Calculation of Efficiencies, etc, from Beam-Scanning Data

With the goal of eventually reaching an agreed procedure that can be used throughout ALMA, I have written out the expressions that I used to calculate the efficiencies and other parameters, such as the position of the centre of curvature, in a recent look at the data produced by the NSI scanner at the NA FEIC.

I will only treat the case where we are working with complex far-field data, i.e. I am assuming that the conversion from the measurement plane to the far-field has already been done and any corrections needed, e.g. for the probe pattern, have been applied.

I note that we are really concerned with the field at the subreflector, rather than the far-field. The receiver to subreflector distance is about 12 times the Rayleigh distance for Band 3 so using the far-field is a good approximation. If we ever build Band 1 then the fields at the subreflector distance ought to be calculated explicitly.

A second approximation is that I have assumed that the propagation is paraxial – i.e. that the directions of propagation, u and v are sufficiently small that we can write $\sin(u) = u$, etc. I have not estimated the errors that might arise from making this approximation, but I would expect them to be small given that the edge of the subreflector is at only 3.58 degrees off-axis. The data I received was on a uniform grid labelled with what are called the X and Y in degrees. Since the natural units for the output from an Fourier Transform are $\sin(u)$ and $\sin(v)$, I assume that the data has been interpolated, but I haven't looked into the details of that. My feeling is that it would really be better to work with the values coming straight out of the Fourier transform rather than have this interpolation step in the processing. One could then do things a bit more carefully and avoid this approximation.

I also assume that the scan is made in a plane perpendicular to the telescope's z-axis and that the axes of the scan have been set up so that they coincide with the x- and y-axes of the telescope. If this has been done correctly then we can assess whether the beam is propagating in the right direction to illuminate the subreflector correctly. Any alignment errors in setting up the scanner will of course be reflected in the results.

The final assumption is that the centre point of the scan has been chosen such that if we were measuring a point source at the nominal position of the beam waist, then the phase measured in the far field would be uniform, i.e. the reference phase centre is at the point (0,0,0). Since the scan is not in fact made in the focal plane, but nominally at z = 200 mm, some corrections are needed. In the data that I have at present it appears that this is not being done right in all cases and this is something that we need to sort out, but obviously I will give the expressions that I think apply for the case where the referencing is correct.

I have written the expression as if we had array of complex values $\boldsymbol{E}(u,v)$. In fact the data from the NSI scans comes in the form of "amplitude" in dB's, i.e. $A = 20\log10(|E|)$, and phase φ in degrees, so we first form $|E| = 10^{A/20}$.

The processing steps are then:

1) Find the fraction of the power on the subreflector:

$$\text{Eta\_spill-over} = \Sigma_{sec} |E|^2 \ / \ \Sigma_{all} |E|^2$$

where Σ_{sec}, is the sum over data points for which $r < r_s$ where $r^2 = u_1^2 + v_1^2$ with $u_1 = u - u_0$ and $v_1 = v - v_0$, r_s is the radius of the subreflector, 3.58 degrees, and the Σ_{all} is the sum over all data points.

There are a couple of choices to be made here. First should (u_0, v_0) be the peak of the beam, the point that gives the maximum efficiency, or the nominal direction corresponding to the centre of the subreflector? Since we cannot in fact move the subreflector or realign the receiver, I think that we should use the nominal direction. This means that most of the loss that arises due to any miss-pointing of the receiver will be included in the spill-over loss. I realize that at present there is a separate item in the requirements that allows 2% loss for

misalignment, but from the point of view of sensitivity it doesn't matter what the origin of the loss is. In fact there is a separate reason why the alignment (the "pointing" of the receiver) needs to be good: if the beam is not well centred, there will be an asymmetry in the illumination which in turn causes a linear phase error across the beam on the sky. Recall that the (complex) beam on the sky is, to a good approximation the Fourier transform of the field on the subreflector. According to the derivative theorem of FT's the slopes of the imaginary part of the beam are proportional to the first moments of the field on the subreflector. The relevant parameters therefore are the moments of the amplitude distribution on the subreflector (actually these moments should really be complex quantities but, for the reasonably flat wave-fronts that we should have here, I think this can be ignored):

$$M_u = \Sigma_{sec}\, u_1 |E| \,/\, \Sigma_{sec} |E| \quad \text{and} \quad M_v = \Sigma_{sec}\, v_1 |E| \,/\, \Sigma_{sec} |E|$$

These should be calculated. We need to do some sums to see what an appropriate limit on these values should be. It will presumably be some smallish fraction of the subreflector radius. Note that, since the moments are only calculated for the part of the field that actually falls on the subreflector, they will actually move a good deal less than the centroid of the whole beam does when there is an error in the receiver alignment.

A second point is the definition of Σ_{all}. Should this really include the whole data set right out to the Nyquist angles or something smaller? I think this depends on the quality of the data. Clearly the region over which the data is summed should not extend out into the area where the measurement noise dominates. For example the data sets we are working with contain spurious features due to the backlash on alternate rows. It would not be helpful to include those. On the other hand we do want to include all the energy that is in the beam – there could be significant contributions from sidelobes that are some way out and we do not want to miss those. If one does use a cut-off then I guess that should be stated. I have taken the full pattern provided, but that was in fact limited to a square 20 degrees by 20 degrees.

Finally, we should really do better than simply including or rejecting points that are inside or outside the circle representing the edge of the subreflector. I have in fact implemented a mask m with $m = 1$ for $r < r_s - \delta/2$, $m = 0$ for $r > r_s - \delta/2$ and $m = 0.5 + (r_s - r) / \delta$ otherwise, i.e. a linear taper for points within $\delta/2$ of the edge. One can then set δ to be equal to a little larger than the grid spacing. (I used a factor of 1.2.) With the finely spaced data we are working with at present – grid spacing 0.1 degrees, this changes the numbers at only the 0.01% level. Doing this would presumably have a bigger effect with more coarsely spaced data.

2) If we have a cross-polar map to go with this co-polar one, then we should include the polarization efficiency. This is most simply done by summing the total power in the cross-polar pattern and applying the relevant normalization. (I don't at present understand how the patterns are normalized in the NSI data.) The combination of spill-over and polarization is then found from:

$$\text{Eta\_spill+pol} = \Sigma_{sec} |E_{co}|^2 \;/\; \{ \Sigma_{all} |E_{co}|^2 + \Sigma_{all} |E_{cross}|^2 \}$$

Note that by if we want to have the spill-over and polarisations as separate items, then according to TICRA's formulation we should proceed by:

$$\text{Eta\_spill-over} = \{ \Sigma_{sec} |E_{co}|^2 + \Sigma_{sec} |E_{cross}|^2 \} / \{ \Sigma_{all} |E_{co}|^2 + \Sigma_{all} |E_{cross}|^2 \} \text{ and then}$$

$$\text{Eta\_polarization} = \Sigma_{sec} |E_{co}|^2 \;/\; \{ \Sigma_{sec} |E_{co}|^2 + \Sigma_{sec} |E_{cross}|^2 \}$$

The alternative is to work only with the co-polar pattern in estimating the spill and make a separate estimate of Eta_polarization using the whole pattern:

$$\text{Eta\_spill-over} = \Sigma_{sec} |E_{co}|^2 \;/\; \Sigma_{all} |E_{co}|^2 \text{ and then}$$

$$\text{Eta\_polarization} = \Sigma_{all} |E_{co}|^2 \;/\; \{ \Sigma_{all} |E_{co}|^2 + \Sigma_{all} |E_{cross}|^2 \} .$$

The product of the two, i.e. Eta_spill+pol, will still be the same even though the individual numbers will be slightly different. Since both the spill-over and the cross polar loss end up on the sky and both are small there is no practical effect on the estimated performance.

3) Calculate the "amplitude" efficiency[1]. This is the ratio of the aperture efficiency that would be obtained with this amplitude illumination (and no phase errors) to that which would be obtained with a uniform illumination. This comes out to be:

$$\text{Eta\_amplitude} = (\Sigma\, m|E|)^2 / (\Sigma\, m|E|^2 \,.\, \Sigma\, m)$$

where I have put the mask in explicitly because one does have to be a bit careful if one uses anything other than 0 and 1.

One can also do things like fit a Gaussian and find how wide that is (for comparison with theory) or find the edge taper. For the first of these I simply found the values of r_G and A_G that minimize the sum of squares of the quantity $m|E|^2 - m[A_G \exp(-(r/r_G)^2]^2$. (Note that the mask is included so the fit is only to the power on the subreflector. Instead of minimizing the differences in power one could use the amplitudes – it isn't clear which is more relevant.) For the second I just averaged the amplitude of the data points for which $|r - r_s|$ is less than the grid spacing. The actual edge illumination is found to be substantially lower than that of the fitted Gaussian, e.g. 2 to 3 dB lower, which is what we expect from the optical design.

4) Calculate the Phase Efficiency. This is where take account of any phase errors in the pattern. Since we are calculating the aperture efficiency we need to add up the field over the aperture first and then square it. We can write:

$$\text{Eta\_phase} = | \Sigma\, mE) |^2 \;/\; (\Sigma\, m|E|)^2 = [\Sigma\, mE \cos(\varphi_{err})]^2 \;/\; (\Sigma\, m|E|)^2$$

Here the relevant phase is $\varphi_{err} = \varphi_{measured} - \varphi_{fit}$ where φ_{fit} is a simple function to take account of the fact that the actual phase centre of the receiver is not exactly at the point (0,0,0) as assumed.

To see the correct form of φ_{fit} one can write down the change in path p for a signal travelling in the direction (u,v) resulting from moving the phase centre to $(\Delta x, \Delta y, \Delta z)$. This is:

$$p = -\Delta x \sin u - \Delta y \sin v - \Delta z \cos u \cos v$$

The phase is then $\varphi_{fit} = k.p = 2\pi cp / f$ where k is the wave-number, c is the speed of light and f is the frequency used in the measurement.

One could in fact use the trig formulae above, but it is clear that a fit of the form $\varphi_{fit} = A\,u + B\,v + C\,r^2 + D$ is adequate for the small-angle approximation that we are making. I have done this and then made the connection with the phase centre offsets by $\Delta x = -A/k$, $\Delta y = -B/k$ and $\Delta z = 2C/k$, making the appropriate allowance for units, i.e. degrees/radians, metres/mms, etc. I think that these are the correct signs for a "natural" set of coordinates, but whether or not they actually give the right answers will of course depend on what conventions have been used in things like the phase measurement, the coordinates of the scanner and the signs used in the FT. I suspect that this can only easily be resolved by experiment – e.g. taking one set of data and then shifting the position of the source by a known amount in x, y and z and taking another set.

The constant term D presumably has no practical use. In fact we can eliminate the need for the term D by summing the cos and sin components and including them both, i.e.

$$\text{Eta\_phase} = \{ [\Sigma\, mE \cos(\varphi_{err})]^2 + [\Sigma\, mE \sin(\varphi_{err})]^2 \} \;/\; (\Sigma\, m|E|)^2 \,.$$

In that case one could just fit for the three parameters A, B and C.

At this point we need to decide whether the u and v in the expression above should be the original coordinates (u,v) in the reference frame aligned with the antenna, or the modified ones (u_1,v_1) taken with respect to the nominal "pointing". I think we should use the latter. The main reason for this is that we have a large uncertainty in Δz. This is a result of the fact that the slow f-ratio of 8 means that it takes a big change in z to produce a significant amount of curvature. If

[1] I have adopted the terminology from the TICRA report ALMA-80.04.00.00-026-A-REP section 5.1. I myself would have called this the "taper" efficiency or perhaps the "illumination" efficiency.

we use (u,v) the coupling between the terms in the fit means that any error in Δz feeds through into errors in Δx and Δy. Using u_1 and v_1 means that we obtain (Δx_1,Δy_1,Δz_1), i.e. we measure the offsets in a frame with z_1 along the beam and x_1 and y_1 perpendicular to it. This is also good because Δx_1 and Δy_1 are in fact the right quantities to use in finding the corresponding shifts of the beam on the sky. To do this, simply multiply them by the "plate scale", which is 2.148 arcsec/mm for the 96m focal length of the ALMA 12m antennas.

Since what we want to know is the efficiency at the peak of the beam, it is clear that the fit should be done by maximizing the phase efficiency. (In fact I minimized 1 – eta_phase.) The problem is non-linear and the measured phase is often "wrapped", so minimizing is φ_{err}^2 is not a good thing to try. When maximizing the phase efficiency the wrapping can still be a problem, but this is solved by starting with a more restricted region at the middle of the beam to get a first guess and then enlarging it to the whole of the subreflector. In all the data I have worked with so far, make a first attempt with r_s set to 2 degrees and then increasing it to 3.58 degrees and solving again was sufficient to get to a correct solution. In almost all of those, the correct solutions gave phase efficiencies above 99% whereas the local false maxima were much lower, so it is easy to see when one has gone wrong even without graphics.

The total predicted aperture efficiency (assuming a perfect antenna) is of course the product of the four individual ones:

Eta_aperture = Eta_spill . Eta_pol . Eta_amp . Eta_phase

I think this list – the efficiencies, the moments and the offset (Δx,Δy,Δz) of the phase centre – includes all the important quantitative information, but there is of course a lot more to be learned from plots of the data, both what is originally recorded in the near-field and the far-field patterns and including both phase and amplitude. In general I suppose that, for the far-field phase, the most useful thing would be to plot the phase after subtracting the model derived above.

Richard Hills 16th June 2008 revised 22nd June