Treatment of Non-Repeatable Residual Delay

The present form of the specification (Section 5.6.2 attached for convenience) seems to state that some contributions to the path errors, specifically the "slowly varying ones" can be doubly differenced before inclusion in the error budgets. By "doubly differenced" I mean that the contribution to be counted is only the change over some time Δt (given as 3 minutes) in the difference in the path that occurs due to an angular movement $\Delta\theta$ (2 degrees). I believe that this double differencing should not have been permitted by the specifications.

My basic argument is as follows:

Consider that a particular antenna has a path length residual, $p(t,\theta)$, which is a function of time, t, and some generic angular parameter, θ. (I'll outline the proper expressions later.) The reason that p is described as a residual is that it is the deviation of the path length from that for the nominal antenna, e.g. one with no deformations due to temperature, wind, etc. Note that we do not take any account here of the fact that we are really only concerned with differences between antennas: the path errors are supposed to be considered on each antenna individually. (I shall return to this point later.)

Now suppose that we are doing some form of switching and that this is done so rapidly that the change in p with time over a switching cycle is negligible. The relevant path error is therefore

$$\Delta p_\theta = (\mathrm{d}p/\mathrm{d}\theta)\,\Delta\theta\,.$$

Obviously this shows up as a phase error – effectively we measure the wrong distance between our calibrator and the object of interest. If p were completely constant with time, then this term would only change as a result of the changing geometry as the source moves across the sky – that is to say that Δp_θ is just determined by the details of the dependence on the elevation and azimuth angles. In general this would look the same as an error in the baseline. It would therefore be taken out as one of the parameters fitted in finding the baseline. Strictly speaking, however, this would only apply to errors that were constant right through from the time of the baseline solution until the time of the observation.

If this term Δp_θ does in fact vary with time, then clearly all measurements will be effected. This is obvious in the case of a position measurement – we get a error of order Δp_θ / baseline – and presumably variations on timescales of an hour or two would show up as spurious structure on synthesis maps in the way that phase errors usually do. I see no place here for saying that we are only interested in variations of this term on periods of less than three minutes.

It seems to me that the time period of three minutes is intended to represent the time for which we may wish to integrate coherently on a source without moving to a reference source. It is true that for many observations we will be switching more frequently than that, but we certainly do not want to be constrained to do that all of the time. Recall also that these specifications are intended to cover the ACA as well as the antennas for the 64-element array, and that the ACA will not usually be used in a fast switching mode. Any changes in the path delay will cause a loss of coherence. The figure of 15 microns is comparable to the specifications in other areas (water vapour radiometer, local oscillator stability, etc.)

The term to be included here is presumably

$$\Delta p_t = (\mathrm{d}p/\mathrm{d}t)\,\Delta t\,.$$

This will at least give us the first order effect on the change in path for a slowly varying quantity. This would appear directly as a phase error that accumulates over the observing period Δt. Again, I see no argument for saying that we can difference this term with respect to angle.

Instead it seems to me consistent with our general policy that both these terms, Δp_θ and Δp_t, should appear in our error budget. They are in some sense orthogonal (see below) so the full residual error should be the RSS of these two terms, plus of course the ones due to wind, etc., to which these differencing arguments do not apply.

Various mitigating arguments can be made. For example, in the case of time variation, one could say that it is actually only the deviation from the mean path error during the three-minute period that matters, or even that one can interpolate linearly between the phase calibration measurements at either end of the period so only the second and higher time derivatives are important. Similarly, for the angular dependence one could use several calibration sources around the object and fit a function for Δp_θ. In general, however, I think that our policy is to make the antennas good enough that we do not have to do these complicated things.

Obviously we can't consider changing the specification at the moment, but if others agree with this analysis we will have to fix it up at the earliest opportunity.

More details:

For completeness I have run through the derivation of the terms, replacing the generic angle θ by the Elevation and Azimuth pointing directions E and A.

For convenience, the path can be broken down into three components: p_1, the part above the elevation axis; p_2, the component due to movements of the elevation bearing with respect to the azimuth axis; and p_3, any movement of the azimuth axis.

Now p_1 will contain things like the overall expansion of the dish, together thermal effects in the part of the cabin which lie between the elevation bearing and the BUS. This term will have no direct angular dependence, so it subtracts out completely as we move from source to reference. Note that there could be an indirect dependence if there were for example some gravitational terms in the dish deformations that were not being taken out correctly. The temperature distribution could also be affected by a change in pointing, but one would not expect this to be a large effect for 2 degrees. Neglecting these finer points, the derivative of p_1 with respect to the angles E and A is zero so it contributes only to Δp_t and not to Δp_θ.

For p_2 we will have a vertical and a horizontal components Δz and Δh due to movements of the elevation axis where h is the direction along the Azimuth where the antenna is pointing. Note that these would include things like expansion and bending of the yoke but also the effects of tilts in the base and its expansion. Resolving along the line of sight we get:

$$p_2 = \Delta z \sin E + \Delta h \cos E$$

and clearly the derivative is

$$\mathrm{d}p_2 / \mathrm{d}E = \Delta z \cos E - \Delta h \sin E,$$

with no direct dependence on Azimuth.

(The suggestion that the terms Δp_θ and Δp_t are orthogonal, in some loose sense, comes just from the fact that the sin and cos terms are interchanged between the two equations above. Obviously this is not supposed to be anything rigorous but it is weak justification for applying the usual RSS treatment of the terms.)

Finally we need to consider movements in the base that shift the azimuth axis by amounts Δx and Δy. These will not rotate with the source so they produce a term like

$$p_3 = (\Delta x \cos A + \Delta y \sin A) \cos E$$

This does have an azimuth dependence so we get non-zero derivatives with respect to both azimuth and elevation. In practice however these lateral shifts in the base should be extremely small.

As already noted the forms of the terms p_2 and p_3 are the same as baseline errors.

Returning briefly to the fact that we are really only interested in the differences in these effects between different designs, it is clear that we are to some extent being conservative in insisting that the antennas meet the requirement individually. On the other hand some effects will not be the same at different antennas across the array. This is certainly true for the wind contributions and one can imagine cases where it is true for some thermal effects – for example where one antenna shadows another or when cloud covers part of the site. I would comment that this is a prime example of where having two different antenna designs, with different levels of thermal response and different time constants, would cause a lot of problems. I would re-emphasize the point made above that the movements of the location of the axis crossing point will look like a baseline error and will only be properly removed by performing a baseline fit.

I would also draw attention to the "Repeatable Residual Delay" (5.6.1). This is specified at 20 microns and is again only thought to be of concern over an angular step of 2 degrees. I am concerned that in principle this could allow much larger errors on wider angles and that these would make it extremely difficult to find a proper baseline solution. The implication of the way this is written is that these effects could be measured and stored, perhaps in a look-up table. Is this really what we intend to do? Do the software people know about this?

I note also that this is another example of something that would not cancel with non-identical antennas.

Richard Hills 9th June 2004

5.6 PATH LENGTH ERROR

Path length errors must be considered since the antennas will be used in an array. Path length (also called delay) errors are defined as follows. Consider a plane wave arriving at the antenna from the direction of the boresight. Define the "excess delay" of the antenna to be the difference between the arrival time of that wave at the secondary focus (via the main reflector and subreflector) and its arrival time at an arbitrary reference point fixed with respect to the ground as if the antenna were not present. (It is convenient to choose the reference point along the azimuth axis. If the boresight axis, elevation axis and azimuth axis all intersect, then choosing that intersection as the reference results in an excess delay that is nearly constant.) The excess delay (expressed as a path length) for the nominal antenna in the absence of environmental perturbations, shall be computed as a function of boresight direction over the full range of azimuth and elevation. This shall be defined as the "nominal excess delay function." Now define the "residual delay" as the difference between the actual excess delay of a particular antenna under existing conditions and the nominal excess delay. The residual delay is limited by the specifications of this section. The residual delay has a repeatable and a non-repeatable component.

5.6.1 REPEATABLE RESIDUAL DELAY

The repeatable residual delay is caused by the difference in gravity deformation between an antenna and the nominal antenna (for example, this could be caused by differences in the material properties of an antenna compared to the nominal material properties), axis alignment errors, bearing runout, bearing alignment, and similar errors, which repeat as a function of antenna position and can be corrected using a computer delay model.

It applies:
The repeatable residual delay for an antenna shall not change by more than 20 micrometers when the antenna moves between any two points 2 degrees apart in the sky.

5.6.2 NON-REPEATABLE RESIDUAL DELAY

The non-repeatable residual delay is the delay component that varies with time or is not repeatable as a function of antenna position. It is caused by wind, effects of temperature differences and temperature changes, acceleration forces, bearing non-repeatability and other sources of non-repeatable errors. The Contractor may include metrology equipment in the antenna design that can be used to estimate the residual delay in real time. In that case, both the measured values and the results of a calculation estimating the residual delay shall be provided to ALMA via the digital interface (see Section 6.10.5.1). If this is done, then the estimate shall be subtracted from the actual residual delay for the purpose of meeting the specifications of this section. Further, for slowly varying sources of residual delay, but not for wind induced residual delay, the contribution to the residual delay budget may be limited to the differential residual delay over a solid angle of 2 degrees radius on the sky and then only the change in that differential delay over a 3 minute period when tracking at the sidereal rate.

The non-repeatable residual delay under Primary Operating Conditions (section 4.4.3) must be less than 15 micrometers RSS when tracking an astronomical source at sidereal rate.

5.6.3 COMPUTATION OF THE NON-REPEATABLE RESIDUAL DELAY

The non-repeatable residual delay shall be computed both for nighttime and daytime conditions by use of error budgets.

a) Computation of the wind induced delay:
Calculate the quasistatic delay error for each of the eleven wind directions shown in Table 5.3.2.3-a, with the average wind speed defined for the Primary Operating conditions. Compute the weighted RMS of these eleven wind directions with the weighting factors defined in the table to obtain the steady state component of the wind.
The gust component of the wind can be computed by scaling the quasistatic values obtained for the steady state.

b) Computation of the thermal effects (daytime)
For daytime Primary Operating Conditions a computer thermal model of the antenna shall be used to determine the worst case non-repeatable delay error due to temperature differences and temperature changes in the structure. The structural temperature differences shall be calculated using a computer thermal model of the antenna, assuming a wind speed of < 5 m/sec

c) Nighttime residual delay
The nighttime residual delay shall be the quadratic sum (RSS) of all the individual sources of delay including the steady state and gust wind contribution according to the Primary Operating Conditions of Section 4.4.3. No thermal effect contribution shall be included in the nighttime error budget. The effect of the steady state and gust wind shall be computed according to the methodology and with the weighting factors of Section 5.3.2.3.

d) Daytime residual delay
The daytime pointing residual delay error shall be the quadratic sum (RSS) of all the individual sources of delay error including the steady state and gust wind contribution according to the Primary Operating Conditions of Section 4.4.3 and the thermal effects contribution. The effect of the steady state and gust wind shall be computed according to the methodology and with the weighting factors of Section 5.3.2.3.
The daytime thermal residual delay is defined to be 75 percent of the thermal effects on delay computed for the worst case under point b) above.