ALMA Memo XXX
Image Quality as a Function of the Number of Antennas in ALMA

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

February 20, 2005

Abstract

We perform simulations of ALMA data for array configurations with the number of antennas (N_a) ranging from 40, 43, 46, 50, and 56, adding only thermal noise. For very high SNR observations (ie, 10000:1), the dynamic range shows a strong dependence upon N_a, with 56 antennas yielding far superior images to 40 antennas. For high SNR (ie, 3000:1), the dynamic range shows only a modest dependence on N_a, and at 1000:1 SNR and below, all arrays are basically able to achieve the theoretical dynamic range. However, even though all configurations seem to be able to achieve thermal noise limited imaging from the perspective of dynamic range, configurations with fewer antennas provide image reconstructions with on-source errors which are far from the theoretical maximum. For these simulations, the 56 element configuration produces images which have nearly thermal noise limited fidelity up to the very highest SNR, where at last deconvolution errors begin to limit the image quality instead of thermal noise.

A vision for a future path of ALMA imaging work is presented.

1 Imagine This

The ALMA project has a number of science goals and a Design Reference Science Plan (DRSP) to guide it in its technical requirements. However, most of the experiments in the DRSP and the science goals push the sensitivity, and we are not entirely clear in the manner in which these or other observations will push ALMA's imaging capabilities.

At this point, I encourage the reader to dream of an ALMA which rivals the Hubble Space Telescope in imaging quality. A filled aperture (an optical telescope for example) obtains complete and highly redundant Fourier plane coverage and generally can make wonderful images of arbitrarily complex objects. We in the radio community have been struggling with image quality for decades. We have made great progress in overcoming the limitations of phase and amplitude errors, and we have made inspiring images, judging them to be great in part because we are able to achieve high dynamic ranges. However, the radio community is only slowly coming to terms with a fundamental limitation in radio imaging: the effects of incomplete (u,v) coverage. As we will see clearly in the work presented in this memo, high dynamic range does not imply high image quality as measured by the level of error on-source, or on-source SNR. Having more antennas results in lower on-source errors. ALMA, originally designed to have 64 antennas, would have constituted a revolution in radio synthesis imaging, and would have brought about a revolution in millimeter wavelength astronomy. Cutting back the number of antennas in ALMA threatens that revolution.

I'd like you to imagine making an image with ALMA that is as wonderful and complex and details as an HST image. To be concrete, consider an observation at 300 GHz or 1 mm and 0.1 arcsec resolution. We'll start with an image 2048 pixels across, with 0.03 arcsec pixels. This image would be 60 arcsec across, much larger than the primary beam, so we will need to mosaic to make this image. With pointings separated by $\lambda/2D$, we will need to separate discrete pointings on the sky by 8.6 arcsec, so we could cover our 60 arcsec region with an 8x8 grid of discrete pointings. Now, 0.1 arcsec at 1 mm is equivalent to a 2000 m baseline (actually, with the Gaussian (u,v) distribution which ALMA's array configurations produce, there will be baselines as long as 3000 m in the array which produces 0.1 arcsec resolution, but the coverage out there is very sparse, so "complete coverage" is already out of the question). In order to avoid time smearing on a 2000 m baseline, we will need the integration time to be at most 40s. Or, if we are doing an 8x8 mosaic, we need to go through the entire raster of 64 discrete pointings in 40s if we wish to have (u,v) coverage in each pointing which is essentially as good as a single pointing observation[1]. Hitting 64 discrete pointings on the sky in 40 seconds won't work, but performing on-the-fly mosaicing, sampling better than Nyquist with about 0.2 s integration time will just about work.

In the most compact ALMA arrays, we will have complete (u,v) coverage in a single snapshot, and we'll be able to make wonderful mosaic images of complex sources at low resolution. However, as we go to longer baseline configurations, our cell size will stay the same, but the number of pixels in the (u,v) image will increase, so eventually we won't be able to get good mosaic images from a snapshot on each field. To counter this, we will need to observe each field with longer tracks. The length of the tracks will be limited by the hour angle limit which is appropriate for a given frequency. As the sky becomes more opaque at high frequencies, following the source down to lower elevations becomes a game of diminishing returns as the opacity, noise, phase errors, and amplitude fluctuations all increase. So, as we take away antennas from ALMA, we basically restrict the configuration size, or resolution, with which we will be able to make good mosaic images of complex source structures. With 56 antennas and observations ranging from -2h to +2h, we get fairly complete (u,v) coverage at 0.1 arcsec resolution. Reducing the number of antennas will mean we could still perform good mosaicing, but at lower resolution.

We need to reconstitute the ALMA imaging group, and one of the first things we need to do is to push some large, high resolution simulated mosaics (ie, 8 x 8 pointings at 0.1 arcsec) through AIPS++. For the amount of time we had to perform the current work in advance of the February 2005 ASAC Meeting, this could not be managed, so instead, we made a simulation of a single field at 0.1 arcsec resolution. The source size was taken to be the largest source which we could reconstruct all the flux from with the configuration we used, so this should be considered to be a step in the direction of a Hubble-type image, though these images are only

[1] Having nearly complete (u,v) coverage is essential for good mosaic image quality. Single field imaging usually looks at a source in a small part of the primary beam, and can thus employ a "clean box" or support constraint, effectively reducing the number of unknowns to solve for or increasing the size of the (u,v) cells.

256 pixels.

2 Simulation Details

We have performed detailed numerical simulations of ALMA's single field imaging in AIPS++.

2.1 The Model Images

For model brightness distributions, we have used two images obtained from optical telescopes: the M51 galaxy as imaged in HII, and a single HII region in the M31 galaxy. The M51 image is dominated by many small clumps in the spiral arms, and the M31 image is dominated by extended structure, so they complement each other well. The two models, convolved with an 0.1 arcsec circular beam, are shown in Figures 1 and 2. The M31 model has a maximum extent of about 5.5 arcsec, and the M51 model image has a maximum extent of about 6.0 arcsec. Note that at 300 GHz, $\lambda/2D$ (ie, the mosaic spacing) will be 8.6 arcsec. Model images larger than these could be imaged with single configuration interferometry could succeed with larger model images, but only if we added some compact configuration data. These are just about the largest model images we could make with single configuration imaging at 0.1 arcsec resolution. The image plane noise level for a 4 hour observation with a 56 element ALMA at 300 GHz is 8.17 microJy. To achieve a theoretical SNR of 10000:1, we scaled the model images such that when concolved with the 0.1 arcsec restoring beam, both the M31 and M51 images had a peak brightness of 81.7 mJy/beam. The integrated flux for the M31 image at 10000:1 SNR was 13.5 Jy, while the M51 image at 10000:1 SNR had a flux of 3.82 Jy. For simulations at different SNR, we kept the added thermal noise constant and scaled the model images to fainter levels.

2.2 The Configurations

We optimized configurations for 40, 43, 46, 50, and 56 elements using the Boone algorithm.

In single configuration imaging of large objects, we run the risk of the reconstructed images being dominated by artifacts due to poor short spacing coverage. To eliminate this possibility, we first optimized the inner 250 m worth of Fourier plane coverage by using the Boone algorithm to place 11 antennas, fixing the positions of four antennas which had made the shortest two baselines in Conway's analogous configuration which gave just better than 0.1 arcsec resolution at 300 GHz. Starting with these 11 fixed positions, we then used the Boone algorithm to optimize the positions of the remaining $N-11$ antennas over a -2h to +2h track such that the resolution was just slightly better than 0.1 arcsec. (We ended up with 300 GHz resolutions of about 0.098 and 0.099 arcsec, but all configurations were tapered to the exact 0.1 arcsec resolution in the imaging process.)

2.3 Errors: (u,v) Holes and Thermal Noise

Only two error processes were present in the data: errors due to the incomplete coverage of the Fourier plane, and thermal noise which was added to each visibility. We already have a firm understanding of how the radiometer equation works, and we know that observations with fewer antennas will yield higher image noise. What we are interested in here is how image quality will degrade as we lose antennas from the array. Hence, so as not to confuse these two

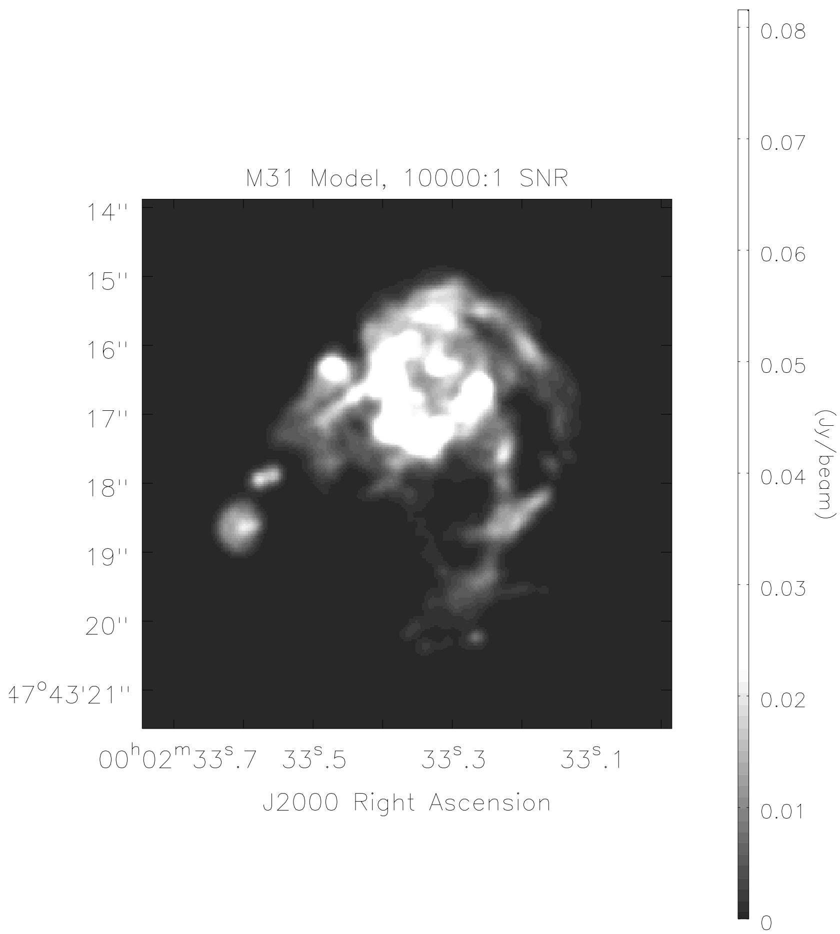

Figure 1: M31 model used for SNR=10000.

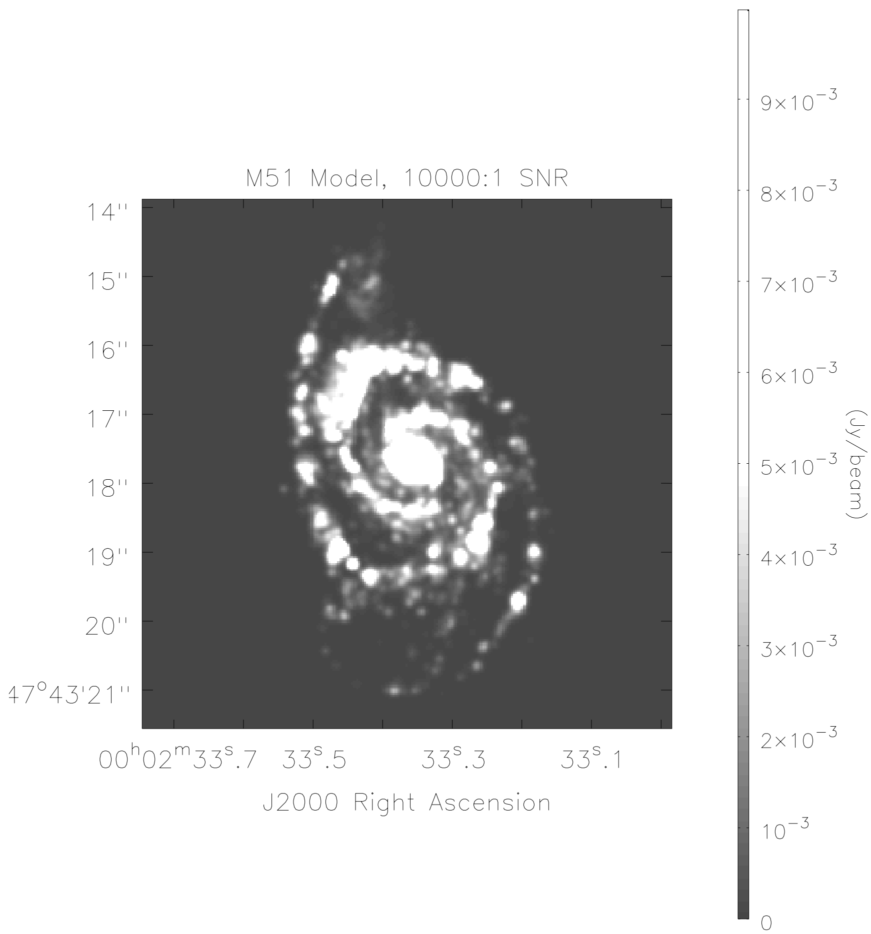

Figure 2: M51 model used for SNR=10000.

N antennas	integration time [hours]
56	4.00
50	5.03
46	5.95
43	6.82
40	7.90

Table 1: The amount of time we needed to simulate the different configurations for to achieve the same noise level. With 40 antennas, we basically need to observe twice as long to achieve the same image plane noise level.

issues, we sought to simulate observations with the same final SNR, and hence the same final image plane noise level. The intention was to limit the hour angle range to -2h to +2h for the 56 element configuration. The configurations with smaller numbers of antennas would then need to observe more than a single transit in order to maintain the same SNR. However, due to complications with the measurementset concatenation function in my rather old version of AIPS++, that strategy was not possible, so we faked this strategy by adding lower amounts of noise to the simulations with fewer antennas such that the final image plane noise level always came out very close to 8.17 microJy. The supposed amounts of integration time are shown in Table 1. As mentioned above, we scaled the model images by different amounts so we could look at the different configurations' ability to image at different SNR. We used SNR of 300:1, 1000:1, 3000:1, and 10000:1. SNR is defined as the peak of the model image convolved with the 0.1 arcsec beam divided by the image plane thermal RMS noise level.

2.4 The Observational Details

We observed the source at -48.0 deg declination, but we accidentally specified the position in B1950 coordinates and imaged in J2000 coordinates, so the images aren't exactly at -48.0 declination. As mentioned above, long tracks from -2h to +2h were used, with 30 s integration times. The observing frequency of 300 GHz and bandwidth of 8 GHz and two polarizations were simulated.

2.5 Image Reconstruction

The exact taper which would result in 0.1 arcsec resolution was found using the "findtaper()" tool. This usually meant tapering the (u,v) data by an elliptical Gaussian of order 0.04 arcsec. This mainly impacted just the outlying visibilities, and resulted in sensitivity loss of about 1%. Past simulations have found a somewhat chaotic behavior of image fidelity with the exact resolution, and we wanted to eliminate that difference when comparing the different configurations' images. After the correct taper for each configuration was calculated, the data were Fourier transformed with natural weight to yield 512 x 512 dirty images and point spread functions.

The dirty images were deconvolved using both MEM and Clark CLEAN algorithms in AIPS++. A mask image (ie, a generalized CLEAN box) which extended about 5 resolution elements (15 pixels) beyond the outline of each source was used to help constrain the deconvolution. The reasoning here is that any project which has been awarded four hours (or more) of observations at 0.1 arcsec resolution has been observed before at lower resolution, and we can use the lower resolution image to provide a mask to constrain where the high resolution flux will go.

MEM accepts an image plane noise level and a maximum number of iterations, and will stop iterating when the fit is below the image plane noise level. We set the maximum number of MEM iterations to 200, and set the noise level to the actual image plane thermal noise level. We found that the thermal noise level was never reached, but the algorithm usually came within about 10% of the thermal noise, so the full 200 iterations were always performed. Essentially no improvement in the image fit was made in the last 40-80 iterations, so we feel that a larger number of iterations would not have improved the MEM images.

The Clean algorithm accepts a stopping residual flux level and a maximum number of iterations. We found that the image quality improved as the stopping residual flux level was decreased, but the improvement essentially stopped at about 0.2 sigma (ie, 1/5 of the image plane noise level). For the higher SNR simulations, or sometimes for the smaller N configurations at somewhat lower SNR, we found we were not able to reach the 0.2 sigma residual level. In those cases, the maximum iterations number was set to 300,000. When unable to achieve the low residual threshold, the algorithm would ofte go into a slow down mode, so doing many more than 300,000 iterations would be futile and costly time-wise.

3 The Results

Recall the dimensions of the phase space we are concerned with: the number of antennas in the single configuration are 40, 43, 46, 50, and 56. We have the M31 model dominated by extended structure and the M51 model dominated by compact structure. We have theoretical SNR limited by thermal noise, with values of 300:1, 1000:1, 3000:1, and 10000:1. And we have used both MEM and Clark CLEAN to deconvolve the images.

3.1 Evaluating the Images: Dynamic Range

For low to moderate SNR, we expect all of the configurations to be able to achieve the theoretical dynamic range equal to the SNR. In Figure 3, we see this to be more or less true, every configuration is close to the 300:1 dynamic range limit. In Figure 4 where the theoretical dynamic range limit is 1000:1, it is still more or less true that every configuration comes close to the theoretical limit, but there is a slight trend where more antennas is better. And last, as we go up to SNR=10000 in Figure 5, we at last see a marked advantage with more antennas, with the 56 antenna configuration outperforming the 40 antenna configuration by a factor of 3-4 in dynamic range. In this case, even 56 antennas wasn't quite enough, it only reaches a dynamic range of about 8000:1.

A quick digression on such high dynamic ranges. ALMA will be susceptible to many kinds of errors in addition to thermal noise: phase errors, pointing errors, primary beam errors, and amplitude fluctuations. These nasties will usually limit the dynamic range to far below 10000:1. However, when SNR permits, radio astronomers will draw upon a growing arsenal of self-calibration algorithms and will be able to get past the limitations of many of these effects. However, we can't really get beyond the basic limitations of Fourier plane coverage. So, while

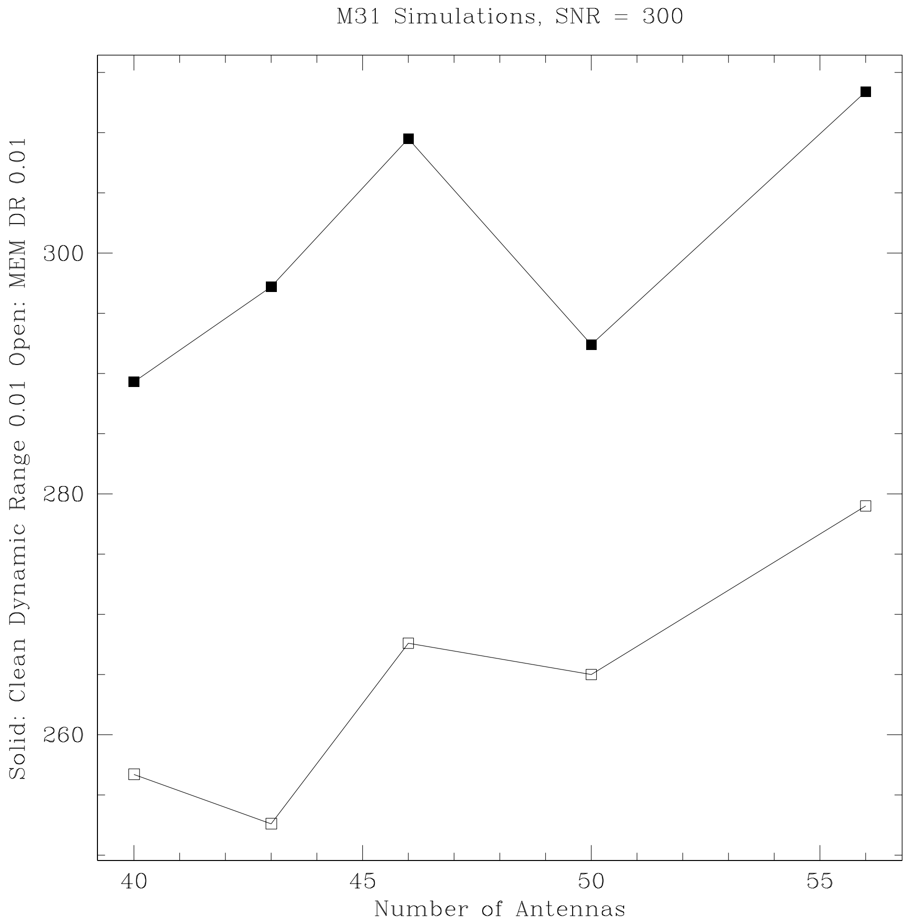

Figure 3: Dynamic range of the M31 model as a function of number of antennas for SNR=300. The open squares are for CLEAN deconvolution, filled squares are for MEM deconvolution. All configurations do similarly well at achieving the theoretical DR of 300:1

we don't expect to get up to 10000:1 for most observations, it is good to have that capability for when we need it.

3.2 Evaluating the Images: On-Source Fidelity

If the dynamic range were the entire story, I would probably recommend that we pack up and go home. Can we justify the difference between 40 and 56 antennas based on their performance of imaging on the small fraction of sources which have potential dynamic ranges in excess of 2000:1? Perhaps, but it would be a hard sell. However, there is another side of this story.

Typically, radio astronomers will pull out self-cal and various other fix-it-up algorithms and apply them iteratively with deconvolution until the off-source RMS approaches the thermal noise, and then call it quits as this is as well as we can do. As all the configurations did similarly well in dynamic range for the 1000:1 SNR series of simulations, we might take a

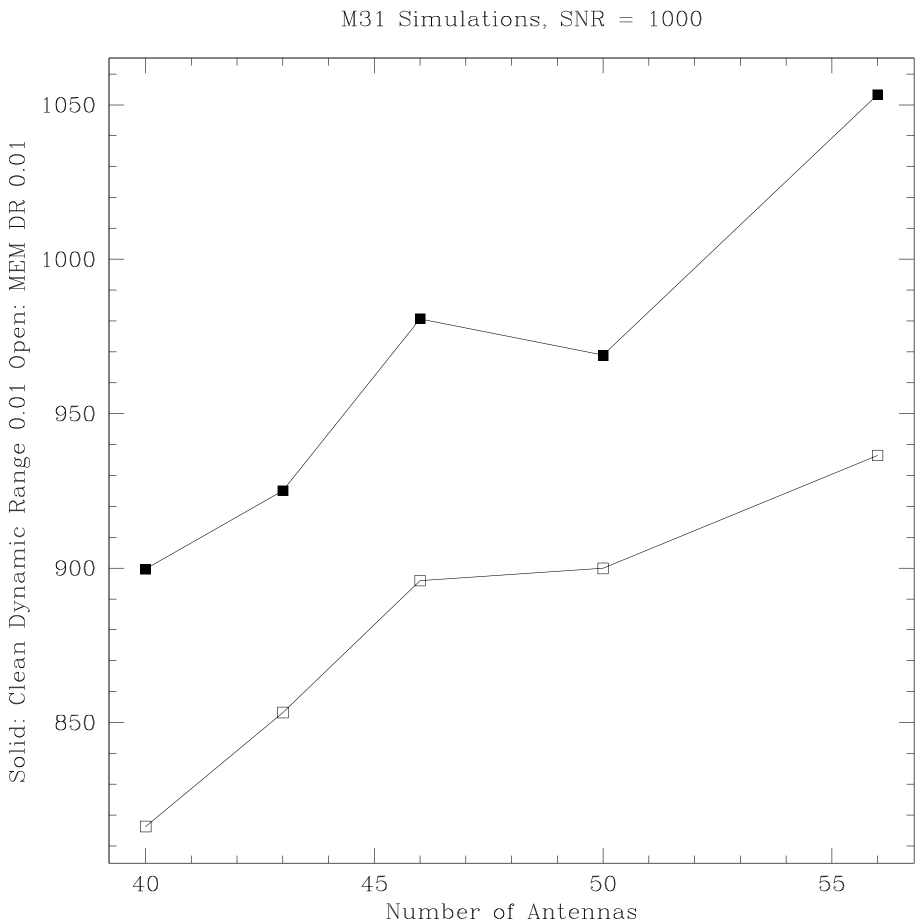

Figure 4: Dynamic range of the M31 model as a function of number of antennas for SNR=1000. The open squares are for CLEAN deconvolution, filled squares are for MEM deconvolution. There is a mild trend, with configurations of larger numbers of antennas doing slightly better at achieving the theoretical DR of 1000:1, but all configurations are similar.

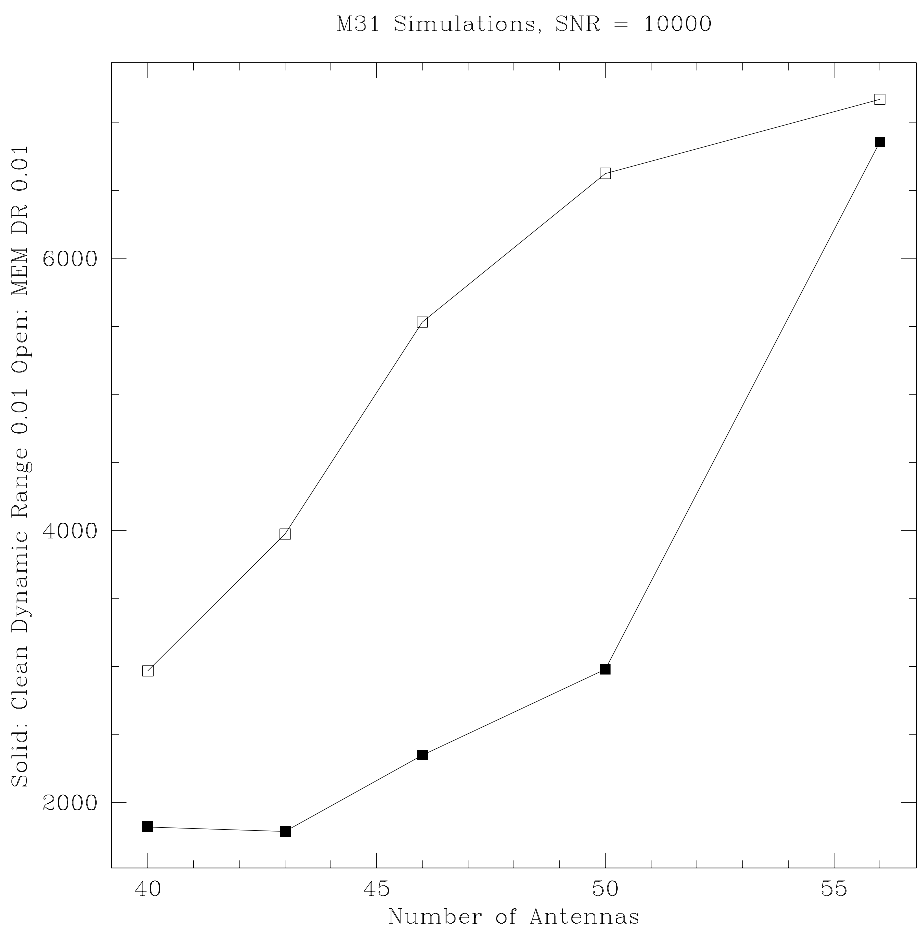

Figure 5: Dynamic range of the M31 model as a function of number of antennas for SNR=10000. The open squares are for CLEAN deconvolution, filled squares are for MEM deconvolution. At the highest potential dynamic ranges, the configurations with more antennas have a distinct advantage, and the configurations with fewer antennas don't do nearly so well.

similar approach and say that we are done. But we aren't done, because off-source errors and on-source fidelity are two very different measures.

The fidelity image is an SNR image and can only be calculated if you know what the source REALLY looks like. It is given by the model image convolved with the clean beam, divided by the error image. The error image is given by the absolute value of the reconstructed image minus the model image convolved with the clean beam. The image fidelity will be a single number derived from and representative of the fidelity image. In this memo, we take all pixels where the convolved model image is brighter than 1% of the peak, and calculate the median of the fidelity image for those pixels.

Further, we can calculate what the best the image fidelity could be given the thermal noise. To do that, we take the perfect model image, convolve with the clean beam, and add the appropriate thermal noise, and feed that image to the function that calculates the fidelity. The theoretical maximum fidelity will be a linear function of the SNR. Any deviations from this theoretical maximum will represent deconvolution errors on source. In this set of simulations, those deconvolution errors can be traced directly back to deficiencies in the configurations.

As an example, Figure 6 shows the image fidelity plotted against the theoretical SNR. The theoretical maximum fidelity is plotted as a dashed line, and below it lie lines for the 56, 50, 46, and 40 element array configurations. **This is the most important point of this paper. Even though there was essentially no difference in dynamic range among the different configurations for the SNR=300 and SNR=1000 cases, we can see significant differences in the fidelity measuring on-source image quality.** What that means is that even though an image looks good because the off-source noise had been beaten down, the image could be pretty poor due to deconvolution errors.

3.3 Evaluating the Images: (u,v) Fractional Errors

In the Fourier plane, we can see exactly where the fidelity limits are coming from. The fractional error in the reconstruction is calculated by:

$$FT(TrueModelImage - DeconvolutionModel)/FT(TrueModelImage),$$

where the DeconvolutionModel is either the unconvolved MEM image or the Clean Component image. The result of this calculation is an image of fractional reconstruction errors in the Fourier plane. We can investigate these errors pixel by pixel, or we can investigate these errors in a more general sense, for example, by binning the plane by radial (u,v) distance. Figure 10 shows the rms fractional error in the reconstruction as a function of UV distance for the M51 300:1 SNR image imaged with CLEAN. The solid line is the image made with 56 antennas, and the other dashed lines are for 50, 46, and 40 (sorry you don't know which are WHICH). The solid line generally has the lowest error. There are two different things going on: first, as you go to longer baselines, you have larger errors, because you have less flux on long baselines, and you have fewer long baseline as well, so the noise (ie, one component of the errors) goes up. Additionally, there are several nasty spikes all through the error plot. These are due to an interaction between gaps in the (u,v) plane, the source structure, and the thermal noise (I will prove this when I update this section). In Figure 11, we show the same plot for the 10000:1 SNR case on the same scale. The SNR is about 30 times higher, so the big noise curve on the long baselines is quite reduced. However, there are still nasty spikes at various (u,v) distances, and the 56 element configuration always does better.

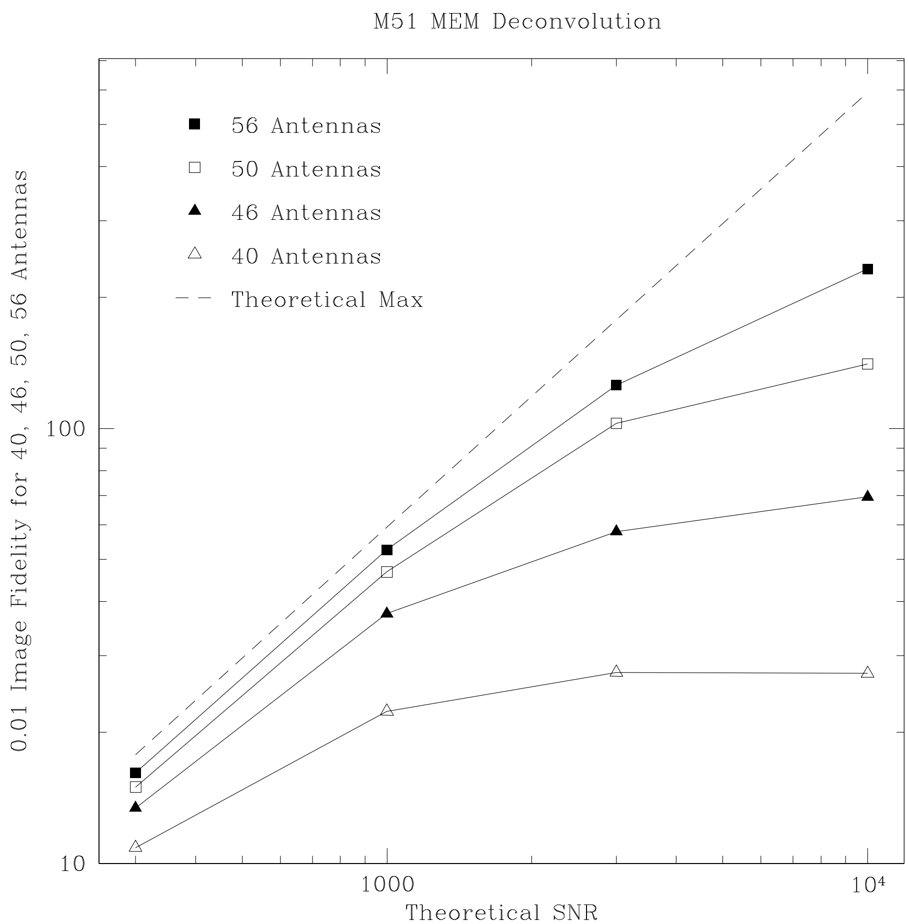

Figure 6: For the M51 model imaged with MEM, the 0.01 image fidelity as a function of SNR for 56, 50, 46, and 40 antenna configurations.

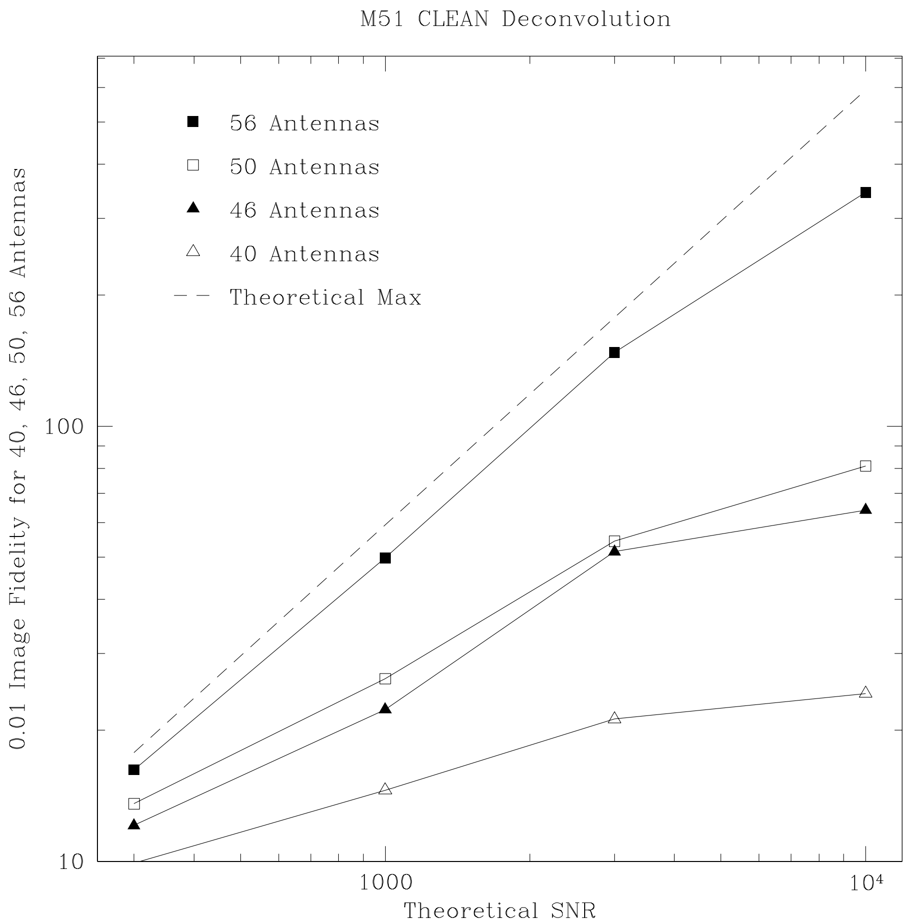

Figure 7: For the M51 model imaged with CLEAN, the 0.01 image fidelity as a function of SNR for 56, 50, 46, and 40 antenna configurations.

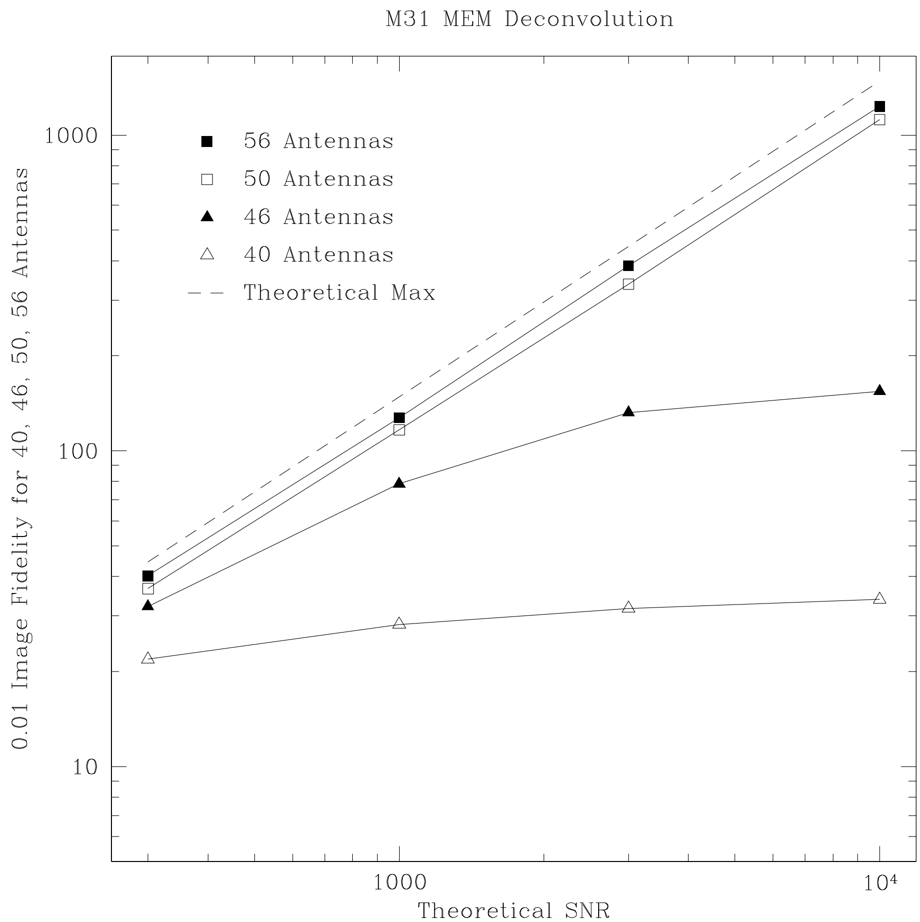

Figure 8: For the M31 model imaged with MEM, the 0.01 image fidelity as a function of SNR for 56, 50, 46, and 40 antenna configurations.

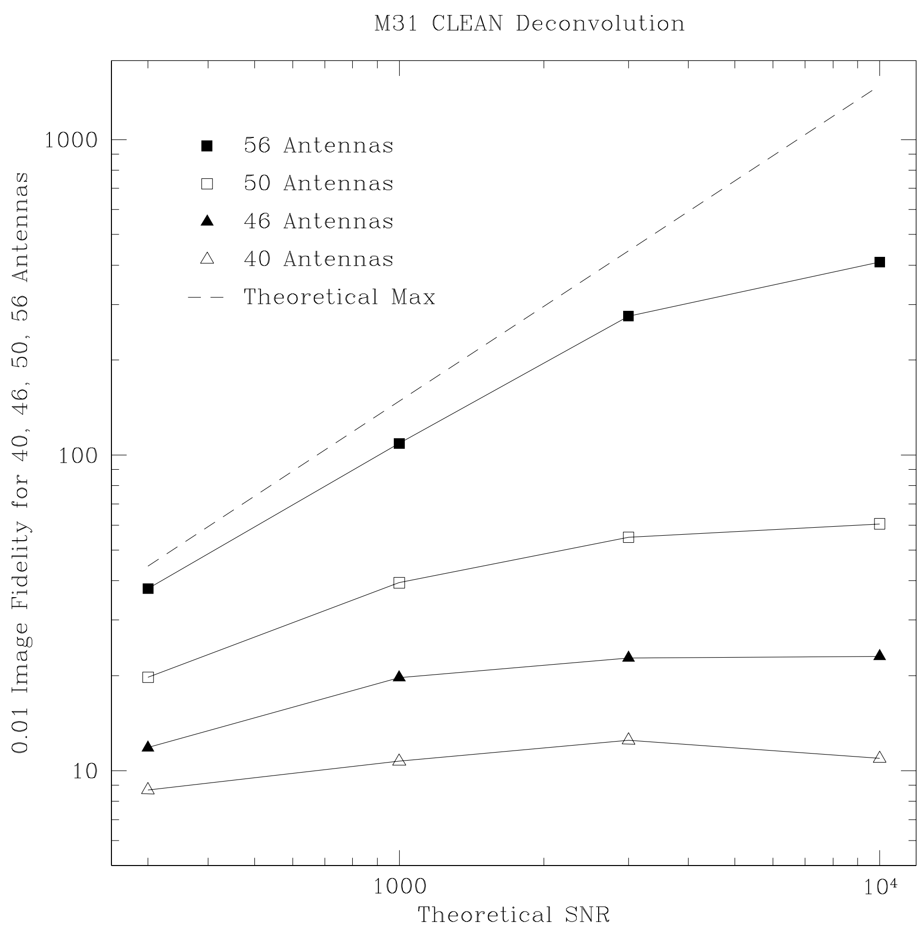

Figure 9: For the M31 model imaged with CLEAN, the 0.01 image fidelity as a function of SNR for 56, 50, 46, and 40 antenna configurations.

So, what is going on that makes some Fourier pixels in the reconstruction go haywire? It is a combination of small value of model visibility amplitude on a particular (u,v) pixel, poor (u,v) coverage on that pixel, and large values of model visibility gradient, as can bee seen in Figure 12. This graph was made from the residual errors and (u,v) coverage for the M31 MEM 3000 SNR 40 antenna simulation, and used Fourier data between 500 and 800 m. This graph plots, on a pixel by pixel basis, the fractional error in the Fourier plane as a function of an arbitrary quantity called "goodness", defined as UV-weight (UVWT) in that pixel (ie, the number of Fourier samples which were gridded onto that pixel) raised to some power, times the source's Fourier amplitude (ModVis) at that pixel raised to some other power, times the GRADIENT of the source's Fourier amplitude at that pixel (GradVis). We made a least squares fit to the exponents and found that the scatter in the plot is minimized if Goodness = $\mathrm{UVWT}^{0.5}$ * $\mathrm{ModVis}^{2.5}$ * $\mathrm{GradVis}^{-0.4}$. Different simulations and different parts of the the (u,v) plane show different best fit exponents (for example, the inner part of the (u,v) plane usually shows a smaller exponent of the ModeVis. The exponent of UVWT is always 0.5. In some cases, we do not find compelling evidence for the Gradient term. Anyway, this way of looking at the errors indicates that there is only a mild contribution from the UV weight, and that errors are caused more often by low points or nulls in the Fourier transform of the image – unless, of course, we have essentially no samples and no UVWT. Due to the gridding process and the way we obtained the UVWT image, those pixels with no samples actually had a very small weight, typically between 0.05 and 0.5. Pixels with essentially no Fourier coverage (UVWT ¡ 0.5) are marked in the plot by filled symbols, and you will note that they tend to have large fractional errors, but are not the worst pixels.

So, this is a very promising method for understanding how the source structure and the error processes interact in synthesis imaging, but it isn't the final word yet. It does suggest that the errors and limitations we see in the simulations are mainly due to the interaction between the Fourier plane coverage, the source structure, and the thermal noise. Hopefully, this sort of analysis can be fruitfull when applied to the other imaging errors which are listed in the FUTURE PLANS section.

4 Conclusions

Very SHORT STILL, sorry! Better Fourier coverage means better images, even if you can't see what is wrong because of dynamic range.

5 A Plan for Future Work

Some of the simulations performed in this document result in fidelities of well over 100. For comparison's sake, a flux scale error of 1% will limit us to a fidelity of 100. It makes sense to want images with higher fidelity if results of calculations on this image alone are improved by the high fidelity. On the other hand, if comparisons with other images at other frequencies dominate the calculations, then having a fidelity much higher than 100 will not be very useful.

Past numerical simulations have indicated that pointing errors will often limit image fidelity to around 20-30. If this is the case, it would not be beneficial to have (u,v) coverage capable of 100 fidelity if we will just be limited by pointing errors. However, there is reason to believe that we may be able to overcome this limitation. We must address that before we know how

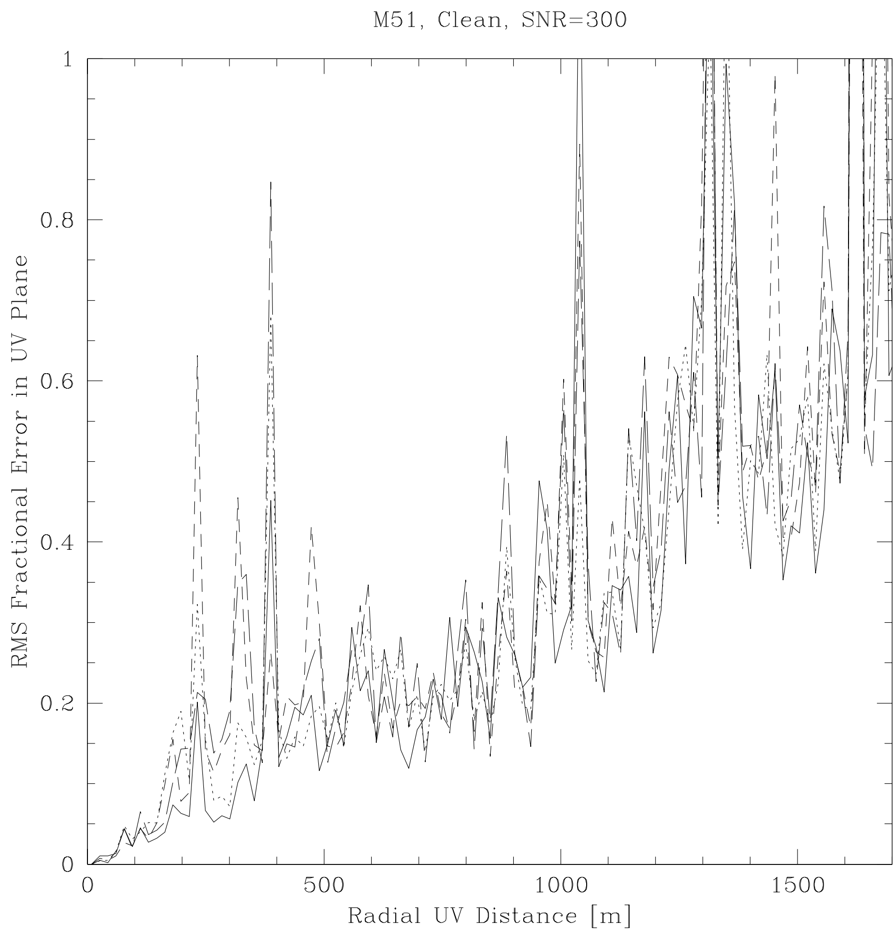

Figure 10: Radial plot of rms (u,v) fractional errors for M51, CLEAN, SNR=300.

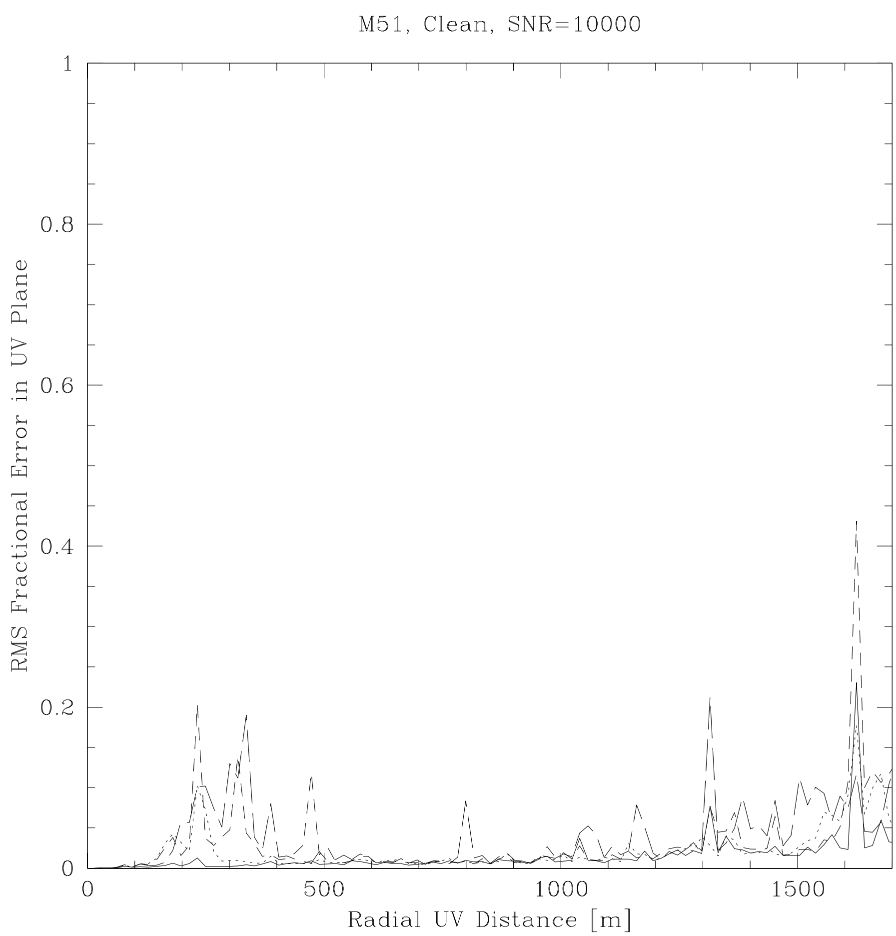

Figure 11: Radial plot of rms (u,v) fractional errors for M51, CLEAN, SNR=1000.

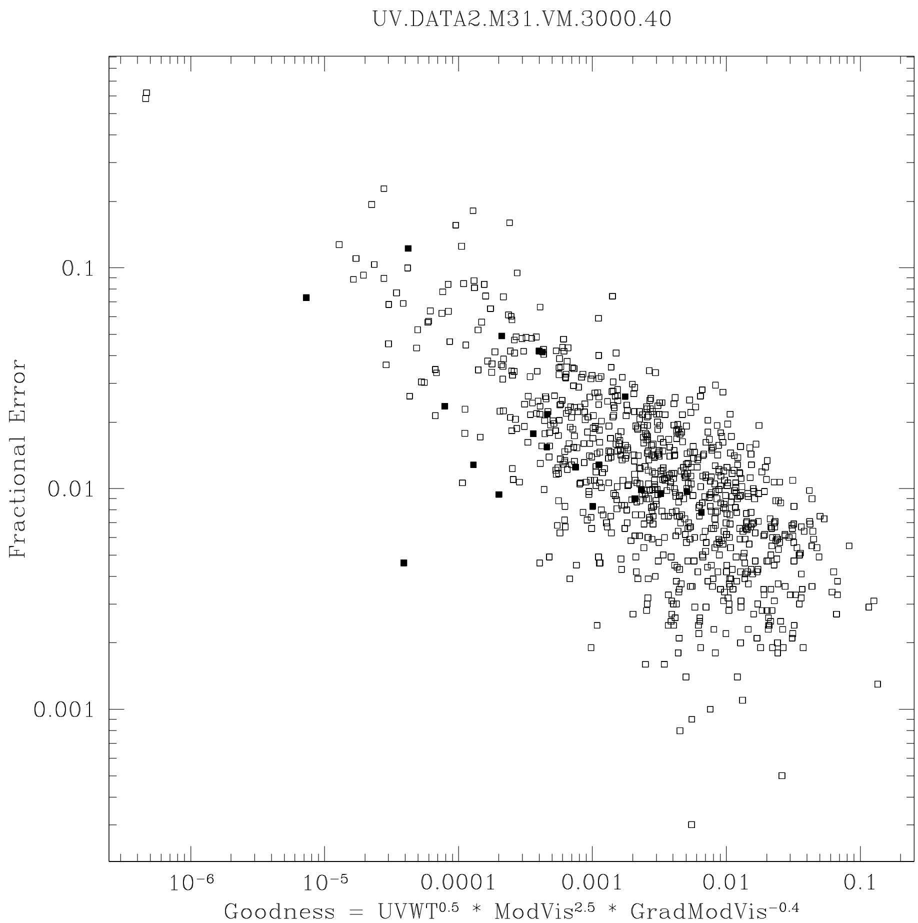

Figure 12: Where do these errors in the (u,v) plane come from? It is an interaction between the (u,v) coverage, unhappy accidents in the visibility structure of the source, and thermal noise. This plot, taken from the M31 MEM 3000 SNR simulation with 40 antennas on baselines between 500 and 800 m, shows a scatter plot of the fraction error in the (u,v) plane as a function of the "best fit" goodness parameter = UVWT$^{0.5}$ * ModVis$^{2.5}$ * GradVis$^{-0.4}$. The UVWT is the sum of natural weights in the Fourier plane, ModVis is the true model visibility amplitude, and GradVis is the maximum gradient in the visibility image. Filled symbols represent pixels which were not sampled by ALMA (due to the way the gridding was performed, these pixels appear to have weights typically of 0.05 - 0.5, rather than 0.0). Pixels most likely to have large fractional errors are pixels with very little flux, very low UV weights, and very large gradients in the model visibilities.

to interpret the very high fidelities we have found in this memo.

Many of the simulations which helped justify or set the MMA and ALMA specifications were performed in SDE (Tim Cornwell's *Software Development Environment*). SDE is a dead package, and we have basically lost access to that functionality. However, with the new pointing error work being performed by Sanjay Bhatnagar, Kumar Golap, and Tim Cornwell, we will very soon have the essential simulation functionality in AIPS++: we will be able to simulate the effects of pointing errors and voltage pattern errors. Furthermore, this new work will result in algorithms permitting us to solve for and correct pointing errors and voltage pattern errors, a quantum step beyond SDE's capabilities.

Past simulations indicated that pointing errors are a major obstacle to high fidelity imaging , limiting the image fidelity to well below the limits due to (u,v) coverage. However, with the possibility of pointing self-calibration, we should be able to achieve higher quality images. Furthermore, excellent (u,v) coverage will ultimately help us provide better model images for the self-calibration process, and will ultimately help us get past the various sorts of errors that afflict our data.

While many of the problems we are worried about have been simulated (ie, thermal noise, phase errors, pointing errors, voltage pattern errors), we have not performed any comprehensive simulations of ALMA data which include all of these factors. The time has come to do just that. I propose that we start with the Hubble-like image (an 8 x 8 pointing mosaic, or 2048 x 2048 pixels at 0.1 arcsec resolution), apply all the errors we can imagine according to AIPS++'s data model, and then proceed with the iterative process of imaging and self-calibration. The effects which I foresee including in these simulations are:

- **(u,v) holes -** by the time we perform these industrial strength simulations, we may be closer to knowing how many antennas we will have in the array, but we will still be in the process of optimizing the configurations.

- **thermal noise -** as we discovered in the simulations presented in this memo, it is important to cover a wide range of thermal noise. More importantly, the level of the thermal noise will in part determine how well the various self-calibration algorithms will work.

- **residual phase errors -** the two competing phase calibration strategies, fast switching and water vapour radiometry (WVR), will result in residual phase errors with very different properties. We will have to simulate through a model atmosphere with both of these observing strategies. Observations of a nearby calibrator will be included.

- **amplitude fluctuations -** the model atmosphere will also produce fluctuations in opacity, which will translate into amplitude fluctuations. In the high SNR cases, some of the affects of the fluctuations can be removed via amplitude self-cal. These simulations will indicate when the amplitude fluctuations are a problem and when we can correct for them and how much we can correct.

- **pointing errors -** the model atmosphere will also result in refractive pointing jitter (aka anomalous refraction), pointing errors of as large as 1 arcsec with characteristic time scales of 1 s, the crossing time of the atmosphere over the dish. These very fast pointing errors will be very difficult to correct via pointing self-cal, as the object must have high enough SNR to permit ALMA to detect them on time scales under 1 s. Another possibility is to use WVR to infer the pointing errors from the slope of the phase wedge above the antenna, but WVR is taking data on 1 s time scales, which may not be fast enough. In addition to refractive pointing jitter, we will face other sorts of pointing errors: slow drifts due to thermal effects, wind buffetting, and "frozen in" pointing errors due to the limitations of frequent pointing solutions on a nearby calibrator. This is a huge project all to itself.

- **voltage pattern errors -** our model of the primary beam will be incomplete and in error, in large part because of surface errors from the dish. AIPS++ is on the verge of being able to simulate voltage pattern errors, and a few months after that, AIPS++ will probably be able to perform some sort of voltage pattern self-calibration. Earlier work indicated that at high frequencies (ie, bands 8, 9, and 10), the affects of surface errors would dominate the effects of pointing errors on mosaic images.

- **illumination taper offset -** earlier work indicated that the illumination taper's offset from being centrally aligned with the dish would be a more severe problem than pointing errors at frequencies below 300 GHz. However, a very simple fix, shifting the (u,v) coordinates to reflect the illumination taper's center, basically fixes this problem. However, if we are going to be in the business of simulating all effects on imaging, we should include this effect as well to ensure that it doesn't interact badly with other problems.

- **simulate a range of frequencies and a range of observing conditions -** as in other simulations we've done in the past, we must factor in the range of observing frequencies, and match those observing frequencies to representative conditions on the site as determined by a model of dynamic scheduling. The model of dynamic scheduling needs to consider opacity, phase stability, and wind. I would recommend simulating observations at 115, 230, 345, 490, 650, and 850 GHz. At each frequency, we should also simulate something close to the best conditions we could expect for that frequency (given dynamic scheduling) and something close to the worst conditions we could expect for that frequency.

This is a huge project in scope, and it also represents a large number of simulated observations: 6 observing frequencies, each at 2 sets of atmospheric conditions, with 2 phase calibration strategies, each at 5 different SNR settings, and perhaps 3 different model images[2]. So in all, that would be about 360 different simulated data sets. (For comparison, the simulations presented in this memo included 4 different SNR settings, 5 different configurations, and 2 different model images, for a total of 40 different simulated data sets, each imaged with CLEAN and MEM.) There will be further multiplication when we consider different imaging algorithms and self-cal strategies. And finally, most of these problems are made worse when observing sources as larger or larger than the primary beam. We should also investigate how these problems decline in severity as we image sources smaller than the primary beam.

There should be many important results from these simulations. We will see how many of these errors interact with each other. We will see what errors limit the imaging at the different frequencies and atmospheric conditions and source types. We will see under what conditions

[2]For models, I suggest something dominated by extended structure like the M31 model, something dominated by several compact features such as the M51 model, and something dominated by a single bright compact source surrounded by lots of low level extended structure.

the various self-calibration algorithms work and fail. And we will likely provide input to the algorithm developers, pushing them in the direction of more useful versions of their algorithms.