in a wide range of physical contexts. Unprecedented opportunities are presented in virtually every area of astrophysical investigation.

[6]Zuckerman, B. (1986), *Astrophys. J.* **311**, 345.
[7]Schloerb, P. (1985), in preparation.
[8]Clancy, R. T. and Muhleman, D. O. (1985), *Astrophys. J.* **273**, 829.

Planetary Surfaces: At different wavelengths one probes to different depths in the crust—roughly several to several hundreds (for pure water ice) wavelengths deep. The measured brightness temperatures depend upon the physical temperature of the subsurface layers and on the radiative transport of the thermal emission outward. The physical temperature depends upon the solar insolation and the thermal conductivity and thermal inertia of the material. Radiative transfer outward is limited by the emissivity and the absorption/scattering characteristics of the material, both of which are highly dependent upon wavelength. Thus probing the surface layers at different wavelengths allows mapping of the temperature distribution with surface depth. This, together with high resolution images in polarization, as well as total intensity, will ultimately allow one to deduce the composition and characteristics of the surface (e.g., solid rock, loose dust, gravel; ice in clumps, wet ice; etc.). The importance of observations at submillimeter and millimeter wavelengths is that one typically probes in the region of the diurnal solar heat wave—this gives the information necessary to derive the thermal characteristics (conductivity, thermal inertia, dielectric constant) of the material—properties unique for each substance.

Circumstellar Shells: The production of dust grains in the stellar photosphere of red giant stars and the subsequent ejection of the outer layers of the extended stellar atmosphere by radiation pressure on the grains are intimately related questions. Very high sensitivity and high resolution ($\lesssim 0\farcs1$) images of the thermal dust continuum emission in a sample of O and C giant stars will answer many questions:

(1) dust abundances can be compared with gas abundances as a function of position in the outflow;
(2) regions of grain formation can be identified; and
(3) the regions of envelope acceleration can be identified and compared to regions of grain formation.

In principle, many of these questions can be addressed with infrared observations; the MMA extends such studies to higher angular resolution.

Molecular Clouds and Protostellar Disks: We remarked earlier that the identification of protostars within a large, highly opaque, molecular cloud was best accomplished by means of their thermal dust continuum emission at short millimeter and submillimeter wavelengths. Since the brightness temperature of the dust within condensing regions is proportional to the dust column density, such protostellar clumps will appear as "hot spots" in continuum images of molecular clouds. Here we can appeal to the IRAS survey to demonstrate that this expected brightness distribution is indeed observed. The IRAS survey has identified hundreds of IR bright spots generally embedded in dense clumps of gas which have no optical counterparts. Those clumps with very low 60–100 μm color temperatures are likely to be the youngest objects. Notable among these sources is the isolated globule B335. Observations with the MMA will complement the IRAS observations by providing an image of the material

structure near the proto-star with sub-arcsecond resolution.

There is growing evidence that gas and dust disks are common features around young stars. Such dense disks lying in the equatorial plane of a very young star serve to deflect and channel outflowing material from the star into collimated, oppositely directed, flows. The resultant bipolar outflows are so commonly observed around newly formed stars[9] that one is driven to the conclusion that they are a necessary phase of very early stellar evolution. An understanding of the nature of the circumstellar disks—their masses, structure, and stability, the role of the magnetic field in their maintenance—rests on an ability to make very high resolution ($< 1''$) observations of their thermal dust continuum radiation. The MMA continuum imaging capability at 250 and 350 GHz is uniquely suited to provide the needed observations.

Galaxies: The ability to provide arcsecond and sub-arcsecond images of thermal dust emission will be of equally great import for studies of the dust content, distribution, and gas-to-dust ratio in both nearby and quite distant galaxies. Since the thermal dust radiation is optically thin and relatively insensitive to temperature at 250 and 350 GHz, the images obtained from the MMA will be a reliable measure of the dust mass distribution throughout the galaxies. Specifically, the MMA will be sensitive enough to provide images of dust complexes in galactic nuclei at cosmologically interesting distances (200 Mpc!). Imaging of such regions at a resolution of 100 pc will be possible for galaxies at the distance of the Virgo cluster.

Primeval Galaxies: A fundamental problem in astronomy is the origin of galaxies. We have available a range of models for the formation of galaxies—but no direct observations of this process, unlike the situation for stars. That is because galaxy formation occurred in the distant past. A variety of arguments suggests that most galaxies form their first generation of stars at an epoch corresponding to redshift $3 \lesssim z \lesssim 30$. The presently available upper limits on fluctuations in the cosmic background make the interval $3 \lesssim z \lesssim 10$ more plausible.

Models of primeval galaxies suggest a large energy release during the formation period, of order 10^2–10^3 times the present luminosity of galaxies over a period 10^8–10^{10} years. A major component of the luminosity is the emission from massive stars; hence a very high UV flux in the rest frame of the objects is predicted. At $z = 3$–10, the UV emission is shifted to optical wavelengths. Optical searches for primeval galaxies, however, have failed to find them. There is a natural explanation for this failure: absorption of the UV photons by interstellar dust in the star-forming galaxies. The dust must re-radiate thermally with a peak at submillimeter wavelengths.

Strong support for this basic picture is provided by our best local analogues of primeval galaxies, the star-burst galaxies. These are bright IRAS sources at 100 μm, and in many cases the spectra are still rising, suggesting a dust

[9]Lada, C. J. (1988), in *Galactic and Extragalactic Star Formation*, R. E. Pudritz and M. Fich, Eds., (D. Reidel: Dordrecht), in press.

temperature < 30 K. Even these local sources are highly luminous—of order 100 Jy at 100 μm at distances of 100 Mpc.

Whether we adopt these figures, or make projections from models of primeval galaxies, we arrive at the same conclusion: primeval galaxies at $z \approx$ 3–10 may be detectable at 100–300 GHz using the MMA. The angular scale of the sources will be $> 5''$ (depending on the intrinsic properties, z, and cosmological parameters like H_0 and q_0). The number per square degree also depends on the cosmological parameters, but is large enough (of order 100 deg^{-2}) to permit searches for such sources in regions free of other sources. The expected flux densities lie in the range 0.3–30 mJy.

Two distinct advantages of the MMA over other planned and current instruments are its ability to map a two-dimensional region of the sky at adequate sensitivity and the freedom of millimeter-wave observations from emission by Galactic IR cirrus. In addition, a wide range of available frequencies is an advantage since the peak of thermal dust emission depends on two unknown factors, the redshift of the primeval galaxies and the characteristic temperature of their dust:

$$\nu_{\rm peak} = \frac{60T}{1+z}\,{\rm GHz}.$$

If these expectations are correct, dusty primeval galaxies at high redshifts will be the dominant contributors to source confusion at λ = 1–3 mm. The MMA is uniquely suited to these astronomical investigations.

4. Wideband/Multi-Band Capability

The ultimate goal of astrophysical research is to reach an understanding about the nature and evolution of objects in the universe. Rarely is such an understanding reached with observations at only one particular wavelength. So it is perhaps not surprising that the scientific demands placed on the millimeter-wave array emphasize wideband and multi-band capabilities. Wide bandwidths increase continuum sensitivity, while a true multi-band facility allows the astronomer to explore the relation of information in one band to that in another. Both capabilities are important to the scientific value of the MMA; together they make it unique.

1. Multi-band continuum sensitivity. With an instrument for imaging continuum radiation, one would, of course, usually like to use wide bandwidths to maximize the sensitivity. But in addition, access to this sensitivity at two (or more) well displaced frequencies will more severely constrain the physics of the phenomenon under investigation. All types of continuum imaging observations will benefit from sensitive, multi-beam observations.

Millimeter-Wave Flares on the Sun and Stars: Solar and stellar flares at centimeter and decimeter wavelengths arise in the corona and release stored magnetic energy as thermal bremsstrahlung emission from heated plasma. Nonthermal emission at low harmonics of the gyro-frequency are also observed. These two physical processes are distinguishable by their spectral shape and

spectral evolution through the duration of the flare. We can anticipate similar phenomena at millimeter wavelengths, with the important difference that millimeter-wave flares have their origin in the low chromosphere. Simultaneous multi-band continuum observations with the MMA (at 35, 90, and 250 GHz, for example) will provide images of the variation of brightness temperature with frequency and time, which can in turn yield information on the temporal evolution of the temperature and density gradients in the chromosphere. From this one can infer the rate of thermal conduction.

Photospheres of Normal Giant and Main Sequence Stars: The ν^2 dependence of the flux density of the thermal emission from a blackbody, such as is characteristic of a stellar photosphere, favors the detection of normal stars at sufficiently high millimeter-wave frequencies. About 600 stars, a few on the main sequence, but most of them giants, will have 250 and 350 GHz flux densities greater than 1 mJy simply from their thermal photospheric emission; these stars will be detectable with the MMA, and their positions will be established with astrometric accuracy.

Because $S \propto T_{\rm eff} {R_\star}^2$, observations at millimeter wavelengths can check the canonical values (mostly inferred indirectly, few measured) of $R_\star$, or with less sensitivity, $T_{\rm eff}$. One should expect surprises, with radio sizes or effective temperatures differing from canonical values.

In addition to simple detections, several other studies become possible with high enough sensitivity, for example:

(1) starspots and/or active regions would modulate the emission at the stellar rotation period. Long-term monitoring will allow direct determination of the way that cycles of stellar activity relate to spot numbers and sizes;

(2) stellar pulsations, expected in some regions of the HR diagram, may modulate the flux enough to be detected;

(3) metrology of star positions over periods of $\gtrsim 1$ year, with a resolution of a few milliarcseconds, may reveal the binary nature of some stars and perhaps demonstrate the existence of low mass (planetary) companions.

Winds of Hot Stars: Observations at 90, 250, and 350 GHz are of fundamental importance to studies of winds from hot stars. Current problems with wind fluxes at centimeter wavelengths (variability, spectral oddities, and failure to fit expected visibility curves with model temperatures) argue that data at millimeter wavelengths will be critical in separating thermal from nonthermal radiation.

The simple theory of radio emission from winds predicts $S_\nu \propto \nu^{0.6}$ and $R_\star \propto \nu^{-0.6}$, with typical 2 cm values $S_\nu \approx 100$ mJy and $R_\star \approx 0\farcs1$. As shown below, this rising spectrum allows detection of many wind sources at millimeter wavelengths. Effects of nonthermal electron populations are much smaller at millimeter than at centimeter wavelengths, so that "pure" wind radiation should be more common, and uncertainties in derived mass-loss rates should be smaller.

Because of the $\nu^{0.6}$ spectrum, the flux densities of wind sources are roughly ten times larger at 250 GHz than at 6 cm. The mass-loss rate $\dot{M}$ goes as $S_\nu^{3/4}$. If the sensitivity at 1.3 mm were the same as currently available at 6 cm, then mass-loss rates a factor of six times smaller could be observed, or sources currently detectable at the VLA could be seen three times farther away than now possible. All known galactic Wolf–Rayet stars could be detected, roughly doubling the sample size. Ninety new O–B star detections should be added to the 16 now detected with the VLA.

The Abundance of Dust Relative to Gas: In many astrophysical environments, from circumstellar shells to molecular clouds to galaxies, the ratio of gas to dust is of great chemical and physical importance. In the case of circumstellar shells the sensitivity of the MMA to thermal dust continuum emission at 250 and 350 GHz will allow astronomers to investigate the formation sites of dust grains in the outflow. Useful comparisons at the same angular resolution can be made between the MMA images and those from IR speckle interferometry, the latter of which suggest that the dust around low-mass stars is distributed in disks.[10,11] The linear extents of the disks, a few hundred A.U., are suggestive of pre-planetary systems and are accordingly of particular interest. Differences in the dust/gas ratio in molecular clouds in galactic disks may be indicative of real variations in the chemistry applicable to different environments. Sensitive, multi-band continuum images of the dust emission are needed to investigate these questions. Observations at 250 and 350 GHz provide the most unambiguous answers.

Cosmologically Distant Sources: Finally, we mentioned above the utility of searching for redshifted thermal dust emission from primeval (z = 3–10) galaxies. Here we emphasize the need for broadband observations, 35–350 GHz (to establish the thermal character of the emission and our estimate of the redshift), and the need for great sensitivity owing to the very weak signal expected (0.1–30 mJy). A similar statement is also germane to observation of fluctuations in the microwave background and to searches for a decrement in the background radiation toward clusters of galaxies. For this type of observation, the greater the continuum sensitivity the greater the scientific return.

2. Multi-band spectroscopy. The detection of a molecular spectral line gives the astronomer two pieces of information directly: (1) knowledge that the molecular species is present in the direction being observed, and (2) the velocity of the molecular material. Usually the scientific question being posed requires observational insight of a more specific character, and when this is the case observations of more than one spectral line are needed. The ability to observe multiple lines simultaneously amplifies the utility of the MMA as a scientific instrument.

---

[10]Beckwith, S., Zuckerman, B., Skrutskie, M. F., and Dyck, H. M. (1984), *Astrophys. J.* **287**, 793.

[11]Grasdalen, G. L., Strom, S. E., Strom, K. M., Capps, R. W., Thompson, D., and Castelaz, M. (1984), *Astrophys. J. (Letters)* **283**, L57.

Generally if one seeks to understand the kinematics of a particular molecular cloud or condensation, any spectral line is suitable, so one chooses CO $J = 2$–1 so as to obtain maximum sensitivity. However, if opacity in the surrounding cloud is likely to distort the emission profile from the ambient cloud, then lines of lower opacity or higher characteristic density can be used to discriminate against the emission from the more extensive materials. For the highly inhomogeneous molecular regions one usually studies—circumstellar shells, planetary atmospheres, galactic disks, collapsing protostars, bipolar flows—one needs to obtain images in abundant species such as CO to map the temperature distribution; in the rarer isotopomers, such as ^{13}CO and $C^{18}O$, to map the distribution of column density; and in high dipole-moment species, such as CS, to map the density distribution.[5] Clearly it is an advantage and an economy to obtain all this information simultaneously through multi-band spectroscopic observations; often it is crucial to the science to do so.

Abundance Gradients in Galactic Disks: The opportunity to map the abundance of enriched material with galactocentric radius in nearby spiral galaxies will be possible with multi-band MMA spectroscopy. Variations in the abundance of D, C, O, N, and S in galactic disks will illuminate those areas with elevated rates of nuclear processing and will lead to an indication of the large-scale chemical evolution of galaxies. For these purposes, images of the emission from CO, ^{13}CO, $C^{18}O$, DCN, DCO^+, HCN, and CS are needed. In addition, images of the molecular ion HCO^+, which is sensitive to the local cosmic ray energy density, are needed to map the galactic cosmic ray intensity. The MMA will have sufficient sensitivity to provide such images in dozens of nearby galaxies, but some integrations will be long. For this reason it is appropriate to employ a spectrometer of sufficient flexibility to allow all images to be made simultaneously.

Disks and Outflows Around Young Stars: The chemical composition of the ubiquitous disks seen around stars is of great interest not only because of the role the disks play in focusing the stellar outflow but also because these disks are often invoked as being the pre-planetary environment. The chemical heterogeneity seen in our own planetary system may have its origin in a more common, highly-structured chemistry of the environs of young stars.[12] To investigate these questions images are needed of many disks and of many chemical species in each disk. Particularly crucial are the relative distributions of ^{13}CO and $C^{18}O$; DCO^+ which avoids hot cores and so traces ambient gas; SO which is found in turbulent shocked regions; SiO which probes hot outflows; and the dust. One image, without the others, is of limited diagnostic utility. Multi-band spectroscopy is clearly needed here.

Interstellar Chemistry: To understand the nature of interstellar chemistry itself, several different types of regions must be studied in the lines of many molecular species. The interface regions between HII regions and ambient molecular clouds are important in understanding the central role of carbon

[12]Vogel, S. (1985), private communication.

chemistry, particularly that of atomic carbon itself. Such regions are very small in angular extent and require the high resolution of the MMA. The cold cores of dark clouds, which are the precursors of star formation, have their own unique chemistry, including the remarkable long-chain carbon species (cyanopolyynes). Except for study of the nearest few such objects, the resolution of the MMA is essential for these objects as well.

Chemistry of Circumstellar Shells: The chemistry of circumstellar shells differs fundamentally from interstellar chemistry in that it originates largely in hot, dense regions close to the photosphere. The stellar wind produces the shell by propelling these molecules outward. As the molecules move out in the shell the rapidly falling density lowers chemical reaction rates to levels where they are unimportant and the relative abundances of molecules become "frozen" and independent of distance from the central star. The equilibrium breaks down at sufficiently large radius because some species are absorbed into grains and others have their abundance modified through the action of the interstellar radiation field as they are exposed near the surface of the stellar envelope. Images of many species are needed to unravel the effects, and the relationships, of these processes.

Photo-Chemistry of Planetary Atmospheres and Comets: The need for true simultaneous multi-band spectroscopic observations is nowhere more evident than in the study of planetary atmospheres and comets. We noted earlier that a determination of the abundance and excitation of CO on Mars and Venus requires observation of the $J = 3$–2, $J = 2$–1, and $J = 1$–0 lines. Owing to the time-dependent (diurnal and seasonal) nature of the photochemistry in these atmospheres, images of the three CO transitions must be obtained simultaneously.

A similar need is evident when studying trace constituents in planetary atmospheres. Here the altitude distribution of a particular constituent (e.g., PH_3, H_2S, HCN, SO, SO_2, HC_3N) provides a clue as to its place of origin. For example, species seen deep in the atmosphere are likely to have been brought up by convection, while those found high in the atmosphere may have fallen in from outside (from rings or satellites) or may have been created by dissociative photolysis.

Multi-band spectroscopic observations of comets are especially important for deciphering the composition of the early solar system. Images of the outflow and extent of primary constituents such as HCN, HC_3N, and other nitriles and hydrocarbons may be related to those of secondaries such as CN, C_2, C_3, CO, and CS—to provide a clear picture of cometary photochemistry. These are highly time-dependent phenomena for which the rapid, multi-band imaging capability of the MMA is crucial.

III. General Design Principles

The scientific possibilities in the millimeter-wavelength region of the electromagnetic spectrum outlined in Chapter II are important and exciting, and affect almost every area of modern astrophysics. However, the demands on the design of an instrument to do most or all of them are quite challenging. Some form of compromise between conflicting goals is vital since the whole range of observations described are not feasible without a versatile instrument. The final design for the array will probably be neither just a single aperture nor just a simple interferometric array. Instead, some hybrid design encompassing the virtues of both may be desired. Such a hybrid design will also require innovations in the techniques used for data processing. In fact, the formulation of initial design concepts for the MMA has already stimulated promising new ideas—some of which (e.g., maximum entropy mosaicing) have been applied to existing instruments to solve long-standing observational problems.

In this chapter, we will outline some of the general principles of designing a suitable hybrid instrument. We will try to address the problems very generally, so that the discussion will apply to any instrument designed to solve them. In Chapter IV, we will describe a *particular* solution based on principles discussed in this chapter.

When designing an instrument such as the MMA, there are a number of goals which may partially or totally conflict. The final design must arise as a compromise between these factors—a compromise determined by considerations of cost, technical feasibility and scientific importance. With this in mind, we have found it useful to break down the goals as follows:

- Sensitivity
- Angular resolution and field of view
- Imaging speed
- Frequency coverage
- Spectral line and continuum capabilities
- Low systematic errors
- Solar imaging
- Data handling.

We will now discuss each of these items in detail.

1. Sensitivity

We have adopted the target of a collecting area of 1000–2000 square meters suggested by the Barrett committee as a crude guide to the *point-source* sensitivity of the array. The sensitivity of the array to *surface brightness* is determined by the distribution of this collecting area in the u-v plane. We have not placed any specific constraint on this latter form of sensitivity, and so the actual array configuration is driven by other factors, such as the sidelobe level, and by cost considerations.

2. Angular Resolution and Field of View

Perhaps the most challenging problem in the design of the array is the wide range of angular scales we wish to study. Scales from degrees to less than an arcsecond are implied in Chapter II. In many situations we wish to study the fine-scale structure of very large extended regions with good surface-brightness sensitivity. Inevitably though, for a fixed collecting area, the surface-brightness sensitivity of any telescope degrades as the resolution improves. Therefore, in some cases, the need for high surface-brightness sensitivity and high resolution will prohibit certain experiments because of the huge amount of total observing time necessary, especially if narrow bandwidths and mosaicing are important.

Let us divide the remaining types of observations into two classes:

(a) observations of large fields at moderate resolution; and
(b) observations of small fields at very high resolution.

(a) Observations of Large Fields at Moderate Resolution. One basic dilemma of imaging at millimeter wavelengths is that the requirements for large field of view and for good sensitivity drive the optimum element size in opposite directions. Good sensitivity necessitates a large amount of collecting area, which in turn demands either a small number of large elements or a large number of small elements. The latter is totally impracticable: at a wavelength of one millimeter, a thirty-centimeter antenna would have a field of view of about 15 arcminutes, which is still too small for many interesting objects; to provide the total collecting area specified by the Barrett subcommittee (say 1500 square meters) would require 21,000 such elements, implying huge costs in receivers, the correlator, and computing. The former option is more viable: only twenty 10 m elements would give the same collecting area, but the field of view would be 0.5 arcminutes, implying one-thousand separate pointings to image a 15 arcminute field.

Thus, failing a dramatic drop in the costs of electronics, we believe that the only currently feasible solution involves an intermediate number of intermediate-sized elements. This implies three specific and intertwined requirements for the array. First and most fundamentally, in order to image objects larger than the primary beam, we must be able to combine observations from many different pointings of the array into a single image. We use the term *mosaicing* when referring this technique. To understand mosaicing, consider first single-dish mapping: when imaging an object with a single aperture, one must scan the object, taking samples at spatial separations determined by the Nyquist theorem. The same is true for an interferometric array imaging an object larger than the field of view of the individual elements: we must point the array at many positions across the object and then combine the data later in the computer, correcting each pointing for the individual antenna power patterns. In the u-v plane, the effect of mosaicing is to separate out estimates of some of the range of spatial frequencies sampled by each interferometer. Ekers and Rots[1] showed that this separation is possible if the illumination pattern of

[1] Ekers, R. D. and Rots, A. H. (1979), "Short spacing synthesis from a primary beam scanned

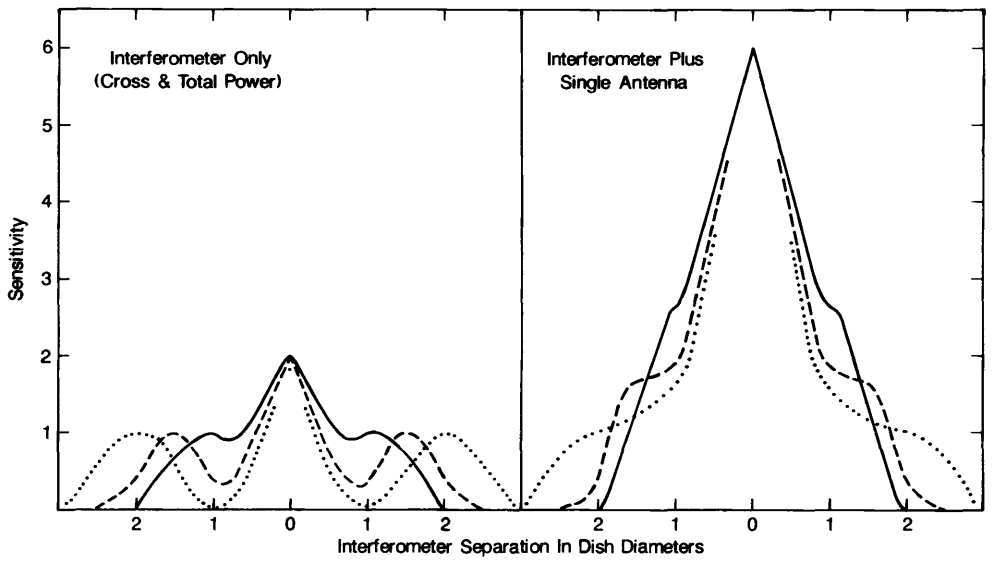


Figure 1. The sensitivity function for a two-element interferometer (left panel), and for that interferometer supplemented by a single dish of twice the diameter of the interferometer elements (right panel). The sensitivity function is shown separately for interferometer separations of 1.0, 1.5, and 2.0 dish diameters by the solid, dashed, and dotted lines, respectively. The total-power measurement from the interferometer elements has been included. All dishes are uniformly illuminated.

the antennas is known, and that it can be effected by Fourier transformation of the sampled visibility function, followed by division by a transfer function. A mosaicing algorithm achieves the same result, but rather more flexibly since actual regular scanning is not required (see Chapter V). The limit on the accuracy with which a spacing can be recovered depends upon the uncertainties in the illumination pattern and upon the sensitivity function of the instrument. The success of mosaicing will depend upon a number of trade-offs: the illumination pattern of the individual elements must be known to some accuracy, and, most importantly, the power in variable sidelobes must be minimized, while not degrading the beam efficiency.

Second, in order to completely specify an image we must include data for "all" spatial frequencies, including data for baselines shorter than the shortest baselines well-sampled by the array. That is, we must solve the so-called "short spacings" problem in interferometry. We can solve this problem by including single-dish scans of the field in our mosaicing process. Figure 1 shows that a single interferometer of sufficiently short baseline can, in fact, sample in this range. The total-power measurements from two uniformly illuminated dishes, plus the response from the same two dishes in interferometer mode and separated by one dish diameter, produce good sampling for all the low spatial frequencies. The difficulty is to arrange for a sufficient integration time for the shortest interferometer spacings without having too much shadowing. For

interferometer", in *Proc. IAU Colloquium No. 49 on Image Formation from Coherence Functions in Astronomy*, C. van Schooneveld, Ed., (D. Reidel: Dordrecht).

example, an interferometer observing a setting source spends only a limited time sampling these spacings, and eventually shadowing of one element by another may occur. Shadowing changes the effective illumination pattern of the elements, and thus considerably complicates mosaicing. Although some degree of compensation may be possible, it would probably be too computationally intensive. We must therefore design the array to obtain reasonable integration time on these short spacings. There are three possible solutions to this:

(1) We can separate the elements sufficiently that shadowing does not occur for any plausible zenith angle. To avoid shadowing of the surface of one element by another, the minimum element spacing must be greater than about sec z-antenna diameters, where z is the zenith distance, which must be maximized to obtain good u-v plane coverage and long integrations, and, since sec z is the airmass, must be minimized to alleviate the effects of atmospheric emission and absorption. Taking, say, an airmass of 2 as the limit, the minimum baseline would be two diameters. The sensitivity function of the interferometer then has a null at a spacing of about one diameter. To fix this we can add measurements from a system consisting of different size elements. As an example, Figure 1 also shows the improvement in sensitivity upon adding the measurements from a single dish of twice the canonical diameter, although one could also add measurements from an interferometric array consisting of *smaller* elements.

(2) We could design the array with a range of short spacings, and accept the shadowing when it occurs. Some loss of efficiency is inevitable in this scheme.

(3) We could eliminate shadowing by mounting the most densely packed elements on a structure which can track the source. Thus, shadowing never occurs, and the same canonical element size can be used down to spacings of one diameter. The null in sensitivity at a separation of about one diameter is thus avoided.

The first scheme is basically that used by currently existing interferometric arrays, such as WSRT, in which the short spacings are obtained from single-dish images made with large apertures such as the Effelsberg 100 meter telescope. The conventional wisdom is that the ratio of element sizes must be at least 2 to 3 since heavily-tapered illumination patterns are used in single dishes. However, the other two schemes are more innovative, since the resulting array could be made from elements of one size. The choice between these three possibilities will be determined by essentially practical matters: for example, the loss in integration time of the first scheme, the cost of the backing structure in the third scheme, etc.

This discussion leads us on to the third major requirement for the array. In order to sample the missing spacings and have the best surface-brightness sensitivity possible, we must use the individual elements as single, beam-switching antennas. Usually these measurements are discarded, since systematic errors

may dominate, and also because they may not be needed if the object is sufficiently small. However, for objects extending over many primary beams, these measurements certainly are required. The most important obstacle to single-dish observations at millimeter wavelengths will probably be the fluctuating emission and absorption of the neutral atmosphere. Beam-switching techniques will be vital to overcome this effect. We should note a related affliction of short-spacing interferometers operating a high frequencies: the sky emission common to both elements of an interferometer can cause a false fringe. Here again, switching techniques are necessary.

In summary, to image wide fields, we need mosaicing, sufficient integration time on short spacings, and a single-dish mode.

(b) High-Resolution Observations. For high-resolution observations, the sensitivity of the array is determined primarily by the total collecting area. Most sources discussed in Chapter II which require, and could reasonably be detected at resolutions smaller than one arcsecond, will be smaller than the field of view of a reasonably sized antenna, say 10 meters. However, at resolutions somewhere between 1.0 and 0.1 arcseconds, depending upon conditions, the seeing at millimeter wavelengths requires that the source be detected on most individual baselines in an atmospheric coherence time. Only if this is so can self-calibration be used to correct the atmospheric phase variations (see Chapter V). Thus sensitivity on each baseline is all-important. The three ways to improve sensitivity within a given time are by increasing either the collecting area or the bandwidth—or both. Low-noise receivers and a site with very little atmospheric extinction are beneficial in all cases.

There may be ways around the self-calibration difficulty that was mentioned in the last paragraph. In the case of very simple, weak sources, it may be possible to correct the effects of the atmosphere by techniques related to optical-speckle imaging.[2] Also, fast switching between sources may allow use of phase calibrators several degrees away, although the calibrator must itself be detected in an atmospheric coherence time. Finally, future refinements in water-vapor radiometry may allow direct correction of the observed phases for variations in atmospheric pathlength. However, with the exception of these possibilities, high-resolution observations demand the highest possible sensitivity on each baseline. We must thus increase one or both of the element size and IF bandwidth. One elegant way to accomplish both these enhancements is to group elements and sum the outputs before correlation. For example, we could configure twenty elements in five clumps of four elements each. By summing the IFs we would increase the sensitivity on each baseline by a factor of four. We could, at the same time, increase the bandwidth for continuum observations on each baseline since more correlator channels would be available, and we would thereby gain as much as another factor of four in sensitivity on each baseline. Thus, at the expense of u-v coverage and the resulting com-

[2]Cornwell, T. J. (1987), "Radio-interferometric imaging of weak objects in conditions of poor phase stability: the relationship between speckle masking and phase closure methods", *Astron. Astrophys.* **180**, 269–274.

plexity of the source which could be mapped, we can gain a factor of 16 in the strength of the source which could be detected. Of course such a mode would place requirements on the bandwidth of the IF signal from each telescope and would require additional delays.

3. Imaging Speed

Fast imaging (i.e., in much less than 12 hours) appears to be a requirement of the array for at least three reasons. First, mosaicing demands that a good image of each sub-field be obtained in a fraction of the total observing time. In some cases, this may be only a few minutes, so the instantaneous (snapshot) imaging of the array must be good.

Second, high-quality data may be obtained only sporadically. The increase in atmospheric attenuation and emission at low elevations will limit the useful range of hour angles, often to much less than 12 hours. Furthermore, atmospheric attenuation and phase stability both can vary on time scales much shorter than 12 hours. Fast imaging will be essential to optimal use of good data that are obtained.

Third, the sheer volume of science which can be done with the MMA dictates a fast instrument so that many projects can be undertaken each year, as in the case of the VLA.

These three requirements strongly favor an instrument with a large number of elements so that many u-v points can be sampled at each time interval.

4. Frequency Coverage

The experiments discussed in Chapter II require frequency coverage in all accessible bands between 30 and 350 GHz. The atmospheric transparency in this region of the spectrum divides it naturally into windows at 30–50 GHz, 70–120 GHz, 140–170 GHz, 200–300 GHz, and a band around 330–360 GHz. Our current understanding of the science possible in these windows sets a rough priority on their importance. The 200–300 GHz band is clearly the most important, followed by the 30–50, 70–120, and 330–360 GHz bands with roughly equal importance, while the 140–170 GHz band currently seems least important.

The significance of the 200–300 GHz band makes clear the necessity of low atmospheric attenuation. This generally means a high site. Unfortunately, most high sites (e.g., above 10,000 feet) tend to be small, which conflicts with the desire for long baselines. Also, moderate size high sites (3 to 15 km in diameter) tend to be in remote areas—in some cases outside U.S. territory. It seems likely that some sort of compromise will be necessary.

5. Spectral Line and Continuum Capabilities

The science described in Chapter II demands that the instrument must be able to do both spectral-line and continuum work. On the one hand, the spectral-line observations require a large number of spectral channels to be calculated for each baseline, which makes heavy demands on the data storage

and post processing. On the other hand, good sensitivity for continuum observations demands the widest possible bandwidths. Both considerations cause the design of the correlator system to be of prime importance.

6. Low Systematic Errors

The cosmological experiments require lower systematic errors than have typically been achieved with existing single dishes. To take advantage of the redundancy in an array any systematic errors should, if possible, be uncorrelated. Although not stressed in Chapter II, low-resolution spectral-line work also will benefit from the the same considerations for detection of broad, weak lines. Careful planning of the optics and minimization of the asymmetric structures in the inner sidelobes of the antennas will be necessary to achieve these goals.

7. Solar Requirements

Observations of active regions on the sun probably require special considerations. Instantaneous fields of view of a few arcminutes, and high time resolution (0.1 seconds), are desired. By under-illuminating some antennas and reducing the bandwidth these goals may be achievable, although at the expense of lower efficiency. The second goal implies either a low duty-cycle of observation, or, preferably, a very high-speed imaging capability, possibly via a special-purpose processor. These capabilities, while necessary for solar observations, will doubtless be useful for many other experiments. The need to shield the telescope optics and receiving equipment from focused solar IR and visible light poses an additional complication for solar observations.

8. Data Handling

The enormous volume of data produced by an MMA made up of several tens of antennas brings with it a need to streamline the data storage and access procedures. If this is not achieved, the MMA astronomer will be in danger of being swept away by a torrent of data. Perhaps a reasonable statement of the goal here is to make the editing, calibration, and analysis of MMA data no more burdensome (for the astronomer) than is presently the case for single-dish millimeter-wave observations.

IV. Practical Design Considerations

1. Overview

The scientific requirements for the MMA outlined in Volume I are in general accord with the outline presented by the Barrett Subcommittee. But the scientific emphases in the Green Bank proceedings have led us to a refinement of the strawman design concept for the MMA. The current design concept can be summarized as follows:

Sensitivity: The received signals at millimeter wavelengths from astronomical sources of either line or continuum emission are weak. Sensitivity is always at a premium, and for this reason the Barrett Subcommittee suggested a substantial total collecting area for the array of between 1000 m^2 and 2000 m^2. (For point source detections a collecting area this large is needed for the array to be competitive with existing international telescopes—1000 m^2 corresponds to a single 36 m aperture, whereas 2000 m^2 is the collecting area of a single 50 m antenna.) For an array of N individual antennas of diameter D, a requisite total collecting area A necessitates the following number of elements:

D (meters)	$A = 1000$ m^2	$A = 2000$ m^2
10	13	26
9	16	32
8	20	40
7	26	52
6	36	71
5	51	102.

The participants in the Green Bank workshop noted that the range and scope of new science available to the MMA expands as the instrument sensitivity is increased. For this reason a total collecting area closer to 2000 m^2 than 1000 m^2 was favored. As can be seen in the table above, this requires several tens of antennas.

In the context of Fourier synthesis, "sensitivity" has more than one definition. Collecting area is one measure; brightness sensitivity is another. The strawman array accommodated the need for brightness sensitivity by allowing for the antennas to be movable among three different array configurations:

(1) a packed array of diameter 90 m;
(2) an array of 300 m extent; and
(3) a 1 km (or greater) array.

The Green Bank workshop participants endorsed the concept of transportable antennas and multi-configuration arrays, so as to tailor the brightness sensitivity of the array to the angular scale of the object being studied.

Wide Field Imaging: The need to image fields many times larger than the primary telescope beam in angular extent was given high scientific priority. One of the primary modes of operation of the MMA will be the formation of

mosaic images from individual snapshots. This places two demands on the instrument design. First, to provide good snapshot images the array must have good, i.e., complete, instantaneous u-v coverage. In addition, the very low spatial frequencies need adequate sampling. In the strawman array a "multi-telescope" platform provided the "zero spacing". But this may not be necessary if the array antennas can also operate in a total-power mode as discussed in Chapter III.

High Frequency Capability: The scientific emphasis placed on measurement of the $J = 3$–2 and $J = 2$–1 lines of CO means that the array design should incorporate provision to image effectively at 345 and 230 GHz. A capability at 345 GHz is a subtle but important specification of the Barrett Committee's general recommendation for good imaging capability at 1 mm. But, because superb atmospheric opacity is so crucial at 345 GHz, observations at this frequency can be made only at a high altitude site. Thus, the array should be located at high altitude.

High Resolution Capability: The Barrett Committee recognized the need to image with 1 arcsecond angular resolution at 115 GHz. The scientific workshop emphasized that 1″ is inadequate for protostellar disks, circumstellar shells, and galactic nuclei; they refined this criterion to 0.″1 at 230 GHz. This implies array baselines as long as 3 km. Only limited need was seen for much longer baselines, but by the above-mentioned considerations dealing with brightness sensitivity, the need for more compact configurations is vital.

These are the principal refinements to the MMA design derived from the Green Bank scientific workshop. Their realization is found in the sections below.

2. Multi-Band Frequency Agility

The need to separate CO opacity and excitation effects from abundance variations brings with it the need for multi-transition observations. Such research is facilitated by provision for simultaneous multi-band observations. While not mentioned explicitly by the Barrett Subcommittee, multi-band observations are an important scientific criterion for the MMA design to accommodate. We here propose three frequency bands for MMA operation, in recognition of the importance of observations of CO and its isotopomers. An additional band centered on the 9 mm atmospheric window is to be provided principally for sensitive continuum observations and to tie-in the higher frequency MMA observations with the lower frequency VLA and VLBA spectral information. The four proposed MMA frequency bands are:

9 mm:	36–48 GHz
3 mm:	70–115 GHz
1 mm:	200–270 GHz
0.8 mm:	270–350 GHz.

Achieving these wide tuning ranges with single receivers may not be possible, so that compromises will likely be necessary. The frequency ranges shown here

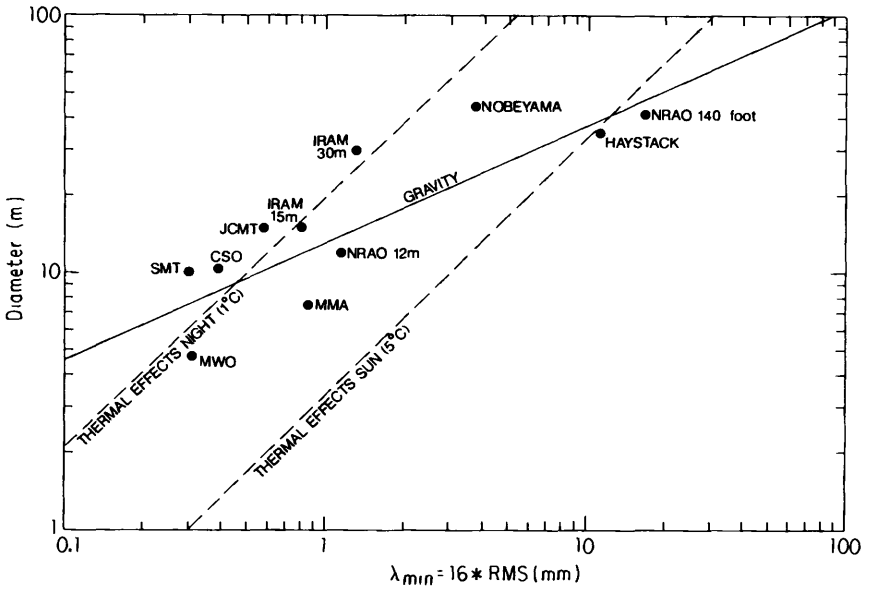


Figure 1. A plot of antenna diameter vs. minimum wavelength at which high-precision observations are attainable.

should be regarded perhaps as goals rather than requirements. The receiver systems are described in Chapter VI.

3. Antennas

The general antenna specifications call for a transportable antenna that operates well at frequencies at least as high as 350 GHz. In order to provide precision operation at the highest frequency the combination of

(1) surface panel irregularities,
(2) panel setting,
(3) backup structure deflections,
(4) gravitational deformations, and
(5) thermal flexure

should be less than (0.85 mm)/16 = 53 μm. Further, the antenna pointing accuracy should exceed one-tenth the full-width to half-power (FWHP) beamwidth, or less than about 3 arcseconds.

The difficulties one may expect in meeting the surface tolerance specification are summarized in Figure 1. This plot of antenna diameter versus minimum wavelength of precision observation (defined as 16 times the cumulative surface error) illustrates the deleterious effect of thermal imbalance and gravitational deformations. For example, when the temperature differential between two members in the antenna exceeds 5 degrees C, as is common when the antenna is exposed to full sunlight, then the antenna performance will be limited to those areas of the plot below and to the right of the 5° C dashed line. The limitations imposed by gravity are similarly indicated.

Neither the thermal nor the gravitational effects are insuperable; with several of the antennas shown on this figure have been made the appropriate

design allowances. For example, the Nobeyama, IRAM and SMT telescopes incorporate homology principles in their design, which allows them to "exceed" the gravity limit. Both SMT and IRAM (30 m) also exceed the 1° C thermal limit, but both do so by means of an enclosure—a dome in the case of the SMT, and an enclosed back structure in the case of the IRAM telescope. The IRAM 15 m telescope employs carbon fiber materials. All such measures are achieved at the expense of increased cost.

Figure 1 shows that it should be possible to build an inexpensive MMA antenna for operation at 350 GHz. If the aperture is less than 11 m it need not be homologous because it then does not exceed the gravity limit. In full sunlight its performance may be compromised somewhat if temperature differences as large as 5° C are allowed to persist in the structure, but if these differences can be held to less than 3° C then we can expect precision operation to the highest frequency. A simple antenna of approximately 8.5 m diameter or less which needs neither an enclosure nor sophisticated carbon fiber construction meets these specifications.

To be specific, let us consider a 7.5 m antenna. A straightforward modification of the VLBA telescope design to allow for transportability produces a 7.5 m antenna design with the following surface error and pointing error budgets:

Surface r.m.s. error due to:		
Gravity		0.013 mm
Surface plate		0.025
Plate setting		0.025
Wind, etc.		0.021
	RSS	0.043 mm
Pointing error due to:		
Position indicator		1.5 arcsec
Acceleration		0.2
Limit cycle		0.4
Position update		0.4
Mechanical rotation		2.0
Wind, etc.		1.5
	RSS	3.0 arcsec.

The surface plate accuracy can be achieved either with TIW cast aluminum or with ESSCO bonded plates.

In the final antenna design we will have to pay particular attention to maximizing the aperture efficiency. For a symmetric antenna with a geometric blockage of 4% one could hope to obtain an aperture efficiency as high as 92% with two shaped surfaces (*viz.*, primary and secondary). A still higher efficiency would require an asymmetric design; the trade-off then would become one of efficiency against optical design (to facilitate multi-band observations) and cost.

4. Array Sensitivity and Speed

Antennas: Number and Size. The optimum number of individual antennas in the array and the optimum antenna diameter can be estimated from sensitivity and cost considerations. The sensitivity calculation outlined in Chapter I of Volume I demonstrates that for point source studies (Eq. 8) the time needed to reach a given sensitivity is proportional to $N^{-2}D^{-4}$, so that the optimum array employs antennas of the largest practical diameter. On the other hand, if the goal is to image a field with an angular diameter equal to or greater than the primary beamwidth, then the speed of the array is proportional to N^2D^2, the square of the product of antenna diameter and number of elements (Eq. 9). An array of many small elements is as fast as one composed of fewer large antennas—as long as ND is the same for both arrays. In either case, sufficiently many antennas are needed to provide good instantaneous u-v coverage, while the cost of the antennas will, of course, increase with their number. Let us estimate the antenna diameter by choosing that diameter which, for given imaging speed $(ND)^2$, minimizes the total cost of the array. We can then use the Barrett Subcommittee guidelines, or other estimates of the total collecting area needed, to compute the number of antennas in the array.

The cost of each of the constructional components of the MMA (e.g., antennas, receivers, computers, buildings, etc.) can be parametrized as a fixed cost, plus a cost term which depends on the number of antennas in the array, plus a term that depends on the square of the number of antennas. Skipping ahead to the MMA budget discussion, Chapter VII, we see that the total construction cost estimate of the MMA can be expressed as

$$\text{cost(k\$)} = 6860 + \left(541 + 890\lambda^{-0.7}(D/10)^{\alpha}\right)N + 4.5N^2\,. \qquad (1)$$

Here λ is measured in millimeters, and $\lambda/16$ is the desired surface error of each of the N antennas of diameter D meters. The cost of each antenna is estimated to be

$$A(\text{k\$}) = 890\lambda^{-0.7}(D/10)^{\alpha}\,, \qquad (2)$$

which is normalized to 1000k\$ for a 10 m antenna with $\lambda/16 = 53\ \mu$m (corresponding to precision operation at 350 GHz). The exact value of the exponent α which relates cost to antenna diameter is not known, but it would be very surprising if α were less than two or greater than three. (Scaling with $\alpha = 2$ corresponds to enlarging or contracting the antenna as a 2-dimensional surface, whereas $\alpha = 3$ is equivalent to thinking of the antenna as a fully 3-dimensional expansion or contraction).

If we now fix the speed of the array (i.e., set ND = constant) and take the partial derivative of Equation 1 with respect to antenna diameter we can determine the value of the antenna diameter which minimizes the cost of the array. The equation to be solved for that diameter is

$$D^{\alpha+1} - \frac{0.608 \cdot 10^{\alpha}\lambda^{+0.7}D}{\alpha - 1} - \frac{0.010 \cdot 10^{\alpha}\lambda^{+0.7}ND}{\alpha - 1} = 0\,. \qquad (3)$$

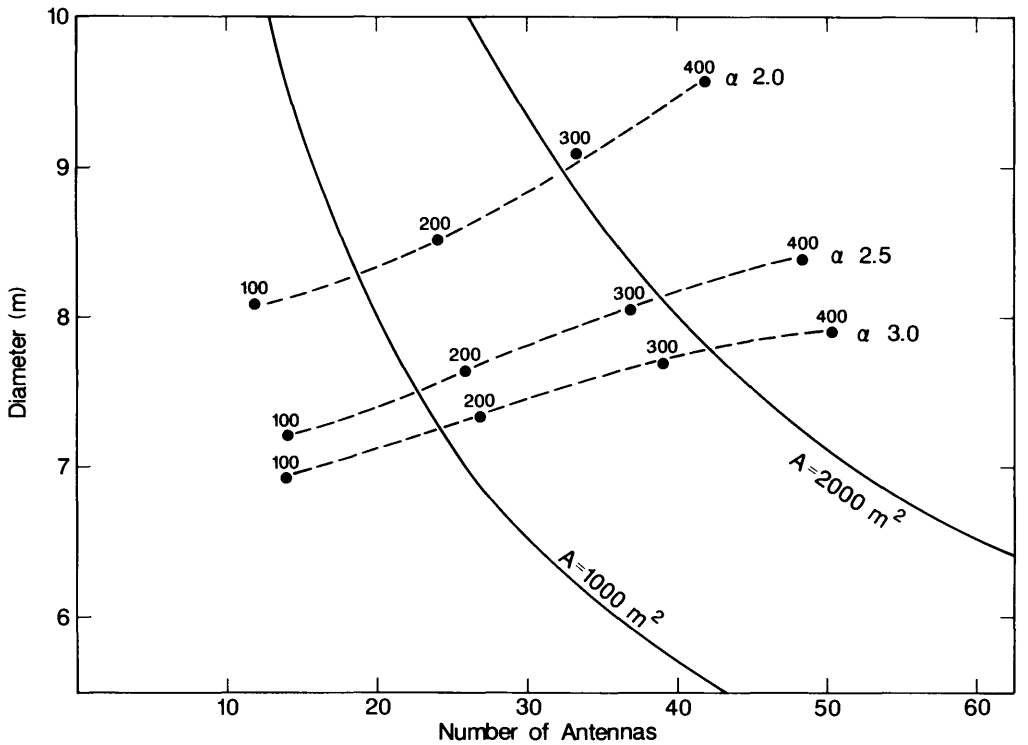


Figure 2. The minimum cost curves for the antenna diameter and number of elements in the MMA (assuming $\lambda = 0.85$ mm), for three scaling exponents, α, of antenna cost with diameter. The dashed lines correspond to loci of minimum total array cost as a function of the product ND. (N^2D^2 is the array speed.) Points corresponding to $ND = 100$, 200, 300 and 400 are noted along each curve. The solid curves are the D vs. N curves which give total collecting areas of 1000 m^2 and 2000 m^2, respectively.

The solutions can be parametrized by the ND product and α; for each selected value of α, the minimum cost antenna diameter $D_{\min}$ is a function of the constant ND, and thus from each $D_{\min}$ we obtain N, the number of antennas. Solutions to Equation 3 are shown by dashed lines in Figure 2 for $\alpha = 2.0$, 2.5, and 3.0. If we can establish the antenna cost scaling with antenna diameter and can decide on a desired speed of the MMA for imaging extended objects, then Figure 2 will provide both the optimum antenna diameter and the appropriate number of array elements.

At this point it is well to introduce one other consideration that figured importantly in the Barrett Subcommittee discussions, *viz.*, total collecting area. That committee recommended 1000–2000 m^2 total collecting area in the array (equivalent to a 35–50 m single dish). The Green Bank workshop discussions emphasized the need for collecting area and favored values closer to 2000 m^2 (or even greater). The boundaries of the Barrett Subcommittee recommendations are shown by solid lines in Figure 2. If we respect the Barrett limits and choose a midrange value for α, $\alpha = 2.5$, then the minimum cost MMA has anywhere from twenty-three 7.5 m antennas ($ND = 170$) to thirty-nine 8.1 m antennas ($ND = 316$). Referring to Figure 1 we see that 7.5–8.1 m antennas

for precision operation at 350 GHz can indeed be made inexpensively, because they need neither homologous design nor extreme thermal stability.

For an instrument, such as the MMA, designed to provide significant new scientific capabilities for studies of compact objects and for imaging extended fields, both total collecting area $\pi ND^2/4$ and imaging speed N^2D^2 are important. An array of approximately forty antennas of 7.5 m diameter ($ND = 300$, collecting area 1767 m^2) compromises neither of these capabilities, and yet appears near the minimum cost curves. We therefore adopt, for the remainder of this discussion, these values of N and D.

The equations describing the sensitivity of the array that were presented in Volume I remain valid. However, the high altitude site and improved antenna illumination should improve the numbers quoted there for two reasons. First, the antenna efficiency using two shaped surfaces should provide an aperture efficiency of 0.8–0.9. Second, the high site implies a better sensitivity when conditions are good, since the attenuation and the contribution to the system temperature by the sky will be less than was assumed in the strawman concept. Thus, for example, an eight hour synthesis in the continuum with "natural" weighting using both polarizations at 230 GHz should be able to reach an r.m.s. noise level of about 28 μJy under typical conditions. In the same time with a spectral resolution of 10 km/s the smallest array will reach an r.m.s. of less than 1 mK. Making the same assumptions about the receiver temperature as in Volume I, the average sensitivity of the array should be a factor or two better than in the 1985 array concept. The overall speed for the present design to image a given piece of sky as large as or larger than the antenna beam, to a given sensitivity level, should be about eight times faster than with the earlier array concept.

5. Array Configurations

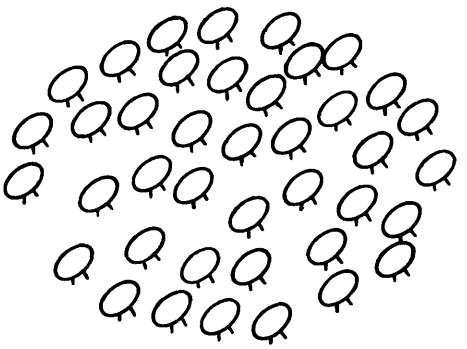
As discussed in Volume I, many geometric configurations have been considered for the array. The best configurations for uniform instantaneous sampling of the u-v plane are circular arrays, with the antennas arranged quasi-randomly around the circle. For observations away from the zenith, we expect this concept to generalize to an elliptical configuration, with the major axis of the ellipse about a factor of two larger than the minor axis, and in a north–south orientation. For the smallest configuration, the array more closely resembles a filled circle or ellipse with about fifty percent of the surface area covered. Another type of "good" configuration, which is optimum for cases where greater sensitivity is needed for larger size scale structures, is the VLA-like Y. Antennas are then arranged so that the nth antenna on each arm has a power law dependence like $n^{1.7}$ on its distance from the center. Practical considerations, such as simple paths for transportation and communication with antennas and a significant reduction in the number of required antenna stations, tend to argue for Y-like configurations. Other practical considerations dictate that the most compact arrays will be a highly packed circle/ellipse as shown in Figure 3a.

Implications of a Changed Concept: The concept of the MMA has changed from the "strawman" arrays presented in Volume I. The current emphasis—a high site allowing configurations $\leq$ 3 km, with twice the number of antennas, each of which is 3/4 the size assumed in the early configuration studies—forces some changes in the choice of array configurations. The most significant change may be that the topography of the best site will dictate modest departures from the symmetry of theoretically ideal configurations. Another consideration, that comes in part from the potential for better high frequency work, in part from the strong dependence of system temperatures on air mass, and in part from the need to make images that are many antenna beams in area in short periods when the atmosphere is ideal, is the need for excellent imaging characteristics during short observations. The last consideration, together with a possible need to reduce the number of antennas stations and communication paths, leads to a generalization of the three-arm radial configurations that are VLA-like Y's, to five- or seven-arm configurations that we shall refer to as star configurations.

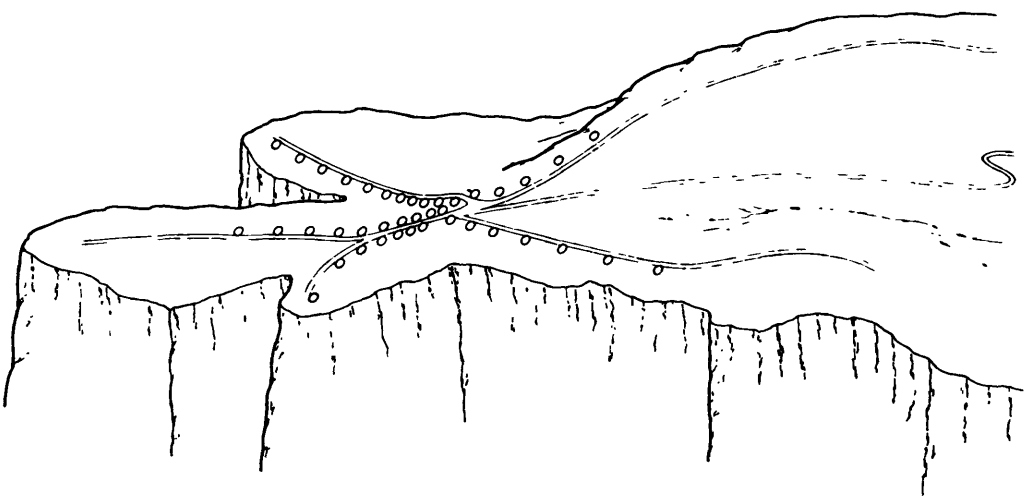
n-Pointed Star Configurations: The properties of VLA-like Y's are due to three principles: the radial power law dependence of the locations of antennas on each of the three spokes of the Y; the fact that a VLA-like power law allows one to reduce the total number of stations when configurations on different size scales are needed; and the simplification of antenna transportation and control/communications systems when one can align things along a small number of straight paths. However, since the studies discussed in Volume I, it has been realized that the 120° angle between arms in a three-armed configuration leads to gaps in the u-v coverage that cannot be filled without making observations that are longer than desirable. These concepts, and the problem of packing 40 antennas into relatively compact 300 m and 1 km arrays with the basic properties of Y's, make n-armed configurations, where n is an odd number like 5 or 7, the most likely alternative to circular/elliptical configurations for the MMA. We will call these *n-pointed star configurations.* The 5-pointed star array can afford hybrid configurations for observations of objects at southern declinations, in an analogous manner to the hybrid arrays of the VLA where the southeast and southwest arms are in one configuration while the north arm has a size scale 3.28 times larger. The hybrid 5-pointed star array would have four arms in one configuration while the fifth arm, pointing to the north or to the south, is larger by a factor of 3.28. At the price of increasing the number of antenna transportation and communication paths from three to five, one achieves all of the advantages of the Y-like arrays discussed in Volume I, but with both significant improvement in instantaneous imaging properties and ease in placing a larger number of antennas into more compact configurations.

Further investigations into theoretically ideal arrays will continue; however, as we will now discuss, future approaches to configuration studies may be dominated by the topography of the potential high altitude sites.

Topography-Limited Arrays: The characteristics of the high altitude site eventually chosen for the MMA may have a significant effect upon the larger



(a)



(b)

Figure 3. Examples of MMA configurations: (a) the most compact configuration for a fixed number of antennas, a packed circle/ellipse; (b) a schematic mountain-top array where topography significantly affects the larger 3 km, and possibly the 1 km, arrays.

configurations. Without going into any of the specific topography-limited configurations that are being investigated, certain principles are clear. Any plateau-like site with capabilities to put antennas anywhere inside an area of size $\approx$ 3 km will not require topography-dependent configuration changes. However, while it may be relatively easy to select a good site for symmetrical 90 and 300 m configurations, topography may significantly affect larger configurations. Preliminary studies, and obvious principles, dictate that even a topography-limited site must have adjacent areas or ridges to which antennas can be moved in configurations that must be roughly equal in two horizontal dimensions. It is also obvious, because of the need to improve imaging at southern declinations by north–south elongation of arrays, that longer topographical dimensions oriented north–south are desirable. In Figure 3b we show a schematic sketch of antennas in a probably acceptable large configuration. The ability to tailor the configurations to the unique topography of each site will be a major factor in evaluating the acceptability of sites.

6. Measurement of Low Spatial Frequencies

In the strawman concept, measurement of the most extended emission was left to a specially designed instrument, the multi-telescope (MT) array. This rotating platform of very small antennas measured all spatial frequencies intermediate between the size of the MT elements and approximately 2.5 times the diameter of the larger array antennas. But it did so at the expense of additional electronics and correlation hardware. It may not be needed (see the discussion in Chapter V). Instead, low spatial frequencies may be measured by the individual array elements themselves. Two unusual provisions need to be incorporated into the antenna design. First, each antenna should be capable of total-power observations. Second, the feed illumination pattern must be nearly uniform across the aperture so that measurements at the cross-over point—projected separations equal to the dish diameter—are made with high signal-to-noise.

While this new approach is attractive theoretically, many practical obstacles may arise. For example, uncertainties or variations in the illumination patterns lead directly to errors in the reconstructed image. Thus in the design of the antennas and the feed system, there must be a compromise between the illumination pattern stability, desired for imaging of bright objects, and uniformity, which is important for adequate signal-to-noise when imaging weaker objects. Pointing errors are equivalent to errors in the illumination pattern, and will have similar effects on the reconstructed images. Both cross-talk and correlated emission from the troposphere may also be limiting factors under some circumstances. It is important to realize that all these problems would also afflict the multi-telescope design. For example, the multi-telescope array would be subject to the level of pointing errors achievable with a backing structure of size up to 20 m in diameter. Furthermore, flexing of the backing structure would lead to poor collimation.

Work to examine these questions is proceeding along three paths: analyt-

ical investigations, computer simulations, and experimentation with existing centimeter- and millimeter-wavelength arrays.

7. Site Selection and Testing

1. Site criteria and evaluation. The scientific goals of the array require that it be located on a high, dry site offering the possibility of baselines of at least 3 km in length. The atmospheric qualities are driven by the desire to minimize the increase in system temperature contributed by emission from the sky and to minimize the loss of signal caused by atmospheric attenuation. Both of these effects will be serious in the 1 mm band unless the site is at high altitude. In addition, it is hoped that the array can be operated at submillimeter wavelengths for some reasonable part of the time. These goals require that the total precipitable water vapor over the site be frequently 2 mm or less, and suggest that the site be at least 9,000 ft in elevation. Even higher sites are preferred so as to minimize the absorption in atmospheric oxygen. The physical size is derived from the goal of 0.1 arcsecond resolution at 230 GHz.

There are very few sites which meet these criteria. In a survey of maps of the western United States south of latitude 38 degrees, only one site was found having size 3 km on a side and having any type of existing access road. This site, the Aquarius Plateau in south-central Utah, is at 11,000 ft and would allow baselines of up to 15 km. To extend the list of possible sites, one or another of the criteria must be eased.

There is an interesting possibility in Colorado, the Grand Mesa near Grand Junction. It is only slightly further north, at latitude 39 degrees, is at 10,000 feet elevation, and would allow baselines of 3 km or greater. The VLA site itself offers considerable advantages from the standpoint of logistics and support and offers baseline lengths in excess of 30 km. It is however at only 7,000 feet elevation. A site near, but not at, the summit of Mauna Kea, Hawaii would have very good atmospheric characteristics. However, even if elevations as low as 12,000 feet are considered, it is difficult to find a site on this mountain that would permit baselines in excess of 1 km. A site in the Magdalena mountains near Socorro, NM lies at 10,500 feet, is close enough to be supported from the existing NRAO facilities, and offers the possibility of baselines of up to 3.5 km. Finally, there is a possible site at Sacramento Peak, NM with a level circular area of diameter $\approx$ 2.5 km at an elevation of 9,200 feet. We have no information on the possible availability of these sites for the MMA.

It appears that one or another of these sites is suitable for the millimeter array, though with a compromise either in the amount of observing time available at the highest frequencies or in the achievable resolution. It is desirable however to make a more detailed search of the topographical maps of the U.S. Southwest to see if there are any other sites that could be used with less sacrifice to the design goals of the array. Therefore, the search will concentrate on identifying possible sites having elevation in excess of 9,000 feet, a circular area of a least 1 km in diameter, and the capability of baselines in at least some directions of 3 km. The first condition is set by the requirement to gain

a significant improvement in elevation over that afforded by the VLA site; the second is an attempt to improve over the conditions obtaining at Mauna Kea and the Magdalena Mountains; and the third is a statement of the minimum consistent with the scientific goals set for the array.

All sites found in the re-survey will be evaluated on the basis of environmental conditions such as wind and snow, and other criteria like ease of access. The most severe test will be whether they offer a significant gain in performance over one of the six possibilities that have already been identified. Any site passing all of these tests will be added to the list of sites for which more detailed studies are being conducted.

The selection of the best site from the list of best possibilities will be on the basis of a number of criteria. Cloud-cover data will be developed from the currently existing satellite data bases and will be given high weight because of the importance of cloud-free skies. The sites will be monitored with the 225 GHz radiometer as described below, and those data will give estimates of atmospheric transparency and stability. Since these factors will be the principal limitation to the ultimate performance of the array, they also will be weighted very strongly. Each site will have at least a crude array plan, so that the coverage in the Fourier transform plane can be evaluated. This will allow a weight based on resolution and quality of synthesized beam to be developed. Finally, the remoteness of the site, the ease of access to it, and the cost of developing it will be combined to give the last weighting factor.

2. Atmospheric transparency tests. One of the most critical aspects is the atmospheric transparency and its stability. Unfortunately, it is also the one for which data are the most difficult to obtain. A limited amount of low resolution data (both low time resolution and low spatial resolution) is available from satellite work, and additional material is available for those sites that are close to radiosonde launch facilities. Some time ago, however, it was evident that the existing data were not sufficient for our needs, and the NRAO therefore began development of a site-testing radiometer. This device has been completed, and observations at two of the potential sites have begun.

The site-testing radiometer operates in the continuum at 225 GHz, with a bandwidth of 1 GHz. The calibration of the temperature scale is made with reference to two internal thermal loads. For measurement of the atmospheric transmission an observation is made once every ten minutes, when a rotating mirror is stepped so as to direct the beam in a series of steps between the zenith and the horizon. Opacity is deduced from the calibrated "tipping curve". A rough check on the procedure is obtained from the measurement of the zenith temperature itself; the consistency between the measured zenith temperature and that predicted from the inferred value of the zenith opacity has been excellent.

Figure 4 shows an example of the type of data now being obtained with these radiometers. Plotted is the fraction of the total number of hours in the period November 10, 1986 to April 15, 1987 for which the opacity as measured at 225 GHz at the Magdalena Mountain site (denoted as South Baldy on

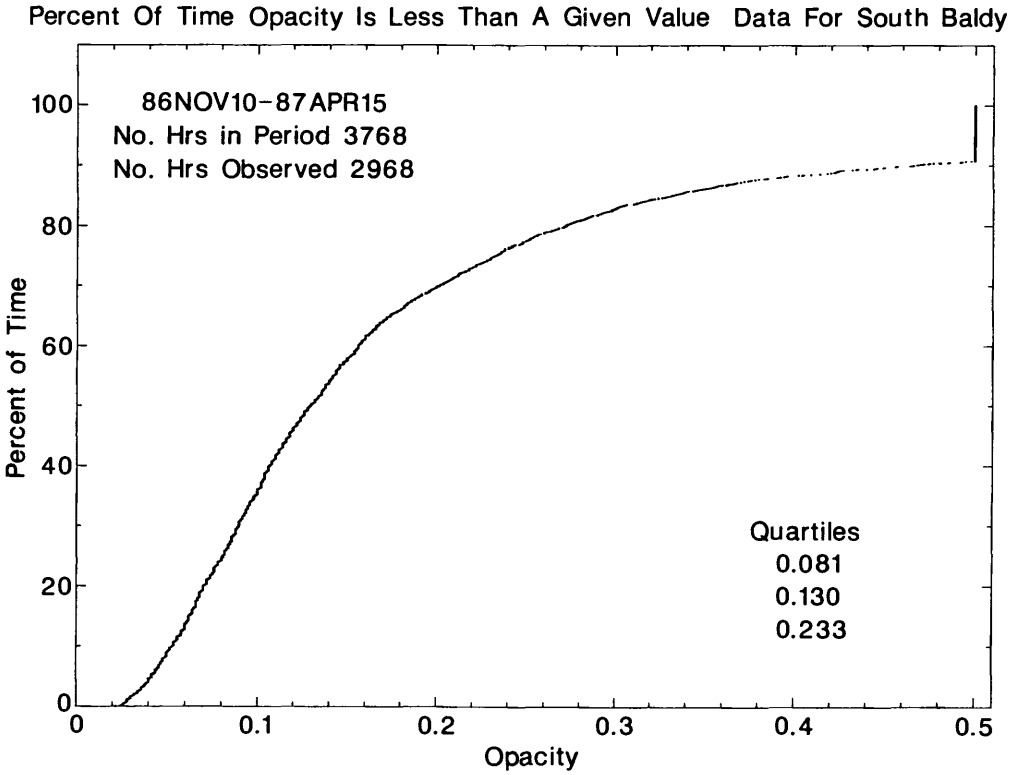


Figure 4. Water-vapor radiometer data obtained at the Magdalena Mountain site during the period between November 1986 and April 1987.

the plot) has a value less than a given amount. Observations were made for about 79% of the time. The quartile values of the opacity are 0.08, 0.13, and 0.23 as shown. For reference, these opacities correspond approximately to 1.4, 2.2 and 4.1 mm of precipitable water vapor in the (zenith) path through the atmosphere. It is planned that similar data will be obtained for at least a year at each of the most promising sites.

Another way of looking at the same data is to ask how long the periods of favorable atmospheric conditions persist. This is an important requirement for a synthesis telescope, which often observes a source for a duration of six to twelve hours in order to build up a representative picture of the source. If, as an illustration, a "good period" is defined to be one in which the opacity at 225 GHz stays at 0.1 or less for at least six hours, the data from Figure 4 can be summarized in terms of good periods, as presented in Table 1.

These radiometers can also in principle be used to assess the stability of the atmosphere using a technique described by Smoot *et al.*[1] Periodically the radiometer output will be sampled rapidly, perhaps every 10 seconds, while it

[1]Smoot, G. F., Levin, S. M., Kogut, A., De Amici, G., and Whitebsky, C. (1987), "Atmospheric emission variation measurements at 3, 0.9, and 0.33 cm wavelengths," *Radio Science* **22**, 521–528.

Table 1. Periods of low opacity at South Baldy. These data were collected over the period November 10, 1986 through April 15, 1987. A "good period" has opacity less than 0.10 for at least six hours; it can encompass time spans of up to three hours when the opacity is higher; however, the worst opacity must be less than 0.15 during a "good period". In all, there were 40 good periods during this time. The total number of hours involved in these periods is 980. An average period lasts 24 hours and has a median opacity of 0.069.

Run	Start	Stop	Duration	— Opacity —		
	y/m/ d/hr	y/m/ d/hr	(hours)	Median	Average	Minimum
1	86/11/10/ 11	86/11/12/ 4	42	0.068	0.071	0.038
2	86/11/23/ 22	86/11/24/ 4	7	0.072	0.075	0.062
3	86/11/27/ 4	86/11/28/ 21	42	0.076	0.074	0.037
4	86/11/30/ 1	86/11/30/ 13	13	0.067	0.068	0.036
5	86/11/30/ 20	86/12/ 2/ 19	48	0.060	0.062	0.037
6	86/12/ 3/ 9	86/12/ 3/ 20	12	0.090	0.089	0.079
7	86/12/11/ 22	86/12/15/ 17	92	0.055	0.059	0.025
8	86/12/19/ 1	86/12/19/ 6	6	0.083	0.086	0.079
9	86/12/23/ 9	86/12/23/ 14	6	0.046	0.046	0.039
10	86/12/26/ 2	86/12/27/ 3	26	0.064	0.068	0.044
11	86/12/29/ 11	86/12/31/ 21	59	0.055	0.054	0.028
12	87/ 1/ 2/ 12	87/ 1/ 3/ 7	20	0.070	0.068	0.031
13	87/ 1/ 4/ 8	87/ 1/ 4/ 16	9	0.090	0.091	0.086
14	87/ 1/ 8/ 21	87/ 1/ 9/ 7	11	0.053	0.054	0.035
15	87/ 1/20/ 18	87/ 1/23/ 5	60	0.080	0.075	0.036
16	87/ 1/24/ 11	87/ 1/24/ 20	10	0.079	0.079	0.066
17	87/ 1/25/ 15	87/ 1/27/ 8	42	0.067	0.069	0.050
18	87/ 2/ 1/ 5	87/ 2/ 1/ 14	10	0.088	0.087	0.078
19	87/ 2/ 2/ 2	87/ 2/ 3/ 1	24	0.080	0.078	0.058
20	87/ 2/ 3/ 6	87/ 2/ 3/ 11	6	0.098	0.094	0.080
21	87/ 2/ 5/ 16	87/ 2/ 6/ 7	16	0.052	0.053	0.044
22	87/ 2/21/ 15	87/ 2/23/ 5	39	0.047	0.053	0.034
23	87/ 2/26/ 21	87/ 2/28/ 3	31	0.060	0.061	0.035
24	87/ 2/28/ 7	87/ 3/ 1/ 21	39	0.082	0.076	0.034
25	87/ 3/ 3/ 9	87/ 3/ 6/ 10	74	0.042	0.049	0.024
26	87/ 3/ 6/ 13	87/ 3/ 6/ 21	9	0.054	0.060	0.044
27	87/ 3/11/ 2	87/ 3/11/ 12	11	0.076	0.078	0.059
28	87/ 3/12/ 2	87/ 3/12/ 16	15	0.076	0.080	0.060
29	87/ 3/13/ 3	87/ 3/13/ 8	6	0.090	0.090	0.082
30	87/ 3/13/ 20	87/ 3/14/ 12	17	0.064	0.070	0.050
31	87/ 3/15/ 5	87/ 3/15/ 10	6	0.082	0.084	0.072
32	87/ 3/17/ 22	87/ 3/18/ 11	14	0.085	0.085	0.065
33	87/ 3/18/ 21	87/ 3/19/ 12	16	0.075	0.078	0.064
34	87/ 3/23/ 9	87/ 3/23/ 16	8	0.049	0.051	0.036
35	87/ 3/26/ 19	87/ 3/27/ 7	13	0.068	0.065	0.022
36	87/ 3/28/ 6	87/ 3/28/ 11	6	0.061	0.068	0.054
37	87/ 3/30/ 9	87/ 3/31/ 17	33	0.069	0.067	0.040
38	87/ 4/ 2/ 16	87/ 4/ 3/ 12	21	0.060	0.059	0.026
39	87/ 4/12/ 20	87/ 4/14/ 15	44	0.054	0.056	0.027
40	87/ 4/14/ 21	87/ 4/15/ 13	17	0.068	0.071	0.053

is looking at the zenith. If an entire hour is devoted once a day to this it will be possible from the Allan variances to infer the fractional fluctuation in the total pathlength through the atmosphere over timescales of between 10 seconds

and 10 minutes. This feature will be incorporated into the radiometer routine soon, and an index of atmospheric stability will be generated for each of the tested sites.

The quantitative testing of the VLA site and of the Magdalena Mountain site started late in 1986. Observations at Mauna Kea will be undertaken in 1988. Measurements at Grand Mesa and other sites that are identified in the re-survey will begin shortly thereafter. The goal is to have the information necessary for the selection of the array site assembled by the end of 1989.

8. Summary: MMA Design Concept

Antennas:	40
	7.5 m diameter shaped reflectors
	Total surface error < 0.043 mm r.m.s.
	Transportable design
	Beam efficiency > 80 percent
	Total collecting area 1767 m^2
Configurations:	90 m (packed circle/ellipse)
	300 m (circle/ellipse, star, or other array)
	1000 m (circle/ellipse, star, or other array)
	3000 m (circle/ellipse, star, or other array)
Frequency Bands:	36–48 GHz
(desired)	70–115 GHz
	200–270 GHz
	270–350 GHz
Angular Resolution:	0.1 arcsecond at 230 GHz

V. Data Processing for the MMA

1. Overview

The MMA will be capable of imaging objects ranging in size from degrees to milli-arcseconds. Such high-resolution, wide-field imaging—with a large amount of collecting area—requires a combination of two different techniques of data analysis:

- First, the incomplete sampling of spatial information forces the use of *deconvolution* algorithms to estimate the missing data. Although such deconvolution algorithms have been in use in radio interferometry for a number of years, and so are relatively mature and well-understood, the scanning of the array, which is vital for wide-field imaging, requires special deconvolution algorithms to correct for the power patterns of the antennas. These recently developed *mosaicing* algorithms, which form a deconvolved image from a collection of data taken with different pointings, are now in use at the VLA.

- Second, both fluctuations in the Earth's atmosphere and pointing errors of the antennas corrupt the calibration of the data. In existing radio interferometers, these effects are corrected by a technique, known as *self-calibration*, in which the unknown calibration is estimated in the deconvolution. At the VLA, images very high in quality and dynamic range are routinely obtained by self-calibration. For the MMA, self-calibration will be complicated by three factors: the weakness of the emission from typical objects, the increased influence of the atmosphere and of pointing errors (both of which scale roughly as frequency), and the differing antenna sizes.

In the MMA, at least one and perhaps both of these techniques—deconvolution and self-calibration—will be necessary in most forms of observation. An additional complication is that variants for spectral line imaging will also be required.

In this chapter, we discuss these types of data analysis and their possible extensions, together with their impact upon the computing requirements for the MMA.

2. Deconvolution

Interferometric arrays can be designed to "almost" completely sample all spatial frequencies up to a limit determined by the maximum spacing, in increments related to the element size. Thus, apart from the very shortest spacings, for which shadowing occurs, all the information relevant to an object can be collected. In recent years, it has been recognized that this classical sampling is far too conservative, and that by exploiting some well-established properties of the radio sky, sparser sampling is acceptable. The CLEAN algorithm[1]

[1]Högbom, J. (1974), "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astrophys. J. Suppl. Ser.* **15**, 417–426.

can correct for sparse sampling by requiring that, roughly speaking, the deconvolved image be mainly empty, whereas another algorithm, the Maximum Entropy Method (MEM) enforces smoothness and positivity.[2] Both algorithms are now extensively used in radio interferometry. One major obstacle in their application to the MMA is the very small field of view of the elements in the current array design. We consider the impact of this in the next section.

3. Mosaicing

In Chapters III and IV, we saw how the problem of measuring size scales ranging from degrees down to arcseconds, with good sensitivity, can be solved. The whole array scans the source, much as single dish would, and the resultant data are combined to form a high-resolution image of the entire object.

The mosaicing problem splits into two cases, depending upon the signal-to-noise ratio of the observations. For very weak objects, the sidelobes are unimportant, and so no deconvolution is necessary. Only appropriate weighting of the various data before addition is required: the visibility data for each pointing center are transformed into images, and then the collection is assembled into a mosaic by adding all images with a spatially varying weight dependent upon the noise level. The processing for this weak-source mosaicing is thus very simple and relatively inexpensive.

For strong sources, deconvolution is vital. An obvious approach to strong-source mosaicing is to deconvolve all the data separately, and then correct for the individual reception patterns, in a similar fashion to weak-source mosaicing. The images are therefore combined *after* deconvolution. This will be successful to some degree but it neglects the nonlinearity of the deconvolution.[3] Algorithms such as CLEAN and the Maximum Entropy Method are inherently nonlinear, and, in general, produce superior results from greater quantities of data. Thus, since the reception patterns cannot be corrected on the raw visibility data, the next best approach is to include knowledge of the reception patterns in the deconvolution itself. Both the cross-correlation data and the total-power measurements from the interferometer elements can be treated simultaneously in the deconvolution. One obvious virtue is that where different measurements contain roughly the same information, consistency checks can be performed.

To test this philosophy, one implementation based upon the Maximum Entropy Method[4,5] has been developed.[6] In the application of the MEM to a simple interferometer array, the deconvolved image is found by maximizing its

[2]Cornwell, T. J. (1986), "Deconvolution", in *Course Notes from an NRAO Summer School Held in Socorro, New Mexico August 5–9, 1985*, Eds. R. A. Perley, F. R. Schwab, and A. H. Bridle, (NRAO: Green Bank, WV).

[3]*Cf.* Cornwell, T. J. (1987), "A comparison between a mosaiced VLA image and a conventional Penticton image", NRAO Millimeter Array Memo. No. 43.

[4]Gull, S. F. and Daniell, G. (1978), "Image reconstruction from noisy and incomplete data", *Nature* **272**, 686–690.

[5]Cornwell, T. J. and Evans, K. F. (1985), "A simple maximum entropy deconvolution algorithm", *Astron. Astrophys.* **143**, 77–83.

entropy subject to the data constraints, which in this case are the visibility measurements. There now exist algorithms to perform this optimization in about 20 to 40 data comparisons: the main work in such an algorithm is to compare the current iterate to the observed data, involving Fourier transformation to the u-v plane, subtraction from the visibility data, and inverse transformation of the residuals (see, e.g., Cornwell and Evans[5] for further details). The resulting residuals are used to steer the iterate towards the optimum image. The reception pattern can easily be corrected by accounting for it suitably in the data comparison stage: the current iterate is tapered with the reception pattern, the result Fourier transformed, subtracted from the visibility data, the residuals inverse-transformed and then tapered by the reception pattern again. The extension to many pointing centers is then trivial: one simply performs a data comparison for each pointing center and sums the residuals. Single dish data can be accommodated with the appropriate data comparison. Although only an MEM implementation exists at present, a CLEAN-based approach also should be possible, and will be preferred for mosaicing images of relatively simple objects for which CLEAN is much more efficient.[6]

An example of the application of this mosaicing method to VLA observations is given in Figure 1, where a VLA image of the source Simeis 57 is shown. The object is about two degrees across and, at the observing wavelength of $\lambda = 20$ cm, required 22 pointings, each of which was sensitive to a field of view of about one degree.

The computing requirements of both weak- and strong-source mosaicing observations are described in Appendix 1. Strong-source mosaicing dominates the processing. For an observation of a large field of view, requiring 100 pointings of our 7.5 m elements, with the resolution of the 300 m configuration, the reduction of data for 512 spectral channels can keep up with real time if performed on a machine with six times the capacity of the CONVEX C-1 with at least 1 Gbyte of fast memory. As a check we also considered perhaps the most extreme mosaicing observation: that of a 30 arcminute cloud with arcsecond resolution at 230 GHz, and 512 channels, would require only seven CONVEX C-1 equivalents to keep up with real time. About 10^4 pointings would be required, forcing about fourteen days of observing, and the final image would be a cube of size $4096 \times 4096 \times 512$.

Again, we should emphasize that these are very rough estimates, good probably only to within a factor of two or three. They do, however, indicate the overall scale of the computing problem. In fact, they are an overestimate since while we have neglected both self-calibration, which may be required in some cases, and general analysis of the processed data, which will quite often be somewhat time-consuming, we have also neglected the limited duty cycle of the instrument at high frequencies, where the computing requirements are heaviest.

[6]Cornwell, T. J. (1984), "Mosaicing with the millimeter array", NRAO Millimeter Array Memo. No. 24.

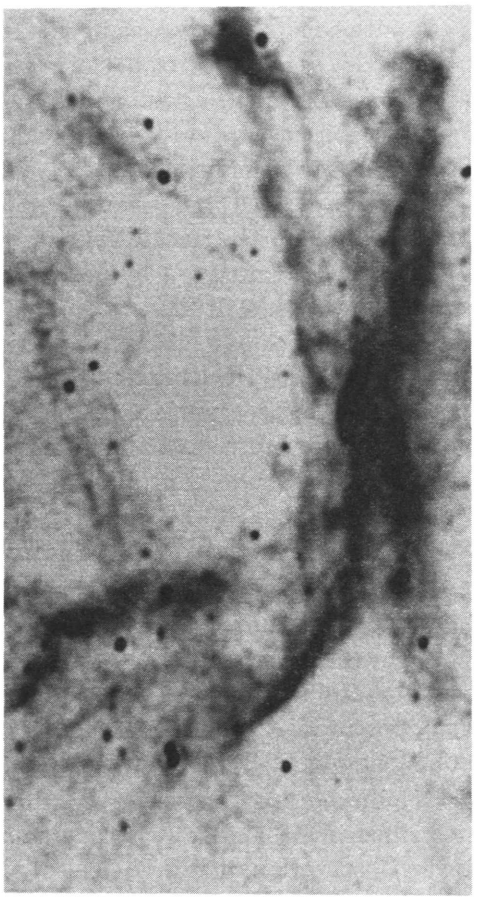


Figure 1. A mosaic image of the radio source Simeis 57, observed with the VLA at 20 cm wavelength.

4. Beam-Switching and Self-Calibration

The troposphere will corrupt the calibration of the MMA at all frequencies. The simplest effect is that the emission from the troposphere will vary on a time-scale of a few minutes and thus corrupt any total-power measurements. In single-dish observations, the best countermeasure involves differencing the power from two feeds pointing in slightly different directions. An image of the sky brightness can be reconstructed very satisfactorily from such dual-beam observations if the source is not too large with respect to the beam separation.[7] In interferometric observations the total power and its variations can be discarded directly. Much like smaller dual-beam observations, an image of the sky can be reconstructed if the object is not much bigger than the broadest sidelobe.

[7]Emerson, D. T., Klein, U. and Haslam, C. G. T. (1979), "A multiple beam technique for overcoming atmospheric limitations to single-dish observations of extended radio sources", *Astron. Astrophys.* **76**, 92–105.

A more complicated set of problems comes from the propagation of the signal through the troposphere. Both the opacity of the atmosphere and electrical path length of the signal through the atmosphere will fluctuate sufficiently rapidly that calibration on a standard source will be problematical some large fraction of the time. There are basically three ways around this:

(a) wait for good weather;
(b) correct using water-vapor radiometer measurements; or
(c) calibrate on the object actually being observed.

The first strategy is guaranteed to work eventually, but the amount of observing time lost may be large. Experiments with (b) have not been encouraging. For strong sources, the third option, self-calibration, is possible.[8,9,10] Self-calibration treats the unknown and variable calibration as something to be determined from the observed data. Since the calibration errors originate at array elements, whereas the visibility function can be measured between all *pairs* of elements, only relatively few extra degrees of freedom are thus introduced. These can be balanced either by forcing the array design so that redundant spacings are obtained or by enforcing known properties of the final deconvolved image such as positivity and confinement. Both of these approaches have been used to similar effect.

The computational cost of self-calibration is outlined in Appendix 2, where it is shown that the net added computing load for spectroscopic observations will usually be minimal. For continuum observations, the total computing load is approximately multiplied by the number of self-calibration iterations required, usually fewer than three or four.

5. Software and Hardware Requirements

The possible MMA designs which have been described would allow high resolution spectral line imaging of large objects. Realizing the scientific potential of such an instrument requires that the computing facilities available to the astronomer be similarly flexible and powerful. With this in mind we set the following goals for the computing system:

(1) The system should be capable of keeping abreast of the reduction of observations.
(2) The system must be capable of performing all the types of processing discussed in this chapter, and must be sufficiently flexible that variants and new algorithms can be easily added.
(3) Since the MMA will be predominantly used for spectral line observations, it is very important that spectral-line data processing be

[8]Schwab, F. R. (1980), "Adaptive calibration of radio interferometer data", *Proc. S.P.I.E.* **231**, 18–25.

[9]Cornwell, T. J. and Wilkinson, P. N. (1981), "A new method for making radio maps with unstable radio interferometers", *Mon. Not. Roy. Astron. Soc.* **196**, 1067–1086.

[10]Cornwell, T. J. (1986), "Self-calibration", in *Course Notes from an NRAO Summer School Held in Socorro, New Mexico August 5–9, 1985*, Eds. R. A. Perley, F. R. Schwab, and A. H. Bridle, (NRAO: Green Bank, WV).

given very high priority. In particular, hardware and software should combine to make reduction, analysis, and display of cubes of data straightforward.

We will now discuss how each of these goals can be achieved.

It is important to note that the actual computing cost will be determined by the mixture of science performed with the array, since some types of data reduction are much easier than others. For example, deep detection experiments can consume large amounts of telescope time, but little computing time; while mosaicing of large strong objects is at the opposite extreme, requiring little observing, and a great deal of computing. Since the overall computing costs depend upon the mixture of observations, we have not attempted to provide a definitive estimate. Instead, in Appendix 1, we discuss two gedanken experiments, one typical and one extreme. The high computing speeds necessary for strong-source mosaicing can be obtained by two possible routes: either by a single fast processor, or many more slower processors, each of which works on part of the problem. In practice, of course, machines available on the market span all possibilities between these two extremes. Programming the highly parallel architectures is an as yet poorly understood art, but they seem to be the cheapest way of obtaining a certain nominal computing power. The inflexibility of such parallel machines is in direct opposition to goal (2), and thus rules out machines like the various hypercubes, at least until the software technology catches up with the hardware. However, it is clear that the mosaicing of many spectral channels is very well suited to parallel processing because the deconvolution of the channels can proceed independently. Thus the required computing could be realized in a number of processors operating synchronously, each on a small subset of the spectral channels. Two currently available machines, the CONVEX C-1 and the the Alliant FX/8, fit this specification. Even for very conservative estimates of the achievable throughput, a small number, say six, of either of these computers, could handle the computing burden in the typical gedanken-observation, while only seven would be required for the extreme example. These would certainly handle the large observations but may be more than is necessary for simple observations, for example those employing just the smaller configurations and a few spectral channels. Indeed some simple detection experiments or spectroscopy of small objects could be processed on a microcomputer. We would therefore envisage a hierarchy of computing capacity, ranging from simple micro-computers, such as IBM PC-ATs, up to workstations such as the SUN IV through to the mini-supercomputers such as the CONVEXes or Alliants. We should emphasize that these models are chosen only for illustration, and that undoubtedly the choice would be different in a few years. In particular, it seems reasonable to assume that the single processor option will become possible at an acceptable price (see Appendix 1).

Goal (2) concerns design of the software for the data reduction. Although NRAO has many man-years invested in a radio interferometric data reduction package, the Astronomical Image Processing System (AIPS), it is not

optimized either for spectral-line reduction or for mosaicing. A new package could take advantage of some recent and anticipated changes in computing: the availability of an almost universal operating system, UNIX; code management systems which ease the problem of maintaining a large body of software; virtual memory, vector processing computers; falling hardware costs; vectorizing compilers for modern languages such as C, Modula-2 or ADA; and, finally, the experience of other groups in producing software for spectral-line interferometry. We would therefore consider the development of a suitable software system to be of the utmost importance for the long-term scientific productivity of the MMA.

Goal (3), the combination of hardware and software to simplify data reduction and analysis, requires exploitation of modern graphics and display devices to allow the astronomer to fully comprehend his observations. For example, large memory, fast-refresh display devices can be used to represent the three dimensions of right ascension, declination and frequency via "animated" movies or rotating views of isophotes. Vector graphics can be used for the display of spectral-line visibility data, or for ruled surface plots. Many forms of hardcopy will also be required, such as video tapes of movies for subsequent analysis.

A very rough estimate of the cost of the hardware for off-line processing is between \$3M and \$6M. We will adopt \$4M as our current estimate. Given the current rapid evolution of hardware costs and algorithmic demands, we do not believe that attempting a firmer estimate would be worthwhile.

6. On-Line Processing

Operation of the MMA will be dependent upon the weather; if the phase stability or atmospheric opacity are unfavorable then an observing program at high frequency may have to be re-scheduled at short notice. In fact, dynamic scheduling will certainly be necessary to exploit periods of good weather. The on-line system will therefore have to allow real-time data quality evaluation, probably by direct inspection or by imaging, including deconvolution and self-calibration.

Independent of the need for dynamic scheduling, the on-line system should also be able to apply some initial calibration, and to detect and report bad data.

The budget for the on-line computing is roughly estimated at \$2M.

Appendix 1. Computing Cost of Mosaicing

We will first consider strong-source mosaicing. The CPU work involved in the mosaicing is dominated by the data comparison stages, which are themselves dominated by two-dimensional Fast Fourier transforms (FFTs) required either to transform to and from the u-v plane or to convolve with a single-dish reception pattern. We will express the cost in time taken for CONVEX C-1, since we have a good benchmark timing of a 1024 by 1024 pixel MEM deconvolution as taking 560 sec for thirty iterations. We will further assume that the rapidly-dropping cost of fast memory will allow each (32-bit) CPU to

address its maximum amount of memory: 1 Gbyte. All arrays for one channel can then fit into memory, and thus i/o costs are avoided.

First, for interferometer data, for each pointing center, we do the following:

(i) Taper the current iterate with the appropriately centered reception pattern;
(ii) Fourier transform to the u-v plane;
(iii) Interpolate values for the sampled visibilities and calculate residual visibilities;
(iv) Grid residual visibilities onto the regular grid required for an FFT;
(v) Inverse FFT to image plane; and
(vi) Taper the resulting image with the reception pattern.

In view of the small image size, steps (ii), (iv) and (v) can be performed in memory, and thus only two passes through the visibility data are required. The image or its transform remains in memory throughout steps (i)–(vi). One very important advantage of this is that the visibility data need not be sorted for gridding.

Second, for the single-dish total-power observations:

(i) Convolve the current iterate with the reception pattern of the elements;
(ii) Interpolate from the grid to the sampled positions and subtract from the sampled values to form the residuals;
(iii) Grid the residuals onto the image; and
(iv) Convolve the result with the reception pattern of the elements.

Again all this may be performed in memory. The convolutions may be evaluated using FFTs.

The time taken is then:

$$T_{\mathrm{mosaicing}} = 1.5 \times 10^{-6} N_{\mathrm{chan}} N_{\mathrm{iter}} N_{\mathrm{pc}} (N_{\mathrm{pix}}^2 \log_2 N_{\mathrm{pix}}) \,\mathrm{sec}\,,$$

where N_{chan} = number of channels to be processed, N_{iter} = number of iterations per channel, N_{pc} = number of pointing centers, and N_{pix} = number of pixels along one axis which is needed for one pointing. If B is the maximum baseline, and D is the diameter of the dishes, then:

$$N_{\mathrm{pix}} \approx \frac{4B}{D}\,.$$

The main burden of the work is due to the FFT stage of the data comparison, all other steps (such as entropy maximization, primary beam correction, data comparison for the central element) being negligible.

The most important attribute of this equation is that the amount of work scales as the field of view, rather than a higher power as one might expect. This occurs because the primary beam of the array elements limits the cross coupling of data, and so each data comparison need only consider a small patch of the whole image. In practice, we expect that the number of iterations would increase slowly with the number of fields.

The weak source mosaicing is very much easier. The steps are:

(i) For each pointing, grid visibilities onto the regular grid required for an FFT. Then FFT to obtain the image for each pointing center.

(ii) Form a mosaic image by forming a weighted sum of all the images calculated in (i), where, at each output pixel, the weight of each separate image is proportional to the inverse variance of the noise.

The time is then about:

$$T_{\substack{\text{weak}\\\text{source}}} = 7.5 \times 10^{-7} N_{\text{chan}} N_{\text{pc}} (N_{\text{pix}}^2 \log_2 N_{\text{pix}}) \,\text{sec}.$$

Thus the work required for weak source mosaicing is a factor $2N_{\text{iter}}$ smaller than that for strong source mosaicing.

To get some feeling for the actual processing required for a typical observation, let us consider a spectral-line observation with 512 channels of the a mosaic of 10 by 10 pointings with the resolution of the 300 m configuration. Thus we must observe in the two smallest configurations—90 m, 300 m—for eight hours in each configuration. Each channel will be processed independently, although we can probably save time by using the solution for a previous channel as an initial guess. At the standard sampling of about 2–3 pixels per resolution element, the overall image will have 512^2 pixels. About ten iterations per channel should suffice. We have that $T_{\substack{\text{strong}\\\text{source}}} = 4.7$ days and $T_{\substack{\text{weak}\\\text{source}}} = 5.6$ hours on a CONVEX C-1 computer.

To summarize this relatively modest example: to keep up with the averaged data rate for this case (one complete set of data in 2×8 hours spread over two different configurations), we would require six CONVEX C-1 class computers, each with a large amount of fast memory, say 1 Gbyte. The cost would be about twenty times less if weak source mosaicing were acceptable scientifically.

Perhaps the most extreme example concerns spectral line observations of a molecular cloud of diameter 30 arcminutes, with a resolution of about 1 arcsecond at frequency of 230 GHz. This requires about 10^4 pointings. If we assume that on each pointing, 1 minute of integration time is used, then the observing time in each configuration is about seven days, probably spread over at least 21 days to get low air-mass. The total time is thus at least forty days. The brightness sensitivity would be about 200 mK, with a velocity resolution of 10 km/sec. The product would be a cube of size $4096 \times 4096 \times 512$ or 30 Gbyte. The time required on a CONVEX C-1 would be about 100 days, or about ten times higher that that consumed by the most ambitious VLA project: the imaging of M31 in the hydrogen line. At first glance this looks hopeless; however, to get it down to real time (14 days) requires only seven CONVEX C-1 equivalents. A recent survey of the mini-supercomputer market indicates that currently the cost/performance ratio is halving every two years or so. By about 1993, we can therefore expect to be able to purchase this capacity for no more than about \$2M. Adding on work-stations, display units, and other peripherals, such as a common file system will no more than double this number so we will adopt a figure of \$4M. Given the uncertainties in

predicting future developments in both computers and algorithms, this figure should be regarded as accurate to only about 50%, the true figure probably lying between \$3M and \$6M.

The saturation in required computing speed evidenced in the above examples arises from the assumed minimum integration time of about one minute. Reducing this limit to 20 seconds, say, would increase the computing load to about 21 C-1 equivalents.

Appendix 2. Computing Cost of Self-Calibration

Self-calibration is usually implemented as an additional number of deconvolutions. An initial deconvolution or a model of the object is produced and the calibration adjusted to move the observed data as close as possible, usually in a least-squares sense, to the modeled data. The object must be strong enough that the visibility function is measurable within the coherence time of the calibration error. The errors thus estimated can then be removed from the observed data, and an improved deconvolution performed. This then forms the model for the next cycle of correction (see e.g., Cornwell[10] for further details). The number of such passes required depends upon both the source complexity and the u-v coverage as well as the quality of the initial calibration. For the VLA only a few passes are required, whereas for VLBI many tens of passes are needed for convergence. For a continuum experiment, the overall number of deconvolutions required increases by the number of passes necessary, but for spectral-line observations, the self-calibration is usually performed either on the continuum or, for example, in the case of a maser, on a particularly strong channel. Thus, for a large number of channels the work load is negligibly increased. One exception to this occurs when the signal-to-noise ratio in a single channel is not sufficient. One then must do a three-dimensional least-squares fit, summing over the different channels and also deconvolving the separate channels a number of times.[10] Hence the deconvolution of each channel must be performed once per self-calibration pass. It is not clear, though, how this interacts with the mosaicing algorithm described above, and whether mosaicing is necessary or indeed possible on such a weak object.

VI. Electronics

1. Introduction

Much useful experience in the design of electronics for a large millimeter array has been gained through the design of the VLA, the VLBA, single-dish millimeter-wave receivers, and the millimeter-wave interferometers such as those at Owens Valley, Hat Creek, Nobeyama, and the IRAM facility in France. There is little doubt that much of the electronics is presently feasible but that extensive development work in the next several years will be needed to realize the performance and reliability necessary for the MMA.

In this chapter the electronics areas which are of particular relevance are discussed, with attention to the key parameters which must match the astronomical requirements, and with reasonable cost estimates for both construction and operation. The areas to be discussed are receivers, local-oscillators, data transmission, and the correlator. A final section summarizes the electronics areas where development work is recommended.

2. Receivers

1. General. In the last five years it has become clear that whenever high sensitivity heterodyne receivers are required for the 30–350 GHz range, SIS mixers are the prime choice. Receivers using SIS junctions made by the National Bureau of Standards, IBM, and the Hypres Company have consistently given single-sideband receiver temperatures under 100 K at 115 GHz, and the rule of thumb that $T_{R_{SSB}}$(K) = LO frequency (GHz) now seems to be quite conservative. The newer, all-niobium junctions operate well at 4.2 K (compared with $\sim$ 2.5 K for lead-alloy junctions), and appear to have excellent long-term stability as well as being more resistant to electrical shock and heating than their non-refractory predecessors.

The only practical alternative to the SIS mixer receiver is the Schottky-diode mixer receiver. At 115 GHz the better Schottky receivers have $T_{R_{SSB}}$ = 150–200 K—about a factor of two higher than their SIS counterparts. We expect this factor to increase to about four in the next few years as SIS receivers are designed to take full advantage of their potential for high gain and low shot-noise. Schottky mixers do have two practical advantages over SIS mixers:

(i) they operate happily at $\sim$ 20 K; and
(ii) they have much higher saturation power levels (significant particularly for precise calibration).

These advantages are offset by:

(i) their higher LO power requirements, typically 100 μW at the diode, compared with $\leq$ 1 μW for SIS mixers, depending on factors discussed in Part 4 of this Section; and
(ii) the continuing need for use of relatively unreliable whisker-contacted diodes to achieve acceptable performance (at least above $\sim$ 100 GHz).

The four frequency bands given greatest emphasis at the MMA Science Workshop are centered in the atmospheric windows at 9, 3, 1.3, and 0.85 mm. In each band a large tuning bandwidth is desirable, so that the receiver characteristics do not impose artificial limitations on the science. Ideally, one would like to be able to observe anywhere within the entire frequency range of each of the four atmospheric windows, but this would require either tunable receivers or multiple receivers in each band. Note that even in the most favorable case, in which tunable receivers prove satisfactory, there would be a very large number (320) of receivers in the MMA (4 bands × 2 polarizations × 40 antennas). Tuning range and bandwidth, both important parameters, are further discussed below.

2. Single-sideband vs. double-sideband receivers. If it is practical to use the difference in fringe-rates to separate the upper-sideband ($f_{LO} + f_{IF}$) from the lower-sideband ($f_{LO} - f_{IF}$), then there is no fundamental reason for choosing single-sideband (SSB) operation and thereby restricting the available bandwidth for continuum measurements. For spectral line observations the only disadvantage of double-sideband (DSB) operation is the additional atmospheric noise, received in the image sideband, which is not eliminated by the sideband selection process.

As SIS mixers operate best in the DSB mode, we shall assume that an SSB receiver would consist of a DSB mixer preceded by a diplexer which presents an image termination (preferably cold) to the mixer. Such a diplexer can in principle be realized either in waveguide or as a quasi-optical structure using Gaussian beams. However, if it is to be tunable, the quasi-optical approach is more practical. If substantial IF bandwidths are required, e.g., 1–2 GHz, 2–4 GHz or even 2–8 GHz, an additional tunable quasi-optical filter element will be needed, resulting in a mechanically complex diplexer with two or three movable elements. Assuming such a diplexer is outside the receiver dewar, the effective temperature of a 4 K cooled image termination will be at least 30 K, and more likely 60 K, thereby reducing the advantage of the SSB receiver over the DSB receiver. Considering also the considerable mechanical complexity of the tunable diplexer, there seems little reason not to choose DSB receivers.

3. Bandwidth. The instantaneous RF bandwidth of an SIS mixer is ultimately limited by the $\omega R_N C_J$ product of its junction(s), where ω is the LO frequency, and R_N and C_J are the normal resistance and the capacitance of the junction. Within a factor of about two, $\omega R_N C_J$ can be regarded as the Q-factor of the junction. The best SIS mixers appear to have $\omega R_N C_J$ products near five,[1] which implies approximately a 20% 1 dB bandwidth for a matched mixer. In practice, the instantaneous bandwidth is further restricted by parasitic reactances and matching or tuning circuits within the mixer.

Our experience with SIS receivers at 115 GHz suggests that we can obtain an instantaneous bandwidth of no more than 10% but that the receiver can

[1] J. R. Tucker and M. J. Feldman (1985), "Quantum detection at millimeter wavelength," *Rev. Mod. Phys.* **57**, 1055–1113.

be mechanically tuned over 30% fractional bandwidth with some sacrifice in performance at the extremes. At higher frequencies, both the instantaneous bandwidth and the tuning range are expected to degrade somewhat, to values approaching 5% and 15%, respectively, at 345 GHz.

It is not certain, at present, whether a single SIS mixer with a simple tuning mechanism can be made to cover each desired frequency band with acceptable performance. However, by integrating the main (fixed) tuning elements into the SIS junction design, controlling the junction parameters closely, and employing electronic tuning in the mixer, it may be possible to approach the desired tuning ranges. The use of varactor diode tuning in the SIS mixers would circumvent the need for mechanical tuning, with its servos, position indicators, gears, linkages, and heat-sinking problems (which have been quite difficult to overcome in existing SIS receivers). Research into this area is in its initial stages.

4. Saturation in SIS mixers. In any mixer, saturation occurs when the signal amplitude becomes an appreciable fraction of the LO amplitude. In order to allow precise calibration, the MMA SIS mixers should exhibit gain compression of no more than a few percent when exposed to an ambient (300 K) load. We know of design strategies that we believe will allow this criterion to be met, while maintaining very low receiver temperatures, but this has yet to be demonstrated. These strategies involve the use of several junctions in series and careful control of the IF impedance.

5. LO power requirements and diplexers. The LO power coupled to the junction(s) of a DSB SIS mixer with N junctions in series, optimally tuned, and operating on the first gain step below the gap voltage is given by[1]

$$P_{\mathrm{LO}} = \frac{(Nhf_{\mathrm{LO}}\alpha)^2}{2R_N e^2},$$

where h and e are Planck's constant and the electron charge, α is a pumping parameter close to 1.2, and R_N is the normal resistance of the whole array. Adopting $R_N = 50$ ohms and considering the case of four junctions ($N = 4$), the LO power required by the mixers is less than 0.5 μW for frequencies as high as 350 GHz. In practice, we find losses in the LO circuit to be as high as 10 dB and losses in the LO coupler of approximately 20 dB. Together these increase the LO power needed, but, even with both effects included, less than 1 mW is sufficient at all the MMA frequency bands.

To achieve efficient coupling of LO power to the mixer with low signal-path loss, it is necessary to use a tuned diplexer—either a resonant circuit such as a resonant-ring waveguide filter or its quasi-optical equivalent, or an interferometric circuit such as the Martin–Puplett or Mach–Zehnder structures. In the ideal case all of these have zero loss in both signal and LO paths. The disadvantage of these diplexers in the present application is the difficulty of achieving wide instantaneous signal bandwidths: In the interferometric diplexers the response varies with frequency with a period $\leq 2f_{\mathrm{IF}}$, while for the

resonant-ring diplexer it is hard to achieve, simultaneously, wide tuning ranges and sufficient spacing between successive resonances. All of these diplexers require extremely precise mechanical tuning mechanisms, with their attendant servos, position indicators, and linkages. As LO sideband noise filtering is not necessary for SIS receivers, and the required LO powers are fairly small, the simplest and most reliable choice of LO diplexer appears to be the simple beam splitter or directional coupler.

Because it must give the highest LO power at the highest frequency, the 345 GHz LO system is the most difficult one to produce. Using a Gunn oscillator at 86 GHz driving a ×4 varactor multiplier, it should be possible to obtain the needed LO power at 350 GHz. With the exception of the widely tunable harmonic Gunn oscillators developed recently at Berkeley[2] currently available solid-state sources do not have the desired 25% to 40% tuning ranges. Based on present technology it would probably be necessary to use at least two Gunn oscillators to cover the tuning range of each receiver. It is expected that each Gunn oscillator would have two mechanical tuners, a varactor for phase-locking, and a frequency dependent bias voltage—a total of four tuning adjustments. If more than one Gunn oscillator is needed, a waveguide switch will be needed.

Varactor frequency multipliers are used as local-oscillators in most millimeter wave receivers. Almost all use whisker-contacted GaAs diodes. Above about 100 GHz (output), whisker length and diode chip position become quite critical in obtaining good efficiency and wide tuning range. From the standpoints of reliability and reproducibility, it is therefore vitally important that the development of planar varactor diodes for these frequencies, already under way at the University of Virginia, be continued. It is quite possible that the desired frequency bands will eventually be coverable by use of a single mechanical tuner or none at all.

In summary, it should be possible to provide sufficient LO power using solid-state (Gunn) oscillators with frequency multipliers. LO coupling to the mixers by broadband directional couplers eliminates the need for tunable diplexers and degrades the receiver noise temperature by only a small amount. Work should be started as soon as possible on developing reliable, easily reproducible frequency multipliers using planar varactor diodes and widely tunable Gunn oscillators to cover the required frequency ranges.

3. Data Transmission

The array will require the transmission of wideband (1 to 10 GHz) IF signals, local-oscillator reference signals, and also control and monitor signals. The key parameters are the bandwidth of the IF signals and the phase stability of the LO signals.

Fiber-optic links are the most likely approach for all signal transmission. Free-space radio has severe bandwidth and interference problems, TE_{01} mode

[2]J. E. Carlstrom, R. L. Plambeck, and D. D. Thornton (1985), "A continuously tunable 65–115 GHz Gunn oscillator," *IEEE Trans. Microwave Theory Tech.* **MTT-33**, 610–619.

waveguide (as in the VLA) is not commercially available, and coaxial cable is not competitive in bandwidth and loss with optical fibers. One possibility is the use of coaxial cable only for LO transmission, since narrow bandwidth is required and the loss is reasonable for a few km paths. Path-length correction systems for UHF carriers on coaxial cables are well developed. If the path-length correction system is to include modulators and demodulators, these devices must be reciprocal (they must function for signals transmitted in both directions). This is the case for UHF modulators but is not true for the optical counterparts. Optical fibers themselves have some degree of non-reciprocity. On the other hand, it would be a great convenience and cost saving if the local-oscillator reference could be accommodated on a single fiber of a multi-fiber cable; optical modulator stability and correction schemes need further investigation. Much valuable information concerning the application of optical fiber transmission to radio astronomy arrays is coming out of the Australia Telescope project.[3] Optical transmitters and receivers with 6 GHz bandwidth are available,[4] but A/D converters are not.

The most likely optical transmission medium is single-mode optical fiber with 10 μm core, transmission loss of 0.5 dB/km at a wavelength of 1.3 μm, and allowable distance × bandwidth products of 20 km × GHz. The cost of single-mode fiber is now approximately the same as multi-mode fiber (both[5,6] $\sim$ \$250/km) which has loss of 3 dB/km and an allowable distance × bandwidth product of 0.4 km × GHz. The multi-mode fibers have lower costs for connectors and terminal equipment but would not be appropriate for the bandwidth and distances anticipated in the millimeter array.

A large portion of the cost of a fiber-optic transmission system is in cable-armor ($\sim$ \$1,000/km)[5] and installation cost ($\sim$ \$6,000/km)[7]; it would thus be prudent to initially install a system with a sufficient number of fibers for future growth. As a bench-mark we will assume 2 GHz transmission is required at up to 3 km distance. This could be accomplished with one single-mode fiber. Allowing three additional fibers for use as spares, LO, control, and monitor transmission results in a transmission media cost of \$2,000/km, to be added to \$6,000/km of installation cost at a burial depth (direct plow-in of cable) of 0.5 meter. The terminal costs for IF, LO, control, and monitor are estimated at \$30,000, with design costs at \$400,000, so that the cost equation for data transmission is

$$I\$ = 8DN + 30N + 400\,,$$

[3] *Cf.* A. C. Young, P. A. Miller and T. M. Percival, "High-speed optical fibre transmission for the Australia Telescope," to be published, CSIRO Division of Radiophysics, Epping, NSW, Australia; and see later AT Technical Notes.

[4] Models LDS10-PM-T and PD050-PM-R, Ortel Corporation, (818-281-3636).

[5] T. Dunne, Private communication, Corning Glass Company, (607-974-9000), May 20, 1986.

[6] S. Casey, Private communication, Ericson, Inc., (913-541-7608), May 22, 1986; prices of \$2.21/meter for armored 4 mono-mode fiber cable and \$8.00/meter for armored 24 mono-mode fiber cable were quoted.

[7] T. Kitay, Private communication, Sargent Electric Company, (412-394-7580), May 23, 1986.

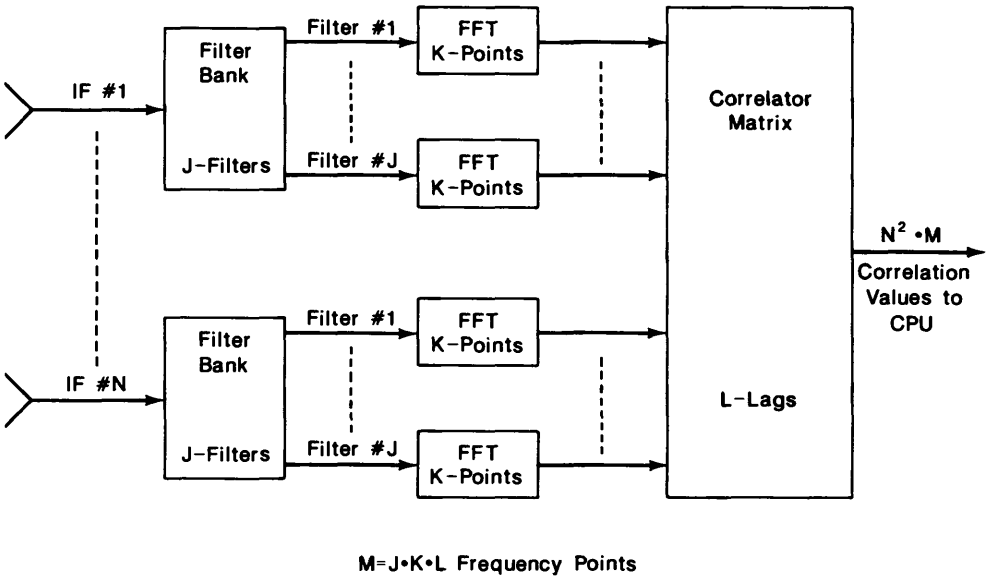


Figure 1. General correlator configuration allowing spectral analysis by analog filters, FFT operations, and cross-correlation. The minimum cost configuration is described in the text.

where $I\$$ is in \$k, D is the average distance in km to an antenna station, and N is the number of antennas.

4. Correlator

The bandwidth required by a millimeter array is an order of magnitude greater than that required by the VLA or the VLBA. This increases the correlator cost to an appreciable fraction (perhaps 10% to 20%) of the array cost, and the minimization of this cost must be carefully considered. Optical processing approaches have been considered (see Millimeter Array Memoranda Nos. 7 and 23) but because of problems of flexibility and dynamic range this is not the favored technique at the present time.

The required bandwidth, B, from N antennas can be correlated in M frequency bands by the configuration of filter banks, fast-Fourier transform (FFT) operations, and cross-correlators shown in Figure 1. The N filter banks each with J filters are characterized by a cost factor C_F, currently estimated at \$900 per filter. This estimate includes amplifiers and local-oscillators associated with the filter, includes labor, and is based upon a 64 channel filter bank recently constructed for a hybrid-spectrometer. The NJ FFT operators of length K are characterized by a cost factor C_T, currently estimated at \$4,200 per 10^9 operations per second, where operation is defined as two complex multiplications and additions (a "butterfly") with the required accuracy. The required accuracy is a question requiring further study but is believed to be in the range of 6 to 12 bits. The estimate of C_T is based upon CMOS gate array technology with 8-bit accuracy and includes power and labor. Finally, the cross-correlator has NL multipliers, where L is the number of lags, and a cost factor C_X, currently estimated at \$500 per 10^9 multiplications per

second—with again a required accuracy which is not known at present but is thought to be in the range of 2 or 3 bits. The number of frequency channels is then given by $M = JKL$.

The total cost, $C_{\rm tot}$, of the correlator system can then be expressed as:

$$C_{\rm tot} = C_F \times \underbrace{N \times J}_{\text{Filters}} + C_T \times \underbrace{K \log_2 K}_{\substack{\text{Operations}\\ \text{per FFT}}} \times \underbrace{N \times J}_{\substack{\text{per}\\ \text{antenna}\\ \text{and}\\ \text{filter}}} \times \underbrace{\frac{2B}{JK}}_{\substack{\text{FFTs}\\ \text{per}\\ \text{second}}} + C_X \times \underbrace{N^2}_{\text{Products}} \times \underbrace{L}_{\text{Lags}} \times \underbrace{J \times K}_{\substack{\text{Freq.}\\ \text{points}}} \times \underbrace{\frac{2B}{JK}}_{\substack{\text{Products}\\ \text{per}\\ \text{second}}}$$

Using $K = \frac{M}{JL}$ and simplifying gives

$$C_{\rm tot} = C_F N J + C_T N 2B \log_2 \frac{M}{JL} + C_X N^2 2BL\,.$$

By taking the derivative of $C_{\rm tot}$ with respect to L, the optimum value $L_{\rm opt}$ is found to be

$$L_{\rm opt} = 1.44 \frac{C_T}{N C_X}\,.$$

An important conclusion can now be reached. With the C_T/C_X estimate at 8.4, the optimum value of L is near unity for $N \approx 24$. It cannot be less than one, so we find that the cross-correlators should be only cross-multipliers, $L_{\rm opt} = 1$, and all spectral resolution should be accomplished by filters and FFT operations. Stated another way, the MMA correlator, if designed as an 'FX correlator' (like the VLBA correlator), is much less expensive than a 'lag' or 'XF correlator'. This is a consequence of the fact that N^2 correlators are needed per lag, whereas the filters and FFT operations are proportional to N.

The remaining step is to find the optimum tradeoff between filters and FFT operations in the FX design. Taking the derivative of $C_{\rm tot}$ with respect to J gives the optimum value of J as

$$J_{\rm opt} = 2.88 \frac{C_T B}{C_F}\,.$$

Typical cost figures based upon the above analysis and cost factors are given in Table 1. The first line of this table corresponds to the FX design, with the number of filters chosen so as to minimize the cost. The second line gives the cost of a lag, or XF, correlator. The final line represents the FX design with the number of analog filters minimized. The figures do not include design costs of \$1160k for 16 man-years of engineering and technician support and \$200k of materials.

Table 1. Correlator Cost (in 1986$), assuming $C_F = \$0.9\text{k}$, $C_T = \$4.2\text{k} \times 10^{-9}$, and $C_X = \$0.5\text{k} \times 10^{-9}$.

Fundamental Parameters			Internal Parameters: Filters		FFT		Cross-Mult.			
N	B	M	J	Total	K	Total	L	Total	Total Cost	
Ants.	GHz	Freqs.	No.	Cost	Points	Cost	Lags	Cost	C_{tot}*	Comments
40	2	1000	29	\$1050k	34	\$3450k	1	\$3200k	\$7700k	Min. cost FX design
40	2	1000	216	7800k	1	0	4.6	14800k	22600k	Min. cost XF design
40	2	1000	8	300k	128	4700k	1	3200k	8200k	Min. filter FX design

*Add \$1360k design costs to these figures.

5. Development Areas

A millimeter array could, in principle, be constructed with present day technology. However, to realize the instrument proposed here, a major developmental effort will be needed in the emerging technology of SIS mixers. In the years preceding the construction of the array, the following items should be investigated:

(1) SIS Mixers: Basic designs for reproducible SIS mixers should be developed for each of the four frequency bands. Limitations on bandwidth and tuning range will have to be explored to determine whether more than four receivers will be needed to cover the desired frequency windows.

(2) Electronically Tunable SIS Mixers: Designs which integrate tuning elements such as varactors with SIS junctions on the same substrate might allow high performance over the $\sim$ 30% band required for each front-end. This would remove the need for unreliable mechanical tuners that must operate at cryogenic temperatures.

(3) Higher Temperature SIS Junctions: Present SIS junctions require cooling to temperatures of 4–5 K, and in some cases performance is improved by cooling to 3 K or less. Future junctions using NbN might perform well at 8–10 K and would thus ease the refrigeration problem somewhat. NRAO must continue to be involved in the development of the technology of such junctions. The new high $T_{\rm crit}$ ($>$ 80 K) superconductors may have fundamental difficulties in their application to low-noise receivers, but we must continue to monitor the intensive world-wide research effort for possible breakthroughs.

(4) Reliable 4K Cryogenic Refrigerators: In view of the large number (320) of cryogenic receivers, it is important to have proven refrigeration systems that can be economically maintained.

(5) Wideband Local Oscillators: While existing Gunn oscillators and whisker-contacted Schottky diode frequency multipliers can deliver the local-oscillator powers needed, the tuning range of the Gunns and the reliability and reproducibility of the varactor multipliers need improvement for millimeter array receivers. Work now under way on planar varactor diodes at the University of Virginia should result in high quality, reliable varactors for tuning and harmonic generation within the next few years. It is expected that pressure from the military markets will lead to improved Gunn diodes, although the

design of widely tunable oscillators using these diodes will probably need to be undertaken by the radio astronomy community.

(6) Wideband, Cryogenic IF Amplifiers: The conversion loss of SIS mixers can be made low, but moderately low-noise (< 30 K) IF amplifiers are still required with bandwidths of the order of 10 GHz. This is possible with balanced or distributed GASFET or HEMT amplifiers, but development work is needed.

(7) Fiber-Optic Transmission Methods: As mentioned, the Australia Telescope is a valuable test-bed for this technology. Local expertise and attention to the particular problems of the millimeter array (e.g., bandwidth and phase stability) are needed. Since a large portion of the cost is in the installation, methods appropriate for a remote radio astronomy site need investigation.

(8) Correlator Development: The cost factors which have been described are important for choosing an optimal design and for arriving at an accurate cost estimate. We must keep abreast of the rapidly occurring developments in digital technology; as we do so, these figures will require frequent revision.

VII. Construction Budget

The construction cost of an instrument such as described in the previous chapters is as yet uncertain. The design is not finalized, the site is still to be determined, and possible savings from advances in technology between 1987 and the start of construction are unrealized. Nevertheless we can estimate the cost in current dollars of various designs, such as the one described in Chapter IV, given specific assumptions about the site. We begin by describing the basic cost equations that have been used in estimating the cost of each major component of the array.

1. Cost Equations

The cost equations have been presented in Memorandum No. 6 of the Millimeter Array Technical Memorandum Series and in Chapter VI of this document. The costs of radio telescopes operating at millimeter wavelengths have also been discussed by Moran *et al.*[1] and in the NRAO VLBA proposals and associated documents. The following are estimates of the cost for the telescopes and receivers in units of 1987 k$ and assuming 1987 technology.

(1) Antenna System Cost, A$. For N antennas of diameter D meters with accuracy $\lambda/16$, where λ is in millimeters,

$$A\$ = 890\lambda^{-0.7}(D/10)^{\alpha}N + 500\,.$$

(2) Front-End System Cost, F$. For N antennas, M frequency bands, each 30% wide, and dual polarization,

$$F\$ = 45MN + 200M\,.$$

(3) LO System, L$.

$$L\$ = 80N + 100\,.$$

(4) IF Transmission, I$. From Chapter VI, for average baseline lengths B in km and N antennas,

$$I\$ = 8BN + 30N + 400\,.$$

(5) Correlator, $C_T$$. See Chapter VI for the derivation which leads to the parametrization below. If we take the bandwidth (2 GHz, or 1 GHz per polarization) and number of channels (1000) from the first example, we can write the cost of the correlator as

$$C_T\$ = 2N^2 + 112N + 1360\,.$$

[1]Moran, J. M., Elvis, M. S., Fazio, G. G., Ho, P. T. P, Myers, P. C., Reid, M. J., and Willner, S. P. (1984), "A Submillimeter-Wavelength Telescope Array: Scientific, Technical and Strategic Issues", Smithsonian Astrophysical Observatory.

(6) Computers, C$. This item is very uncertain, but based on Chapter V and on an estimate for the on-line computers which is based on VLA experience we estimate

$$C\$ = 2.5N^2 + 45N + 200\,.$$

(7) Site, S$. Roughly, we estimate

$$S\$ = 70N + 500\,.$$

2. Cost of Construction

The approach adopted in this section is similar to that used in obtaining cost estimates for the VLBA project. We assume the basic array parameters described in Chapter IV (forty 7.5 meter dishes, baselines up to 3 km, 1 GHz bandwidth, etc.) and adopt a scaling of antenna cost proportional to $D^{2.5}$.

Using the equations in the previous section and the assumptions listed above and in the discussions of the cost (e.g., the correlator), the estimates of the cost of construction are as follows:

1. Telescopes — 40 7.5 m antennas, $\lambda/16$ at 850 μm		\$19.9M
2. Front ends — 4 frequency bands, 2 polarizations		8.0M
3. Correlator — 1 GHz bandwidth/polarization		9.0M
4. LO system		3.3M
5. IF system — 3 km baselines		2.5M
6. Computers — on-line and post-processing		6.0M
7. Site — antenna stations, transporters, etc.		3.3M
8. Buildings		3.0M
9. 20% Contingency		11.0M
	Total	\$66.0M

The cost equation is a sensitive function of the number of antennas in the array. Not only does the total cost of the antenna system depend on N, but so too, of course, does the total cost of the receivers, the IF system, etc. Figure 1 (lower panel) illustrates the dependence of construction cost on the number of antennas. Rather than simply plotting cost against numbers of antennas, it is perhaps more appropriate to plot cost against a quantity more closely related to 'scientific capability'. We have attempted to so by plotting, in the upper panel of Figure 1, the total cost as a function of imaging speed, N^2D^2. This curve is obtained from the minimum cost solution for antenna diameter, which was discussed in Chapter IV and illustrated in Figure 2 of that chapter (using $\alpha = 2.5$). Note here that the cost rises more steeply than N^2D^2, until $N^2D^2 \approx 7.5 \times 10^4$—beyond this point the two are proportional.

3. Uncertainties

This cost estimate should be viewed as a first cut at pricing the MMA. The biggest uncertainty is in the cost of the antennas. The antenna study planned for 1988 should give us a firmer number for the performance described here. If

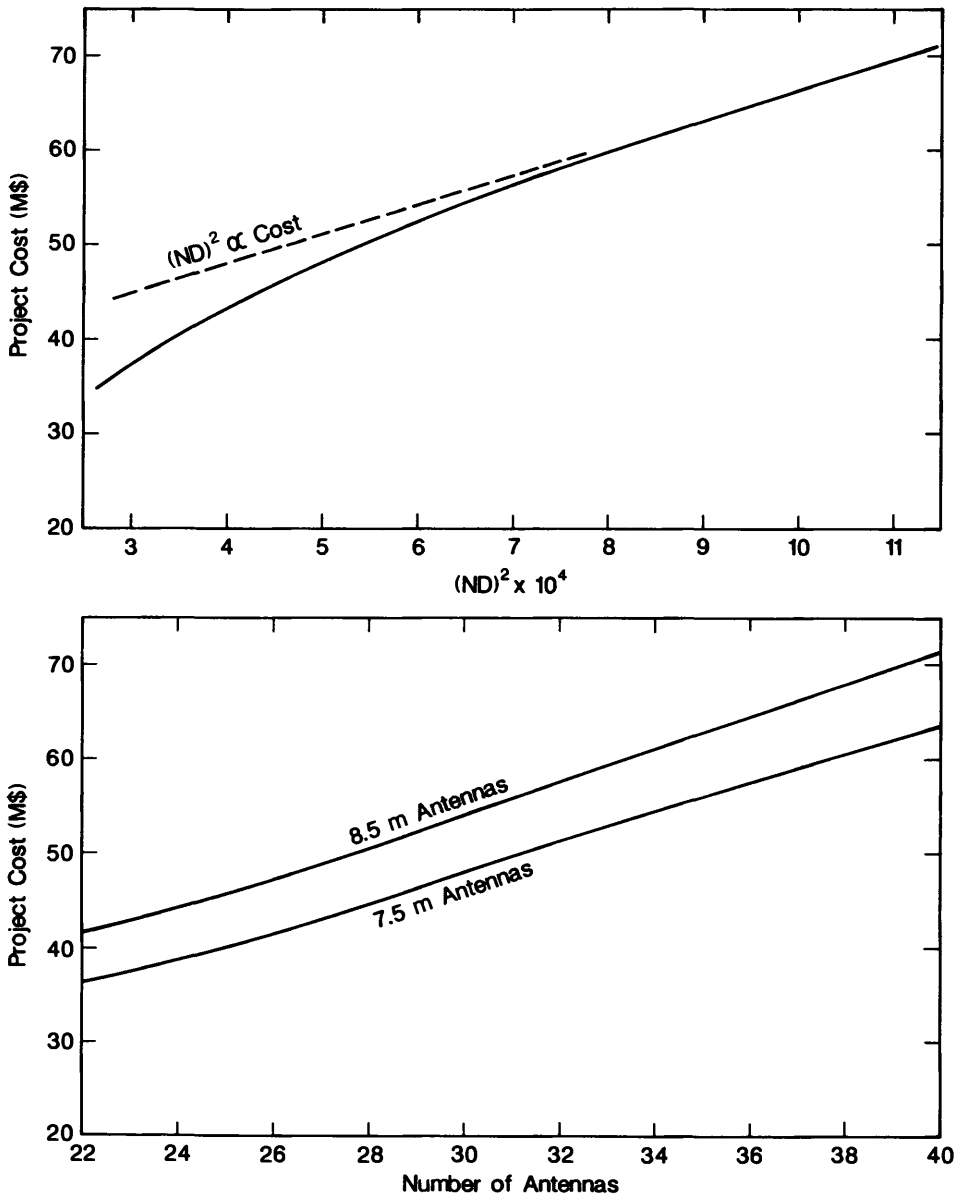


Figure 1. Total project cost, including contingency, as a function of imaging speed, $(ND)^2$. The curve shown here (upper panel) is the minimum cost solution, $\alpha = 2.5$, from Figure 2 of Chapter IV; thus, increasing $(ND)^2$ reflects increases in either N or D, or in both. The lower panel illustrates the total project cost for an array of 7.5 m or 8.5 m antennas as a function of the number of antennas in the array.

we extend the performance of the array through the submillimeter band, the cost of the antennas will double. Another additional cost would be that of a central element, if it is needed. Here this cost is not included. The site cost is obviously indeterminate.

VIII. Steps Toward an MMA Facility

In the previous seven chapters we have outlined the motivation, design considerations, and cost estimates for a national millimeter array. We have described the basic concept of a millimeter array on a high, dry site with a maximum baseline of about 3 km and movable antennas which can be configured in several smaller arrays. We have suggested forty 7.5 meter antennas as a working proposal.

Several questions remain to be studied. First—and foremost—we need to review our current ideas with the community. Is the current plan well matched to the science? How do the latest results from the existing millimeter interferometers affect the design of this instrument? Is a central element necessary? In addition, the antenna design needs a more careful study, in order to define the optimum practical solution for our general needs. The best available site needs to be located. The requirements for the optics need to be studied further.

In our approach toward the final goal of a U.S. national MMA facility, it is useful to have a timetable of target dates for the critical stages in approaching that goal. Although these dates, in common with some of the major elements of the design, will likely be modified over the next few years, this does aid in setting our priorities. The following table is a summary of our current schedule of milestone events:

January–December 1988	—Continue configuration studies
	—Undertake antenna design study
January–June 1988	—Perform cloud cover study
	—Commence stability tests
January 1988	—Receive final report from Working Group on the Central Element
February 1988	—Advisory Committee meeting
May 1988	—Second Scientific Workshop
December 1988	—Finalize all major elements of the design
July 1989	—Begin writing MMA proposal
December 1989	—Circulate MMA proposal
	—Select site
1990–1991	—Undertake detailed design studies
1992	—Begin major construction