MMA Memo. No. 243

Report of the Antenna Size Committee Meeting

Bryan Butler, Bob Brown - NRAO
Leo Blitz, Jack Welch - UCBerkeley
John Carlstrom - UChicago
Dave Woody - OVRO
Ed Churchwell - UWMadison

May 4, 1999

Introduction

Homogeneous array designs (antennas of common size) are currently preferred over heterogeneous designs (antennas of different sizes, e.g., 8 and 15 meters) for the combination MMA/LSA (hereafter referred to as the ALMA - Atacama Large Millimeter Array). Given that preference, it remains to be decided exactly which antenna diameter will yield the "best" ALMA. In an attempt to address this problem, a committee met in Berkeley on November 4, 1998. This report is a description of discussions which took place during that meeting and some subsequent analysis by the members of the committee.

It is not the intent of this memo to provide *the* answer to the question of which antenna diameter to use, but merely to point out and describe the arguments which should provide a framework within which a decision on antenna diameter can be made in an informed manner. These arguments include scientific, technical, and cost drivers for the ALMA.

Scientific considerations

The ALMA will be the principal instrument for astronomy at millimeter and submillimeter wavelengths in the first part of the next century. As such, it must be useful for much, if not all, of the future science that is planned for these wavelengths. These programs will impact all of astronomy. They include characterizing our solar system and the sun, understanding the birth of stars and planets, the late evolution of stars and their envelopes, the physical and chemical evolution of the ISM and stellar content of our galaxy and those nearby, and the evolution of the universe from the earliest times. For many of these programs, observations will need to be made, not only with high resolution, but also over very large fields (minutes of arc or more).

Single Pointed (High Resolution) Observations

It is clear that there is a class of objects which will be observed by the ALMA with single pointings. Such objects include some solar system bodies (NEA's and KBO's, e.g.), O stars and their associated winds, the photospheres of giant stars, observations of the very inner portions (including protostellar and protoplanetary disks) of star formation regions, and observations of the high z universe. For the high z observations, one is likely to mosaic first to form a finder map in the continuum, where the sensitivity of the array is already extremely high and one can detect L^* galaxies at any z. One would then image each galaxy in line emission with one pointing each (with a field of several 10's of kpc at $z > 1$, one pointing per galaxy will be sufficient).

Mosaiced (Wide Field) Observations

Two areas which are of particular current interest are star formation and the study of high red shift galaxies. Sites of star formation are the Myers and Benson cores many of which contain IRAS sources, evidently Young Stellar Objects. Recent interferometer surveys of a number of such cores by Looney and colleagues (ApJ, in press) have found that these typically contain multiple YSO's. Indeed, out of nine objects studied, all show multiple YSO's. Although there is a bias in the survey toward brighter sources, nine out of nine is a good statistic, and it is clear that star formation in small clusters must be an important way in which it happens. The clumps are isolated, and the processes appear approximately coeval in each clump. The YSO's have disks, adjacent envelopes, and overall envelopes. Understanding the formation process requires mapping the entire clump with its contents at the highest resolution. Clump sizes are typically .05 to .2 pc, which correspond to fields of one to several arcminutes. A recent image of HL Tau in ^{13}CO(1–0) made with the BIMA array plus single antenna data for the short spacings from the NRAO 12-m shows that there is much to learn from the study of large fields with high resolution. In this $200''$ diameter field, there is a clearly expanding bubble approximately centered on XZ Tau which dominates the kinematics of the region. Even larger fields are being mosaiced with interferometer resolutions. A recent map in the NGC 1333 cloud covers a region which is $6'$ by $8'$ in size, showing the structure of many bipolar flows with resolutions of a few arcseconds.

Observation of galaxies in the early universe is also an important area. Much of the exciting recent work has been focused on infrared galaxies at moderate red shift and lensed galaxies at z up to nearly 5. These have small angular extents and can be observed in a single pointing. On the other hand, the future may well be in the study of such regions as the Hubble Deep Field. In this field, approximately $2.5'$ in diameter, there are 3000 galaxies with a large range of red shifts. The mean separation is about $2.4''$. The spacing for 10% of the galaxies is about $8''$, and for 1% it is $25''$. One could imagine making a mosaic of the region at modest resolution (a few arcseconds) and then returning to image a number of interesting targets with single pointings.

Although a long integration of the HST was required to obtain this striking image, the NGST with excellent detectors operating in the near-IR with a larger mirror is expected to be at least 100 times faster than the HST. Thus, deep fields of many objects in many directions will be available in a few hours observation with the NGST, and the expected field size is 4 arcminutes. We should anticipate that this instrument may drive much of the science that will be interesting with the millimeter wave array. Another example of this point is the Eagle Nebula, imaged with the HST, which has been recently mapped at a few seconds resolution in both ^{13}CO(1–0) and ^{13}CO(2–1) by Mark Pound. The combination of the excellent optical image and the kinematic and density data that can be obtained with the interferometer array shows the importance of the combined study.

In the nearer neighborhood of our galaxy, interferometer images of the interacting galaxies NGC 6670 and Arp 293 cover fields of 1-2 arcminutes. Grand design spirals, such as M51 and NGC 5194 are extended over regions of several arcminutes.

Even closer to home, images of the envelopes of carbon stars, very old stars destined soon to transform into planetary nebulae, reveal large structures. An example is the star CIT6 which, at a distance of about 200 pc, shows an outer envelope that is at least $100''$ in diameter. It has a core-halo structure. Studies of the whole envelope are important in the understanding of the overall lifetimes of such objects. Measurements of isotopic ratios as a function of radius may reveal the history of the rapid nuclear chemistry that is expected to be in progress in these old stars.

Very large structures also include comets, the sun and moon, and whole molecular clouds. The planets, though not quite so large, will nevertheless have to be mosaiced at all but the longest wavelengths of the ALMA.

Larger antennas will force interferometric observations of the CMB to higher angular scales, i.e., higher ℓ. It is unlikely that ALMA total power observing (single dish mode) will be stable enough to recover the low ℓ data from mosaicing techniques with sufficient sensitivity, at least not without a considerable investment, and so ALMA CMB observations will focus on higher angular scale CMB observations (higher ℓ). Current CMB studies, however, are focused on much larger angular scales, and it is expected that most of this work will be done by the time ALMA is comissioned. ALMA operating at 1 cm will be most sensitive to $\ell > 5000$ to 10,000 for 8 to 15 m antennas where most of the CMB anisotropy is caused by secondary effects such as patchy reionization or the Sunyaev Zel'dovich effect from galaxy clusters and filamentary structures. For these observations the smaller telescopes are strongly preferred, although the larger diameter will still provide useful data.

Required ND^2 and ND

Since it is clear that both single pointing observations (which want ND^2 optimized) and mosaiced observations (which want ND optimized) will be undertaken with the ALMA, and we might envision that the division between the two may be roughly equal, we would like to be able to do both types of observations as well as possible. It therefore seems reasonable to investigate what the minimum desired values are for the quantities ND and ND^2.

Minimum desired ND^2 from high z

Present spectroscopic observations of high redshift objects are limited to those extremely bright (ultraluminous) galaxies, even such galaxies where the observed line flux density is boosted through gravitational lensing. The ALMA goal for spectroscopic studies of the early universe is to be able to image normal galaxies, L* galaxies or galaxies such as the Milky Way. Let us use the goal to be able to detect the Milky Way spectroscopically at redshifts greater than one as a means to set a requirement on ND^2 for the array. At cosmological redshifts the 10 kpc disk of the Milky Way is much smaller than the primary beam so ND^2 is the parameter we need to optimize.

The following sensitivity discussion is derived from the 1998 SPIE paper "Technical Specification of the Millimeter Array" by R.L. Brown; it is available via the MMA www library.

The flux density sensitivity, or rms noise in flux density units, is

$$\Delta S = \frac{4\,\sqrt{2}\,k\,T_{sys}}{\gamma\,\epsilon_q\,\epsilon_a\,\pi\,D^2\sqrt{n_p[N(N-1)/2]\Delta\nu\Delta t}} \quad ,$$

where ϵ_a is the aperture efficiency, ϵ_q is the correlator quantization efficiency, n_p is the number of simultaneously sampled polarizations, and γ is a gridding parameter that we shall set equal to unity. Now simplify this by setting $N(N-1) = N^2, n_p = 2$, and $\epsilon_q = 0.95$. For spectroscopic observations of extragalactic observations we'll use a velocity channel width Δv and write $\Delta\nu = \nu\Delta v/c$. We can then simplify the above expression for ΔS to

$$\Delta S = \frac{45.3\;T_{sys}}{\epsilon_a\,ND^2\sqrt{\nu_{GHz}\Delta v_{km/s}\Delta t}} \quad \text{Jy.}$$

The aperture efficiency is:

$$\epsilon_a = \epsilon_0\,e^{-\left(\frac{4\,\pi\,\sigma}{\lambda}\right)^2} \quad ,$$

which for the goals of the ALMA ($\sigma = 25\,\mu$m, $\epsilon_0 = 0.80$) takes the values shown in table 1 over the range of frequencies planned for the initial ALMA.

We refer T_{sys} to a point outside the terrestial atmosphere and compute it as:

$$T_{sys} = T_{rx}\,e^{\tau_0 A} + \epsilon_l\,T_{atm}\,\left(e^{\tau_0 A} - 1\right) + (1-\epsilon_l)\,T_{sbr}\,e^{\tau_0 A} + T_{cmb} \quad ,$$

where ϵ_l is the fraction of the antenna power that is received in the forward direction (i.e., the fraction that is on the sky in the main lobe and all the forward sidelobes), T_{atm} is the effective atmospheric temperature and T_{sbr} is the temperature onto which the spillover falls. The terms of T_{sys} represent the contributions from the receiver, the sky, the "antenna", and

Table 1: Aperture Efficiency of the ALMA Antennas

Frequency (GHz)	Wavelength (μm)	ϵ_a
115	2600	0.79
230	1300	0.75
345	870	0.70
409	650	0.63
675	440	0.48

the CMB. We assume that $T_{sbr} = T_{amb}$ where T_{amb} is the ambient surface temperature. We assume that $T_{atm} = 70.2 + 0.72 T_{amb}$ (Bevis *et al.* 1992), which has been verified to be a fairly accurate representation of the effective atmospheric temperature by comparison to detailed atmospheric emission models. We assume $T_{amb} = 269$ K, the average surface temperature at the Chajnantor site. For simplicity we will not correct temperatures to radiation temperature by the Planck function.

The goal for the MMA SIS receivers is that they be image separating receivers with a receiver temperature in the chosen sideband (i.e., the SSB noise temperature) of 6 times the photon limit. The unwanted sideband is terminated at 4K. This gives:

$$T_{rx} = \frac{6\, h\, \nu}{k} + 4 \quad .$$

For the second two terms we adopt the MMA antenna goal of $\epsilon_l = 0.95$, i.e., 95% of the received power comes from the forward direction. We will compute T_{sys} at an airmass of 1.3 (50° elevation) and use for the frequency dependent optical depths on the Chajnantor site the opacities produced by a model atmosphere for the site which has 1.5 mm of precipitable water. Note that the opacities produced by this model agree very well with those measured by Matsuo *et al.* (1998).

The terms in the T_{sys} equation above, along with the resultant T_{sys} are shown in table 2. Note that these numbers agree very well with those of Jewell and Mangum (1997) when a common set of assumptions is used.

We can now combine the results for T_{sys} together with the equation for ΔS above to compute the spectral sensitivity of the ALMA:

$$\Delta S = \frac{\Delta S_0}{N D^2 \sqrt{\Delta v_{km/s} \Delta t_s}} = \frac{\Delta S(75\mathrm{km/s}; 6\mathrm{h})}{N D^2} \quad ,$$

where the second line follows from the first by using a spectral bandwidth of 75 km/s and an integration time of 6 hours. Note that ΔS_0 and $\Delta S(75\mathrm{km/s}; 6\mathrm{h})$ both include the aperture efficiency. We tabulate these quantities as a function of frequency in Table 3.

If we wish to detect spectral line emission in a galaxy that is similar to the Milky Way but at redshifts of one or greater then the most obvious candidate to observe is CO. The

Table 2: Estimated T_{sys} for the ALMA at elevation=50°

Frequency (GHz)	τ_0	$T_{rx}\, e^{\tau_0}$	$\epsilon_l\, T_{atm}\, (e^{\tau_0} - 1)$	$(1 - \epsilon_l)\, T_{sbr}\, e^{\tau_0}$	T_{sys}
95	0.017	32.05	5.56	13.75	54.05
110	0.035	37.33	11.62	14.07	65.72
230	0.057	75.62	19.25	14.48	112.06
345	0.205	134.86	76.46	17.55	231.57
409	0.406	206.45	174.32	22.80	406.28
675	1.340	1131.73	1179.53	76.73	2390.69
875	1.412	1605.03	1321.29	84.34	3013.35

Table 3: Spectral Flux Density Sensitivity Coefficients

Frequency (GHz)	ΔS_0 (Jy)	ΔS(75km/s; 6h) (Jy)
95	321	0.252
110	363	0.286
230	448	0.352
345	813	0.639
409	1382	1.086
675	8687	6.825
875	13517	10.620

total CO luminosity of the Milky Way in the 1–0 transition has been estimated by Solomon and Rivolo (1989). This luminosity agrees roughly with the CO luminosities seen in the higher transitions by COBE (Bennett *et al.* 1994; Wright *et al.* 1991). Note that COBE also measured emission from neutral and ionized carbon (CI and CII) and ionized nitrogen (NII). We presume here that the desire is to measure the neutral species, so concentrate on CO. Given the luminosity of the CO 1–0 transition, we can calculate the expected received flux density in any transition as (see e.g. Solomon *et al.* 1992):

$$S_{CO} = 3.08 \times 10^{-8} \frac{L'_{CO}\, \nu_{rest}^2\, (1+z)}{\Delta v_{rest}\, d_L^2} \quad ,$$

S_{CO} is the flux density in Jy, L'_{CO} is the CO luminosity in K km/s pc^2, ν_{rest} is the rest frequency of the transition in GHz, d_L is the "luminosity distance" in Mpc, and Δv_{rest} is the rest line width in km/s. The luminosity distance can be written (Weinberg 1972):

$$d_L = \frac{c}{H_o\, q_o^2} \left[z q_o + (q_o - 1) \left(-1 + \sqrt{2 q_o z + 1} \right) \right] \quad .$$

We use $L'_{CO} = 5 \times 10^8$ K km/s pc^2 for the 1–0 transition, which is slightly larger than that in Solomon and Rivolo (they give 3.7×10^8), but is consistent with the COBE results. We then modify the luminosity as a function of transition and redshift according to a model which accounts for the proper radiative transfer given the higher background temperature at higher

z, and assuming that 90% of the CO is in clouds similar to our galactic dark clouds and 10% is in regions similar to strong PDR's (H. Liszt provided this calculation). We assume $H_o = 75$ km/s/Mpc, $q_o = 0.5$, and the intrinsic width of the lines is $\Delta v = 300$ km/s (Solomon *et al.* 1997). We can then calculate, given any array collecting area, the maximum z to which any of the transitions of CO (we calculate up to 8–7) can be detected by that array. We demand a 4–sigma detection in a 75 km/s channel in 6 hours of integration. With these requirements, and the additional requirement that at least 3 transitions must be observable, the following relationship between collecting area and maximum detectable z is derived:

$$z_{max} \sim \frac{ND^2}{3400\,\mathrm{m}^2} \quad .$$

Thus, to reach $z = 2$, an ND^2 of 6800 m^2 is required. This is satisfied with an array of 68 10-m antennas or 47 12-m antennas. We point out, however, that reducing ND^2 to 6400 (64 10-m or 45 12-m antennas) allows for a z_{max} of 1.9, essentially the same as $z = 2$. Of course, larger values of ND^2 are always desirable, as they would allow us to resolve the line flux density into more pixels (higher angular or spectral resolution) or image to higher S/N more quickly.

We note finally that lines of CI, and redshifted lines of NII and CII will also be observable and will provide important probes of the IMF and the Lyman continuum luminosity from the most luminous stars in early galaxies. However, because so little is know about the luminosity of these lines as a function of redshift in galaxies of differing Hubble type we have not used them here to help us place limits on the needed ND^2 of the array.

Minimum desired ND^2 from protoplanetary disks

Guilloteau (1996) presents an argument on the minimum desirable value of ND^2. We duplicate that argument here, with updated numbers, and a slightly different formulation for the noise flux density. The noise flux density of an interferometric array is given by:

$$\Delta F = \frac{\sqrt{2}\,k\,T_{sys}}{\eta_{sys}\,A\,\sqrt{N_b\,\Delta\nu\,\Delta t}} \quad ,$$

where T_{sys} is the system temperature, η_{sys} is the total system efficiency (including aperture efficiency, correlator (quantization) efficiency, and any other efficiency factors), A is the physical antenna area, N_b is the number of baselines ($\sqrt{N_b} \sim N/\sqrt{2}$ for large N), $\Delta\nu$ is the total bandwidth over all polarizations observed simultaneously, and Δt is the total on-source time.

If we assume that the fluctuations that we are to observe are imposed on a background brightness temperature for which the Rayleigh-Jeans approximation is appropriate, then the noise flux density is related to the brightness temperature fluctuations (ΔT) via:

$$\Delta F = \frac{2\,k\,\Delta T}{\lambda^2}\,\Omega_s \quad .$$

where Ω_s is the solid angle of the synthesized beam. For an image which is restored with a circular gaussian of width θ_s (e.g., the result of CLEAN or relatives), this solid angle is given by:

$$\Omega_s = \frac{\pi}{4\,\ln 2}\,\theta_s^2 \sim \frac{\pi}{4\,\ln 2}\,\frac{\lambda^2}{B_{max}^2} \quad ,$$

for maximum baseline B_{max} at wavelength λ.

Making all of the appropriate substitutions, and assuming that the total system efficiency is about 0.7, we end up with:

$$\Delta T \sim 1.5 \frac{B_{max}^2\,T_{sys}}{ND^2\,\sqrt{\Delta\nu\,\Delta t}} \quad .$$

Note that this is nearly identical to equation 2 of Guilloteau (1996) (substitute $\theta_s \sim \lambda/B_{max}$, and $\theta_p \sim \lambda/D$).

Now, consider an observation of the gas distribution and kinematics in a protoplanetary disk. A velocity resolution of 0.3 km/s might be desired, implying $\Delta\nu \sim 110$ kHz per polarization at a center frequency of 110 GHz. Assume 2 polarizations, and assume that we have 24 hours of on-source time, implying $\Delta t = 86400$ s. Assume also a maximum baseline of 3 km. This leaves us:

$$\Delta T \sim \frac{100\,T_{sys}}{ND^2} \quad .$$

For an observation near 110 GHz, the system temperature during median conditions is ~ 65 K (see Table 2), leaving:

$$\Delta T_{115} \sim \frac{6500}{ND^2} \quad .$$

Given a desired ΔT of 1 K, which should be sufficient to image marginally optically thick lines, we then have the criterion:

$$ND^2 \gtrsim 6500 \quad .$$

For an observation near 230 GHz, we have for median conditions $T_{sys} \sim 112$ K, leaving (remembering that the bandwidth is now a factor of 2 larger for the same velocity resolution):

$$\Delta T_{230} \sim \frac{7800}{ND^2} \quad .$$

For $ND^2 = 6400$, this gives a brightness temperature sensitivity of ~ 1.2 K. Thus, it seems that $ND^2 \geq 6400$ is a good goal to attempt to reach, from both the high z and the protoplanetary disk observation arguments. This is satisfied with an array of 64 10-m antennas, or 45 12-m antennas.

Minimum desired ND

Holdaway (1998) assesses the ability of the MMA to provide precision imaging in all the array configurations and at all the frequency bands planned for the MMA. Here the consideration for imaging is the ability of the instrument to achieve good coverage of the (u,v)-plane. Morita and others have emphasized that excellent imaging, imaging limited by dynamic range, can be achieved when 50% of the (u,v)-cells are filled. Morita calls this quantity FOCC, the "fraction of occupied cells." It is calculated by simulating (u,v)-points taken in short integrations by the array in a particular configuration and gridding them onto the Fourier plane with a cellsize equal to the antenna diameter. The fractional area of the gridded (u,v)-plane out to the diameter of the longest array baselines that is filled by observations is the quantity FOCC. Clearly, FOCC is a function of hour angle that asympotically approaches a value of one.

For the three shorter MMA configurations–the compact array(s), the 240-m and 860-m arrays, the MMA achieves FOCC=0.5 in less than two hours of observations. But for the longest array, the 3000-m diameter A-array, FOCC=0.5 can be realized only with 8 hours of hour-angle tracking. Such long tracks can only be obtained at the expense of observations being made well off the meridian at increasingly low elevation angles. At low elevation the sensitivity is compromised by the increasing atmospheric contribution to the system temperature making the weights of the (u,v)-points so obtained very low. This compromises the fidelity of the image and reduces its signal-to-noise. The situation worsens at higher frequencies where the atmospheric contribution to T_{sys} is larger. The ALMA, however, should have a sufficiently large number of antennas such that FOCC=0.5 can be achieved even in the highest resolution 3000-m array with suitably short hour-angle tracks taken near the meridian where the sensitivity always remains high.

In Figure 2 of Holdaway (1998), the FOCC as a function of hour angle for the 3 longer MMA configurations is shown. As hour angle increases all these curves asympotically approach one as an increasing number of the gridded (u,v)-cells are sampled redundantly. However, as Holdaway notes, the non-redundant FOCC (i.e., the FOCC for "snapshots") for each of the arrays is approximately proportional to

$$\mathrm{FOCC} \propto \left(\frac{ND}{B_{max}}\right)^2 \quad ,$$

where B_{max} is the maximum baseline length in the array. "Hence, to achieve optimal Fourier plane coverage, one wants to optimize the so-called 'collecting length' ND" (Holdaway 1998).

The function that is plotted in Figure 2 of Holdwaway (1998) for FOCC as a function of hour angle for the 3000-m array has the form

$$\mathrm{FOCC} = 1 - e^{-10\,h\,(ND/3000)^2} \quad ,$$

where h is the hour angle and we've explicitly used $B_{max} = 3000$. To achieve the precision imaging for which FOCC$\geq$0.5 is the figure of merit using an array of given collecting length (ND) will require observations made out to the hour angle limits shown in Table 4.

As noted above, observations taken at large hour angles will be given low weight in the imaging owing to the increase in system temperature at low elevation. To avoid corrupting

Table 4: Hour Angle Limits to Achieve FOCC=0.5 for an Array of N Antennas of Diameter D(m) in a Configuration with B_{max}=3000 m.

ND	h
300	6.9
350	5.1
400	3.9
450	3.1
500	2.5
550	2.1
600	1.7
650	1.5
700	1.3
750	1.1

the image with such low weight points, let us restrict our observations to hour angles such that the lowest weight points are reduced from those on the meridian by no more than $\sqrt{2}$. Let us also retain the opportunity to do such imaging in the submm. Then in median meteorological conditions on the Chajnantor site (Figure 4b of Holdaway [1998]) we conclude that we need to observe out to a limiting hour angle range of no more than approximately h=2.0.

With the restriction that $h < 2$ hours, we can return to Table 4 above and conclude that the specification for the "collecting length" ND for the ALMA needs to be approximately $ND \geq 560$. A joint array of 56 10-m antennas or 47 12-m antennas will satisfy this specification.

Conclusions from science considerations

1. It is clear that there will be single pointed observations done with the ALMA. These observations want ND^2 optimized.

2. It is also clear that a significant fraction of observations with the ALMA will involve mosaicing. In particular, some fields which have traditionally been considered single pointing fields will require mosaicing for correct interpretation (e.g., star formation regions). These observations want ND optimized.

3. A vast majority of the science considered so far could be done with any size antenna from 8-m to 15-m, given that it meets specifications (including surface, pointing, and total power capability).

4. ND should be $\gtrsim$ 560 m if possible.

5. ND^2 should be $\gtrsim$ 6400 m^2 if possible.

Technical considerations

In this section, we consider technical considerations influencing the decision on antenna diameter. These include antenna design and construction issues, antenna performance issues (e.g. calibration of several flavors), and other related issues (e.g., number of receivers).

Antenna design and construction

It appears clear that meeting the required telescope specifications (including pointing, surface rms, and slew rate) could be done relatively easily for an 8-m antenna. With a traditionally designed 10-m telescope, meeting these specifications will be difficult. A traditionally designed 12-m telescope will probably not meet the design specifications. Because of this, it seems counterproductive to keep modifying a traditional design to meet the specifications. Rather, it seems better to start from the ground up with designs that avoid the difficulties encountered by the more traditional designs. When starting from the ground up with well understood, but admittedly untested, designs, it seems that the specifiations could be easily met at 10-m and likely met at 12-m. At 12-m, however, there would be very little margin for error or manufacturability, i.e., no design contingency. It seems unlikely that a 15-m design of any type will meet the design specifications. Discussions of various designs and attempts to meet the specifications exist in the MMA memo series (Plathner 1999; Plathner *et al.* 1999; Anderson 1999; Woody and Lamb 1999; Lugten *et al.* 1999).

Number of receivers

As the antenna diameter decreases, the number of antennas will increase, thereby increasing the total number of all related items: receivers, amplifiers, LO's, etc... It has been perceived by some that the maintenance of large numbers of such systems will be impractical at best, and fail at worst. It seems, given the large number of systems now being considered for any of the reasonable antenna diameters, that this is less of an issue. There will be large numbers of systems in any case, and we must figure out how to deal with that.

Size of correlator

With smaller diameters, and more antennas, the size of the correlator also increases (quite rapidly). The size of the correlator should certainly be of concern, but it seems that it will not limit in any fundamental way the number of antennas up to some reasonable limit (which NRAO engineers estimate to be of order 100 or more).

More room in receiver cabin

While the larger antennas may have slightly more room in the receiver cabin, the difference does not favor any antenna size strongly.

Pointing - including pointing calibration

It is important to realize that blind pointing at 1/30 of a beam is not necessary. Only offset pointing with that accuracy is needed. While it is true that large submm telescopes such as the CSO and JCMT have not achieved this dead recogning specification, they do, in fact, meet the offset pointing specification. The argument has been made in the past that larger dishes imply that a weaker pointing calibrator can be used, implying that the pointing calibrator will be closer for larger antennas. This is in fact true, but makes little difference. The difference in distance to a pointing calibrator for 10-m or 12-m antennas is much less than a degree. Therefore, the difference in the quality of the derived pointing solution should be very small. Also, since offset pointing is envisioned to occur only every 15 minutes or so, the small difference in time to go an extra 0.1 degree or so on the sky is negligible.

Calibration

The argument has been made that larger antenna diameter means that a calibrator of smaller flux density can be used for secondary calibration. Again, this is true, but makes little practical difference, since it just means that the angular distance to a usable calibrator increases by some very small value (M. Holdaway estimates that this difference is 0.15 deg for 10-m vs. 12-m antennas).

Anomalous refraction

While anomalous refraction will be a nuisance (and perhaps worse) to deal with, it seems that this effect has no particularly strong bearing on the antenna size discussion. This is because the errors caused by this atmospheric phenomena are essentially random, and thus are less important than constant pointing offsets (see e.g., Holdaway and Woody 1999; Holdaway 1997).

Under-illumination

It has been suggested that larger dishes could simply be under-illuminated at the higher frequencies, in order to essentially relax some of the antenna specifications. This does not seem to be a good alternative, and should be discouraged.

Short spacings

Measurement of short spacings (all the way to the zero spacing) is absolutely critical for the success of the science to be done with the ALMA. Cornwell *et al.* (1993) have argued that it is more desirable to have these short spacings measured by the antennas of the array themselves, rather than a larger central element. The specification that the antennas of the ALMA have very sensitive total power capability is perceived to be possibly hard to achieve, and of great import.

Lowest resonant frequency

In order to be able to satisfy the fast switching requirement for the ALMA antennas, the antennas will be driven with large accelerations. In order for these accelerations to not excite resonant modes in the antennas, we would like the lowest resonant frequency to be as high as possible. It becomes increasingly difficult to avoid this problem as the diameter of the antenna increases. This does not seem to be an overriding consideration in the decision on antenna diameter, but must be considered nonetheless.

Power

When driving the antennas for the fast switching, large amounts of power will be necessary to drive the motors. Larger antennas require more power, as the required torque scales roughly as D^3. However, this does not seem to be an overriding consideration, as large amounts of power will be consumed in any case with the large numbers of antennas now being considered for the ALMA.

Conclusions from technical considerations

Given the above points, it seems that the largest antenna which should be considered is of order 12-m. It is not thought that larger antennas can be constructed which will meet the demanding specifications. Of the 8-m, 10-m, and 12-m diameter antennas, it is clear that the 8-m and 10-m can be designed, and most likely constructed, to meet the specifications. The 12-m is somewhat riskier. The issue is then the tradeoff between the risk of not meeting the specifications with the 12-m antenna vs. the problems associated with large numbers of elements/receivers/etc...

Cost considerations

In this section, we attempt to summarize the costing information that we presently have available for the MMA and to extend that information to assemble an estimate for the cost of the instrument as augmented by the merger with the LSA project (to form the ALMA). The costing estimates in all areas are incomplete and they are likely to remain uncertain through the completion of the Design and Development phase of the project. The uncertainties we've attempted to accommodate by use of contingencies assigned, separately, to each major area of the project. The contingencies will be a permanent part of the project management, although not necessarily at the magnitudes used here. Nevertheless it is important to emphasize at the outset that one should not be tempted to 'spend' the contingency; the contingency is a real cost item not to be regarded differently than the cost of antennas, receivers or any hardware part of the array. Clearly as we make progress in designing the instruments, and defining how all the various pieces of the instrument will be built in production, the contingency can be reduced and the money "saved" in this way can be reallocated within the project. However at each step in the project we need to acknowledge the uncertainty in our cost estimates, and associate a cost with that uncertainty: this is the role of contingency. At the moment the cost estimates in many areas are not reliable and we need to budget contingency to cover that uncertainty.

The cost estimate will be greatly refined over the next few months, in particular by means of a proper 'roll up' of the costs defined through the work breakdown structure of the entire project. At the moment that's not available. Hence the effort below is our own attempt at costing–it is not an 'official' MMA project estimate, it is not the NRAO estimate or the NSF estimate. Rather it is used to help us understand the distribution of costs within a project like the MMA and to guide the merger discussions in the near future.

Costed tasks

We will divide the array tasks to be costed into the following eleven categories:

1. Site
2. Antennas
3. Transporter
4. Electronics at the antennas
5. Central electronics
6. Correlator
7. Monitor and Control
8. Computing
9. System Engineering
10. Management
11. Long Baselines (the 10 km array)

A cost equation is parameterized in terms of the number of antennas in the array, N, the diameter of each of those antennas, D, and the number of frequency bands, N_ν.

1. Site

We use the term "site" to refer to activities on the Chajnantor site itself and for those functions to be located at the array Operations Support Facility (OSF) near the village of San Pedro de Atacama. In both these locations the costs come from a very large number of civil works tasks - roads, power and power distribution, water, buildings and so forth. This information and the associated costs are being compiled now in collaboration with engineering and construction firms in Chile. Refined numbers should be available very soon. Note that the Chajnantor costs include the antenna foundations, and all the site civil works. We use 15% contingency on all of these items. Our current best estimate of these costs can be expressed as:

Chajnantor site: $\$[13.0 + 0.325\ N]$M
OSF: $5.3M
Contingency: $\$[0.15\ (18.3 + 0.325\ N)]$M
Site Total: $\$[21.045 + 0.37375\ N]$M

2. Antennas

The best costing information we have is for a 10-m antenna which we believe in production will cost about $2.0M. We assume that the cost of antennas of other diameters scales from the 10-m cost with the 2.5 power of the diameter. We include in the antenna cost the cost of shipping and installation in Chile and for this purpose we adopt a figure of $75k for each 10-m antenna. Larger antennas will be more costly to ship and handle so we scale this latter cost by the square of the diameter. Finally, the antenna cost is very uncertain and will remain so until we have signed the contract for the production run of antennas and verified that the manufacturer can indeed deliver production antennas that meet the specifications. For this reason we adopt 30% contingency for the 10-m antenna and scale that uncertainty with antenna diameter (i.e., a larger contingency for larger antennas). This gives the antenna cost equation and the equation for antenna contingency as:

Antennas: $\$[0.0063\ D^{2.5}N + 7.5 \times 10^{-4} ND^2]$M
Contingency: $\$[0.0063\ D^{2.5}N(0.03D) + 7.5 \times 10^{-4} ND^2(0.15)]$M
Antennas Total : $\$[0.0063 D^{2.5}N(1 + 0.03D + 0.1369/\sqrt{D})]$M

3. Transporter

We assume that one transporter will be needed for each 15 antennas and each will cost $600k. We use 15% contingency, leaving:

Transporter: $\$[0.04\ N]$M
Contingency: $\$[0.04\ N\ (0.15)]$M
Transporter Total : $\$[0.046\ N]$M

4. Antenna Electronics

We assume that each antenna has one 4K cryogenic dewar that costs \$100k. We use \$75k for the cost of IF switching and monitor/control circuitry. We use \$20k for the LO required at the antenna for each frequency band and use \$75k for each (dual polarization) frequency band including the cost of the control and IF instrumentation. Finally we use 30% as the contingency for this item to reflect the unavoidable uncertainty associated with instrumentation such as this which will be built in-house. This gives:

Antenna Electronics: $\$[(0.175 + 0.095N_\nu)N]$M
Contingency: $\$[(0.175 + 0.095N_\nu)N(0.30)]$M
Antenna Electronics Total: $\$[(0.2275 + 0.1235N_\nu)N]$M

For the MMA the parameter N_ν is 9. We will use this number below but remember that it enters here explicitly and could be changed if that is desired.

5. Central Electronics

This includes the interface electronics to each antenna and the IF processing of 16 Ghz of returned bandwidth from each antenna. We use 15% contingency, leaving:

Central Electronics: $\$[0.150\ N]$M
Contingency: $\$[0.0225\ N]$M
Central Electronics Total: $\$[0.1725\ N]$M

6. Correlator

A conceptual design for the MMA correlator exists and has been costed. This was done for the 40 element MMA. We adopt 20% contingency for that correlator and scale the contingency with N for larger correlators such that the contingency doubles with $N = 80$. This gives:

Correlator: $\$[0.25N + 5 \times 10^{-4}N^2]$M
Contingency: $\$[(0.25N + 5 \times 10^{-5}N^2)(0.20N/40)]$M
Correlator Total: $\$[N(0.25 + 0.00175N + 2.5 \times 10^{-6}N^2)]$M

7. Monitor and Control

The costs for M/C are costs associated with the number of antenna stations (foundations). We assume that the 5 MMA arrays out to 3000m use a total of 4N stations assuming that many of the stations can be used for more than one array. We associate \$50k costs with each station and use 15% contingency:

Monitor and Control: $\$[0.20N]$M
Contingency: $\$[0.03N]$M
Monitor and Control Total: $\$[0.23N]$M

8. Computing

This is a level of effort task with a large contingency. We assume it requires 50 work-years at a cost of \$100k for each work-year and add to that \$5M in array computing equipment to be purchased over the construction phase of the project. We use 25% contingency on the lot. None of this depends in any significant way on N or D. So, we have:

Computing: \$10M
Contingency: \$2.5M
Computing Total: \$12.5M

9. System Engineering

We use a total of 20 work-years for the system engineer and his or her staff and cost that at \$100k per person year. Contingency will be 15%. None of this depends in any significant way on N or D. So, we have:

System Engineering: \$2.0M
Contingency: \$0.3M
System Engineering Total: \$2.3M

10. Management

We assume 50 work-years at \$100k per year and add to that the project travel and communications budget of \$5M over the construction phase of the project. The contingency will be 15%. None of this depends in any significant way on N or D. So, we have:

Management \$10M
Contingency: \$0.15M
Management Total: \$11.5M

11. Long Baselines (the 10km array)

The 30,000 m circumference of this array needs to be connected with salt-stabilized roads (not asphalt), it needs power cable, IF and communications cables, power transformers etc. This leaves:

Long Baselines: $\$[1.8 + 0.115\ N]$M
Contingency: $\$[(1.8 + 0.115\ N)(0.15)]$M
Long Baselines Total: $\$[2.07 + 0.13225\ N]$M

Spares

We cost spares at 5% of the cost of items 2-7 above, and 2.5% of item 8. Spares will not be provided in other areas.

Base costs as a function of N

The antenna diameter D appears only in the antenna cost term. We will therefore compute below the "base costs", contingency for the "base costs" and spares also for the "base costs" of the array as a function of N without the antenna term. After that we will cost the antennas as a function of N and D and then finally combine all the costs. Table 5 shows the base costs as a function of N. We will separately account for the contingency in each of these categories; the result is presented in Table 6. Finally, to complete the base cost without antennas we present in Table 7 the costs budgeted for spares.

Table 5: Base costs, without antennas, as a function of N

N	40	45	50	55	60	65	70
Site	31.3	32.9	34.6	36.2	37.8	39.4	41.1
Transporter	1.6	1.8	2.0	2.2	2.4	2.6	2.8
Antenna Electronics	41.2	46.3	51.5	56.6	61.8	67.0	72.1
Central Electronics	6.0	6.8	7.5	8.3	9.0	9.8	10.5
Correlator	10.0	12.3	13.7	15.3	16.8	18.4	20.0
Monitor and Control	8.0	9.0	10.0	11.0	12.0	13.0	14.0
Computing & Post proc	10.0	10.0	10.0	10.0	10.0	10.0	10.0
System Engineering	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Management	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Long Baselines	6.4	7.0	7.6	8.1	8.7	9.3	9.9
SUBTOTAL	127.3	138.1	148.9	160.7	170.5	181.5	192.4

Table 6: Contingency on base cost, without antennas, as a function of N

N	40	45	50	55	60	65	70
Site	4.7	4.9	5.2	5.4	5.7	5.9	6.2
Transporter	0.2	0.3	0.3	0.3	0.4	0.4	0.4
Antenna Electronics	12.4	13.9	15.5	17.0	18.5	20.1	21.6
Central Electronics	0.9	1.0	1.1	1.2	1.4	1.5	1.6
Correlator	2.1	2.8	3.4	4.2	5.0	6.0	7.0
Monitor and Control	1.2	1.3	1.5	1.6	1.8	2.0	2.1
Computing & Post proc	2.5	2.5	2.5	2.5	2.5	2.5	2.5
System Engineering	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Management	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Long Baselines	1.0	1.0	1.0	1.0	1.0	1.0	1.0
SUBTOTAL	26.8	29.6	32.4	35.2	38.4	41.6	44.7

Table 7: Spares cost on the base, without antennas, as a function of N

N	40	45	50	55	60	65	70
Site	-	-	-	-	-	-	-
Transporter	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Antenna Electronics	2.1	2.3	2.6	2.8	3.1	3.3	3.6
Central Electronics	0.2	0.3	0.3	0.3	0.4	0.4	0.4
Correlator	0.5	0.6	0.7	0.8	0.8	0.9	1.0
Monitor and Control	0.4	0.4	0.5	0.5	0.6	0.6	0.7
Computing & Post proc	0.3	0.3	0.3	0.3	0.3	0.3	0.3
System Engineering	-	-	-	-	-	-	-
Management	-	-	-	-	-	-	-
Long Baselines	-	-	-	-	-	-	-
SUBTOTAL	3.6	4.0	4.5	4.8	5.3	5.6	6.1

Now we examine the antenna costs as a function of N and D. Here we separate the cost of procuring and shipping/delivering the antennas from the costs being carried as contingency and the costs allocated to spares. All of this information is compiled in Table 8 for antennas of diameter 8-m, 10-m, and 12-m.

Table 8: Antenna related costs as a function of N and D

N	40	45	50	55	60	65	70
Base Costs							
8-m	47.5	53.4	59.4	65.3	71.3	77.2	83.2
10-m	82.7	93.0	103.3	113.7	124.0	134.4	144.7
12-m	130.0	146.3	162.5	178.8	195.0	211.3	227.5
15-m	226.3	254.6	282.9	311.2	339.5	367.8	396.1
Contingency Costs							
8-m	11.2	12.6	14.0	15.4	16.8	18.2	19.7
10-m	24.3	27.4	30.4	33.5	36.5	39.5	42.6
12-m	45.9	51.6	57.3	63.1	68.9	74.6	80.3
15-m	99.8	112.3	124.8	137.3	149.7	162.2	174.7
Spares Costs							
8-m	2.4	2.7	3.0	3.3	3.6	3.9	4.2
10-m	4.1	4.6	5.2	5.7	6.2	6.7	7.2
12-m	6.5	7.3	8.1	8.9	9.8	10.6	11.4
15-m	11.3	12.7	14.1	15.6	17.0	18.4	19.8

In Table 9 we accumulate all the costs for arrays of 8-m, 10-m, 12-m and 15-m antennas.

To interpret all these tables we explore two cases: assume that we knew today that we had available to us total funding for the project from all sources of \$350M or \$400M and that we were to be held to that funding amount (i.e., we will not allow cost overruns). In that case we would feel reasonably confident signing up to build, with these constraints, any

Table 9: Array costs, accumulating all costs

N	40	45	50	55	60	65	70
Base Costs							
Base (without Antennas)	127.3	138.1	148.9	160.7	170.5	181.5	192.4
Contingency on the Base	26.8	29.6	32.4	35.2	38.4	41.6	44.7
Spares for the Base	3.6	4.0	4.5	4.8	5.3	5.6	6.1
SUBTOTAL (Base)	157.7	171.7	185.8	199.9	214.2	228.6	243.1
8-m Base & Antennas							
Subtotal Antenna	61.1	68.7	76.3	84.0	91.7	99.3	107.1
TOTAL (8-m)	218.9	240.5	262.2	284.0	305.9	328.0	350.1
10-m Base & Antennas							
Subtotal Antenna	111.1	125.0	138.9	152.9	166.7	180.6	194.5
TOTAL (10-m)	268.9	296.8	324.7	352.8	381.0	409.3	437.7
12-m Base & Antenna							
Subtotal Antenna	182.4	205.2	227.9	250.8	273.7	296.5	319.2
TOTAL (12-m)	340.2	376.9	413.8	450.8	487.9	525.1	562.4
15-m Base & Antenna							
Subtotal Antenna	337.5	379.7	421.9	464.1	506.2	548.4	590.6
TOTAL (15-m)	495.2	551.4	607.6	664.0	720.5	777.0	833.7

of the arrays with the parameters listed below in Tables 10 or 11.

Table 10: Possible arrays with total available funding of \$350M

D	N	ND	ND^2	Array Area (m^2)
8-m	70	560	4480	3520
10-m	54	540	5400	4241
12-m	41	492	5904	4637
15-m	27	405	6075	4771

We point out here that in an entirely independent assesment of cost vs size and number of antennas, numbers were derived which are so close to those shown in tables 10 and 11 as to be indistinguishable (D. Woody, 1998). This adds confidence that we have performed the costing estimate properly.

We point out finally that the total construction cost for an ALMA with 64 12-m antennas using the above costing numbers would be $\sim$ \$520M.

Table 11: Possible arrays with total available funding of \$400M

D	N	ND	ND^2	Array Area (m^2)
8-m	81	648	5184	4071
10-m	63	630	6300	4948
12-m	48	576	6912	5428
15-m	31	465	6975	5478

Discussion

Considering all of the issues outlined above, it seems apparent that arrays of antennas with either 10-m or 12-m diameters should be considered at this point. The 15-m designs (even assuming that they could meet the specifications) do not meet the minimum desirable ND requirement and the 8-m designs do not meet the minimum ND^2 requirement.

In order to make a decision between 10-m and 12-m antenna diameter at this point in time, it seems that there are two classes of questions that we need to resolve:

1. What do we hope to achieve with the array in the long term, and how much risk are we willing to assume to achieve it?

2. What is the proper course for us to pursue jointly in the near term?

The first question has to do with where we place our emphasis and how we weigh the risk. What the tables above show us is that if we build an array of 12-m antennas we maximize our collecting area ND^2 at the expense of a $\sim$10% reduction in our imaging sensitivity ND. On the other hand, if we build an array of 10-m antennas we maximize our imaging sensitivity at the expense of a $\sim$10% reduction in our point source sensitivity. Where is the balance here? Is either design acceptable in terms of the science we hope to achieve?

Now consider the risk: the risk that the as-delivered antenna will not meet the specifications is greater for the larger antennas. If we miss the specifications for pointing, surface accuracy, and slew rate, then we are going to increasingly compromise our ability to image well in the submillimeter (where the primary beam is smallest and the surface errors - the sidelobes - are fractionally greater). How important is good imaging in the submillimeter? How do we weigh the risk of compromising submillimeter imaging against the gain of a 10% increase in collecting area by going to 12-m antennas?

It seems extremely difficult to answer these questions because we don't know how to assess the risk that a contractor can deliver the antennas to specifications. This brings us to the second question we've listed above, how to we proceed in the near term? Let us examine our resources. From the MMA we have enough money to contract for a 10-m prototype antenna but not enough to procure a 12-m prototype antenna. In Europe we may have enough for a prototype antenna but only if that becomes the overriding emphasis of the

D&D work in Europe. However, if we combine the MMA and European resources we have several options:

- We have enough money to procure a prototype of a single 12-m antenna
- We have enough money to commission complete design work from two antenna manufacturers and enough to procure a prototype antenna of 10-m diameter (and *perhaps* of 12-m diameter)

The MMA NSF advisory committee has supported the concept of keeping competition in the antenna procurement process as long as possible in order for us to have the most leverage over the antenna manufacturers. For this reason they like the idea of funding two designs. If we did this, one option we would have would be to support design work from a commercial manufacturer on development of the innovative antenna design that David Woody and James Lamb have been working on that extends many of Dietmar Plathner's ideas. Is this a reasonable way to proceed, fund two designs, one of which we define, and then make a judgement as to what we build as a prototype? In this scenario, we could also consider funding one of the designs internally, leaving open the possibility that NRAO may eventually become its own prime contractor on the antenna construction.

If we agree that there is a real risk that an antenna manufacturer can deliver (as opposed to design) an antenna that meets our specifications, and if we agree that this risk is a concern, and hence that a prototype antenna is absolutely necessary, does it follow that the proper prototype to build is then 10-m thinking that if the contractor can't meet those specifications he surely can't meet the 12-m specifications (and it's affordable in the D&D budget we have)?

Again, the point of this manuscript is not to provide one single answer to the question of which antenna diameter to use, but merely to point out and describe the arguments which should provide a framework within which a decision on antenna diameter can be made in an informed manner.

References

T.Anderson, Feasibility Study for a 12 m Submillimeter Antenna, MMA Memo 253, 1999

Bennett, C.L., and 11 others, Morphology of the Interstellar Cooling Lines Detected by COBE, *Ap. J., 434*, 587–598, 1994

Bevis, M., S. Businger, T.A. Herring, C. Rocken, R.A. Anthes, and R.H. Ware, GPS meteorology: Remote Sensing of Atmospheric Water Vapor using the Global Positioning System, *JGR, 97*, 15787–15801, 1992

Guilloteau, S., Large Southern Array: Feasibility, in *Science With Large Millimetre Arrays*, ed. P.A. Shaver, Springer, 1995

Holdaway, M.A., and D.P. Woody, Yet Another Look at Anomalous Refraction, MMA Memo 223, 1998

Holdaway, M.A., Hour Angle Ranges for configuration Optimization, MMA Memo 201, 1998

Holdaway, M.A., Effects of Pointing Errors on Mosaic Images with 8m, 12m, and 15m Dishes, MMA Memo 178, 1997

Jewell, P.R., and J.G. Mangum, System Temperatures, Single Versus Double Sideband Operation, and Optimum Receiver Performance, MMA Memo 170, 1997

Lugten, J., J. Kingsley, J. Cheng, V. Gasho, and M. Fleming, A 10-m Antenna Design for the Millimeter Array, MMA Memo 240, 1999

Plathner, D., A 12m Telescope for the MMA-LSA Project, MMA Memo 259, 1999

Plathner, D., M. Bremer, and J. Delannoy, The 15m (12.8m) Telescopes for the MMA/LSA Project, MMA Memo 254, 1999

Solomon, P.M., and A.R. Rivolo, A Face-On View of the First Galactic Quadrant in Molecular Clouds, *Ap. J., 339*, 919–925, 1989

Solomon, P.M., D. Downes, and S.J.E. Radford, Warm Molecular Gas in the Primeval Galaxy IRAS 10214+4724, *Ap. J., 398*, L29–L32, 1992

Solomon, P.M., D. Downes, S.J.E. Radford, and J.W. Barrett, The Molecular Interstellar Medium in Ultraluminous Infrared Galaxies, *Ap. J., 478*, 144–161, 1997

Weinberg, S., *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, Wiley, 1972

Woody, D.P., and J.W. Lamb, A Design for a Precision 10-m Sub-Millimeter Antenna, MMA Memo 241, 1999

Wright, E.L., and 21 others, Preliminary Spectral Observations of the Galaxy with a 7° Beam by the *Cosmic Background Explorer (COBE)*, *Ap. J., 381*, 200–209, 1991