NATIONAL RADIO ASTRONOMY OBSERVATORY
VLA PROJECT
CHARLOTTESVILLE, VIRGINIA

November 1977

VLA ELECTRONICS MEMO NO. 164

VLA PHASE STABILITY -- A VISITOR'S LOG

J. Granlund

In spite of almost universal conviction to the contrary, coming from out of town does not make the visitor an expert. In the case of a system as complex as the VLA, the visitor -- who does not carry a detailed block diagram in his head; has not committed to memory the results and implications of the last six months of testing; and even who is unfamiliar with the local language in use for describing components, system variables, and test results -- this visitor is at a distinct disadvantage. What follows is the log of one visitor's stay at the site from 10/17/77 to 10/20/77, in search of causes and solutions to the VLA phase instability problem then current.

At the start, it was clear that several system variables are properly called "phase", preferably with a qualifier. But in conversation the qualifier may be omitted, and worse, the deleterious effect of a 20° change in one phase variable may be equated, by implication, to the effect of a 20° change in any of the others. The phases that count -- those whose instrumentally-caused variations must be minimized, or at least held to the VLA system specification of 1° rms per GHz of observing frequency -- are the phases of the correlation coefficients measured while observing a known point source through a clear atmosphere. For any baseline, the phase abnormalities in the dishes and feeds, front ends, L.O. distribution subsystem, waveguide and I.F. transmission subsystems, 90° phase shifters, samplers, delay subsystem, correlators proper, and fringe rotation subsystem may all contribute to correlation coefficient phase instability. In particular, the proportionality constants interrelating L.O. distribution subsystem and waveguide transmission subsystem phase errors to the correlation coefficient phase error may be quite different from unity.

Before the visit, I learned from the Charlottesville branch of the VLA grapevine that some consider the phase instability problem to be the result of -- or at least to behave like -- a myriad of loose connections in both analog and digital circuitry. Even the most optimistic conceded that there are several causes. Correlation coefficient phase error varies smoothly -- with Fourier components having periods greater than an hour -- and it also contains jumps, several per day. Both of these variations can far exceed the 1°/GHz specification. Larry D'Addario explained that the L.O. distribution subsystem's round-trip phase correction circuitry needed modification when it was received at the site, that the modification has only recently been installed in the vertex rooms of antennas 3 and 5, and that the correction is still not being made. He noted, however, that round-trip phase at 600 MHz is being measured; it exhibits a smooth 10° to 20° peak-to-peak diurnal variation that could probably be predicted and thus removed to within a few degrees; and the amount of this variation is independent of the length of the waveguide run to the antenna in question, so its source is evidently thermal expansion and contraction of the waveguide run from the mainline coupler to the vertex room. I now realize that it is planned for the round-trip phase correction "circuitry" to measure round-trip phase at 600 MHz and to deliver this measurement to the ModComp, which in turn, completes the correction by modifying the fringe phase, or at least that component of fringe phase due to the antenna in question. In other words, the correction loop is finally to be closed through the computer. The missing correction, then, seems to be "merely" the result of incomplete software, not of yet-to-be-developed hardware.

On arrival at the VLA site, I found that antennas 3 and 5 were being used exclusively for testing. The latest modifications -- as suggested above, primarily to the feedback connection around the step-recovery diode in the inner phase-locked loop of the L.O. distribution subsystem -- have been installed at these antennas. So far, all phase stability testing has been done at C-band; the C-band front ends, mixers, and following amplifiers become the I.F. for operation in the other three bands. With trouble at hand, I approve heartily of these testing arrangements: Concentrate effort on one collection of equipment that is as simple as possible, yet still exhibits the trouble, and that contains the latest modifications. I also approve of the current top-priority effort to find and fix the trouble; this will reduce the number of units requiring possible major modification to eliminate the trouble. Finally, I suggested that the current heavy observing schedule may be unreasonably delaying the testing, but I found no one who agreed with me. All felt that keeping the equipment in use constitutes serendipitous testing of reliability, that almost completely separate groups are involved in the observing and the testing, and that tests that would interfere with an observing run -- waveguide transmission through an isolated section of guide, for example -- can well wait until the end of the run.

I was shown two 8-track chart recordings of test data pertinent to the phase stability problem. The first recording was started on 10/4/77 and the second on 10/12/77. Both had a time scale of 1 mm = 1 minute and covered a period of some forty hours. Ernie Caloccia and John Archer showed me two seemingly identical swept-frequency responses of waveguide transmission from the control building to the vertex room of antenna 5 at DW2 and, I think, to the mainline coupler supplying DW2, both in the range between 37 and 38 GHz. (Seemingly identical -- sure, but I'll bet the correlation coefficient between the two even after the slow variations corresponding to the use of an unleveled sweeper have been removed is well less than 0.5.) I also heard as many ideas and descriptions of related tests as was possible.

As a result of this four-day exposure, I have formed a four-day opinion about the VLA phase instability problem that I saw: If the following three causes were substantially eliminated, VLA phase stability at C-band would fall close to and perhaps even within specification:

(1) Waveguide transmission phase ripples are the sole cause of correlation coefficient phase jumps. These ripples need only be reduced enough to eliminate the jumps.

(2) Waveguide line-length changes cause slow variations in correlation coefficient phase that were anticipated in the system design. The round-trip phase correction loop should be closed to eliminate these variations.

(3) Phase changes in the front ends, mixers, and following amplifiers in the vertex room add directly to the correlation coefficient phase. These changes should be substantially reduced. Without having looked, I suspect the paramps.

Caution: This memo has just ceased to be a log. From here on, it can only contain the biased mumblings of the self-styled expert from out-of-town that are intended to justify his instant opinion. I'll discuss the above causes in order.

(1) The connection between phase jumps and waveguide ripples is a bit tricky because of the uncertainty as to whether the 600 MHz loop or the 5 MHz loop at the vertex room is controlling the 5 MHz VCXO there. Normally the 600 MHz loop controls, but if the open loop phase difference $\Delta\phi_5$ between the VCXO and the 5 MHz tone transmitted on the 1200 MHz sideband from the control room exceeds 2° (recently increased from 1°) either way, the 600 MHz loop is opened and the 5 MHz loop is allowed to lock. Immediately thereafter, the 5 MHz loop is opened, and the 600 MHz loop is again allowed to lock. But because adjacent lock positions of the 600 MHz loop correspond only to

$$360° \cdot \frac{5\ \text{MHz}}{600\ \text{MHz}} = 3°$$

changes in $\Delta\phi_5$ and because $\Delta\phi_5$ must have started 2° away from $\Delta\phi_5 = 0$, reached when the 5 MHz loop is locked, the final locked position of the 600 MHz loop must be a whole revolution different from the starting position at which $|\Delta\phi_5|$ first exceeded 2°: $\Delta\phi_5$ will have jumped 3°. A multiple of this phase jump adds to the correlation coefficient phase.

Waveguide transmission phase ripples causing such jumps have not been measured directly. Instead, waveguide transmission magnitude ripples have been measured, as was noted above. The swept-frequency recordings show a strong periodicity with a period somewhere between 5 and 10 MHz and a peak-to-peak variation of perhaps 0.5 dB. By itself, this suggests a double reflection -- or two mode transitions -- producing an undesired signal 30.82 dB down from the desired signal, the two signals producing 1.65° peak phase ripples. If the two discontinuities produce equal reflections of the TE_{01} mode, each with a 15.41 dB return loss, the VSWR due to either reflection would be 1.41, which is certainly excessive; corresponding remarks apply if the discontinuities produce mode transistions. The recordings also show a periodicity of some 100 MHz with a somewhat smaller peak-to-peak variation; I presume that the still larger periodicities are caused by the use of the unleveled sweeper. The similarity of the two recordings suggests that the reflections -- or mode transitions -- are taking place in the 60 mm mainline waveguide, not in the shorter runs from the couplers to the vertex rooms. In the worst case, 1.65° phase ripples can cause a 1.65° error in the 5 MHz phase received from the control room and a 3.3° error in the 600 MHz phase. But since the latter must be divided by 120 before its contribution is added to $\Delta\phi_5$, I'll confine attention to the phase of the 5 MHz tone that is transmitted on the 1200 MHz sideband from the control room.

The first point to be noted is that the two-signal 1.65° peak phase ripples considered above can only cause a 1.65° error in the received 5 MHz phase if the 1200 MHz sideband carrying the 5 MHz happens to be sitting on a peak or a trough of the magnitude ripples and if the ripple period happens to be 20 MHz or an odd sub-multiple -- like 6-2/3 MHz -- thereof. In the table below I have calculated, for 60 mm pipe, the TE_{01} reflection spacings and TE_{01} to TE_{0n} mode-transition spacings, for the first few circularly-uniform modes, that produce a 20 MHz ripple period for a frequency sweep near 37.51 GHz. These spacings and all of their odd multiples are the ones most likely to cause trouble for transmission in waveguide channel 5.

Mode	Cut-off Freq. (GHz)	Source of Undesired Signal	Spacing (m)
TE_{01}	6.094	TE_{01} double reflection	7.60
TE_{02}	11.158	TE_{01} - TE_{02} transitions	468.70
TE_{03}	16.180	TE_{01} - TE_{03} transitions	177.32
TE_{04}	21.191	TE_{01} - TE_{04} transitions	92.77

Next note that, since the 1.65° phase error calculated above is less than 2°, it is not enough to trigger the logic in L5 to seek a new 600 MHz lock position and thereby to cause a phase jump. No wonder only a few phase jumps per day have been found! Are these few due to the ganging up of ripples resulting from two or more pairs of nearly equally-spaced reflections? Or, alternately, how was the 5 MHz phase shifter in L1 set? If it was set to make $\Delta\phi_5 = 0$ at an outside air temperature that maximized the 5 MHz phase error, the system's resistance to waveguide transmission phase ripples can be doubled by resetting the phase shifter when the 5 MHz phase error is zero, i.e., when the 1200 MHz sideband is half way between a peak and a trough of the magnitude ripples. Presuming that $\Delta\phi_5$ passes through one or more ripple cycles in the course of a day, the phase shifter should be set to make the positive and negative peak excursions of $\Delta\phi_5$ equal.

Larry D'Addario showed me an alternative method of resolving L.O. subsystem phase ambiguities that does not require the transfer of VCXO phase control between the 600 MHz and 5 MHz loops, and so avoids phase jumps altogether. $\Delta\phi_5$ from the vertex room and round-trip phase at both 600 MHz and 5 MHz from the control room are used. But as I was leaving, I saw Larry waxing dubious about the required accuracies of measurement of both $\Delta\phi_5$ and the 5 MHz round-trip phase. An obvious, but much more extensive modification that reduces the required measurement accuracy would be to add a 50 MHz "hand" to the clock. But at this point, my recommendation is to chase down the waveguide transmission phase ripples and to experiment with the phase shifter in L1 before making major modifications.

(2) During my visit, I was pleased to see Durga Bagri struggling with the algorithm required to close the round-trip phase-correction loop, but I was disappointed that such struggling was only then starting. Durga asked me to check his work. I did, and I told him that it looked fine to me. I have since had doubts about two aspects of Durga's algorithm that I have investigated as described below. The first produces no cause for alarm, but I shall red-flag the second because it seems to call for major changes in system thinking and organization, if not in hardware.

(2.1) The algorithm does not account for the dispersiveness of waveguide transmission. How should this be done, and what difference would it make? In settling this question, it is useful to note that the purpose of the round-trip phase-correction loop is <u>not</u> necessarily to remove all instrumental phase error associated with the waveguide runs, but rather only that part of it that varies too rapidly to be removed by calibration. Diurnal and more rapid changes should be removed; annual changes need not be. Evidently, as was noted above, the diurnal changes that have been measured in round-trip phase are caused primarily by thermal expansion and contraction of the waveguide runs from the mainline couplers to the vertex rooms. I understand that these runs use the TE_{01} mode in 20 mm round pipe, for which the cut-off frequency is

$$f_c = 18.282\,394 \text{ GHz.}$$

In dealing with the present question, it is then simpler to model the whole run from vertex room to control room as using the smaller pipe. A fixed -- or perhaps annually varying -- phase error may well be introduced by this substitution for the larger pipe of the mainline run, but at least the diurnal variations should be properly accounted for and removed when the loop is closed.

The phase lag introduced by an increase of length $\Delta \ell$ in a waveguide run is

$$\Delta\phi = \frac{2\pi f \Delta \ell}{v_p} = \frac{2\pi f \Delta \ell}{c} \sqrt{1 - (f_c/f)^2} = \frac{2\pi \Delta \ell}{c} \sqrt{f^2 - f_c^2} \equiv T \sqrt{f^2 - f_c^2}$$

in which v_p = phase velocity

f = frequency of propagation in the waveguide

c = speed of light

T = $2\pi \cdot$ (delay increase due to $\Delta\ell$).

I'll use T as the line-length-change unknown in solving for the correction to be added to the channel A - channel C phase difference. I'll use Durga's numbers throughout. Since the phase difference should be zero when the input signals are identical, the correction itself will be $\phi_C - \phi_A$. I'll produce three solutions and three sets of partial results corresponding to

- waveguide not dispersive (or $f_c = 0$) -- Durga's solution
- waveguide dispersive, using channel 5
- waveguide dispersive, using channel 1

and I'll display them as above so they can be compared. I'll work in MHz as Durga did.

The waveguide frequencies used for transmitting the 1200 and 1800 sidebands are

	1200		1800
1200	f_a -- doesn't affect solution	1800	$f_a + 600$
	37,210		37,810
	27,610		28,210

The phase increase -- due to line length increase -- of the 600 phase-reference tone produced at the vertex room by beating the 1200 and 1800 sidebands received there from the control room, where they were transmitted with zero initial phase, is

$$\phi_{600} = \begin{cases} -600\ T \\ -687.\ 153\ 70\ T \\ -794.\ 144\ 89\ T \end{cases}$$

The 600 round-trip phase increase is exactly twice the above set of numbers, in spite of waveguide dispersion. These and other partial results yet to emerge show strong effects of dispersion, but don't jump to conclusions until you see the solutions!

It is damn tedious to work through the whole VLA block diagram, including the modems and the L.O. signals delivered to them, multiplying phase increases by frequencies and adding -- or subtracting -- them as appropriate to represent mixing, in order finally to calculate $\phi_C - \phi_A$. Instead, let me reduce this effort by noting that $\phi_C - \phi_A$ would be zero regardless of line length and dispersion if, somehow, channel A were transmitted back to the control room at the frequency of the 1200 sideband and channel C were transmitted on the 1800 sideband. This would eliminate the need for the 100 MHz and 250 MHz fine-tune offsets supplied from the master oscillator. Returning to the problem at hand, $\phi_C - \phi_A$ is then zero plus the correction to ϕ_C for transmitting below the 1800 sideband minus the correction to ϕ_A for transmitting above the 1200 sideband minus the net correction for using different multiples of the 600 phase-reference tone as L.O. signals to produce the A and C signals that are transmitted through the waveguide. Numerically, these corrections all add, as follows:

$$\text{Channel A Corrections} \begin{cases} 140\ T \\ 160.\ 643\ 94\ T \\ 186.\ 460\ 18\ T \end{cases} \qquad \text{Channel C Corrections} \begin{cases} 210\ T \\ 240.\ 115\ 42\ T \\ 276.\ 498\ 89\ T \end{cases}$$

$$\text{L.O. Corrections} \begin{cases} 350\ T \\ 400.\ 839\ 66\ T \\ 463.\ 251\ 19\ T \end{cases}$$

The total correction is

$$\phi_C - \phi_A = \begin{cases} 700\ T \\ 801.\ 599\ 02\ T \\ 926.\ 210\ 26\ T \end{cases}$$

Durga used a case in which the measured round-trip phase was $2\phi_{600} = 20°$. From this, the unknown T and the correction to be added to $\phi_A - \phi_C$ are

$$T = \begin{cases} -0.016\ 666\ 667 \\ -0.014\ 552\ 785 \\ -0.012\ 592\ 161 \end{cases} \qquad \text{Correction} = \begin{cases} -11.666\ 667° \\ -11.665\ 498° \\ -11.662\ 988° \end{cases}$$

Note that, in spite of the variety of the partial results, the corrections are substantially all the same. I conclude that, unless the diurnal variations in round-trip phase increase by one and perhaps two orders of magnitude, Durga's algorithm that neglects waveguide dispersion will be adequate.

(2.2) What should be the form of the algorithm that is used to close the round-trip phase-correction loop? First, since channels A and C will always carry signals from one observing frequency band, while channels B and D will always carry signals from a different band, 27 x 4 different signals will reach the control room, and as many as 351 x 8 different correlations may be measured from these signals; the second number is 26 times the first. To reduce by a factor of 26 then, the number of corrections to be dealt with, a strong effort should be made to make phase corrections on the individual channels, rather than on the measured correlations. Unfortunately, the number of Fringe Generators -- L7 -- that are used to make phase corrections on the individual channels in a vertex room has recently been reduced from four to two, which seems to block the correcting of the four individual channels. But let me argue once again for the individual channel corrections: If they are not made, and the measured correlations are corrected instead, large phase errors will pervade the system, making smaller errors due to other causes difficult to detect and track down. In short, I suppose that the VLA stratagem "Let's just plug the feed horns into the computer." _is_ feasible, but isn't that carrying things a bit too far?

Next I'll discuss an algorithm for correcting the phases of the individual channels; if the final decision is to correct the phases of the measured correlation coefficients, this algorithm will still point the way. As was implied above, no correction would be needed if channels A and B were transmitted at the frequency of the 1200 MHz sideband and channels C and D were transmitted on the 1800 MHz sideband; phase corrections result only from the offsets from these impossible conditions. For starters, it is certainly impossible to transmit each spectral component from a band of noise 50 MHz wide on the same frequency, regardless of that frequency! It is evidently necessary to identify _which_ spectral component, and it will be convenient to do this by baseband frequency f: f is the frequency of the spectral component in question after the last mixing operation, just before the low-pass band of noise is phase-split into quadrature components and sampled. Durga used f = 40 MHz in his example, but I'll be careful not so to constrain myself. The d-c component at baseband came through the waveguide at a frequency higher than the 1200/1800 MHz sideband by the fine-tune offset f_o, but watch the sign of f_o! Tones of frequency $|f_o|$, derived from the master oscillator, are produced by L17. The table below gives values of f_o corresponding to Durga's example, extended to include channels B and D.

Channel	I.F. Range (MHz)	f_o (MHz)
A	1300 - 1350	+100
B	1400 - 1450	+200
C	1550 - 1600	-250
D	1650 - 1700	-150

As was shown in (2.1), waveguide dispersion can be neglected in dealing with the phase-correction loop, and if so, the coefficients of T used in (2.1) become simply the various frequency offsets. The round-trip phase measured at 600 MHz becomes

$$2\phi_{600} = -(1200 \text{ MHz}) \cdot T,$$

and the phase error -- that must be subtracted this time to complete the correction -- the phase error of any of the four channels becomes

$$\phi_{ch} = -(2f_o + f)\ T.$$

Why not also 2fT? Well, f_o needs to be doubled because the d-c component transmission through the waveguide is at a frequency f_o higher than the frequency of the appropriate 1200/1800 MHz sideband and the net L.O. tone generated from the phase-shifted 600 MHz reference at the vertex room must also be raised by f_o. But, although the component that emerges at baseband at frequency f is transmitted through the waveguide at a frequency higher by f than that used for the d-c component, no vertex room L.O. frequency change is required. Eliminating T from the two equations yields the phase error

$$\phi_{ch} = 2\phi_{600} \cdot \frac{2f_o + f}{1200 \text{ MHz}}.$$

The phase error corresponding to the first term in the numerator,

$$\phi_{ch1} = 2\phi_{600} \cdot \frac{f_o}{600 \text{ MHz}},$$

is perfectly comprehensible, but, for the second term, what's this with a "phase" error proportional to baseband frequency? Well, it's not a phase error at all; it's a baseband delay error: The delay

$$\tau_{ch1} = \frac{2\phi_{600}}{1200 \text{ MHz}} , \quad 2\phi_{600} \text{ in revolutions}$$

should be added to the baseband delay of the channel in question to complete the "phase" correction.

If $2\phi_{600}$ could really be measured -- not only the fractions of a revolution, but the integer number of revolutions as well -- then τ_{ch1} would be the total electronic delay correction to be used for the channel in question. But in practice, only the fractions can be measured; the integer number must be determined by the calibration procedure.

I understand from Art Shalloway that the VLA delay subsystem provides delay steps of 10 ns/16 = 0.625 ns. Considering that the maximum measurable $2\phi_{600}$ is 1/2 revolution, the maximum delay correction associated with the round-trip phase measurement will be

$$\frac{1/2 \text{ revolution}}{1200 \text{ MHz}} = 416.67 \text{ ps} = 0.667 \text{ delay steps},$$

so that seldom, if ever, will the round-trip phase measurement call for a one-step delay change! The delay steps must then be pretty coarse. Suppose that, at worst, the delay error is 1/2 delay step, and suppose that the d-c end of the baseband has been properly phased. Then the phase error at the 50 MHz end of the baseband is

$$50 \text{ MHz} \cdot 1/2 \cdot \frac{10 \text{ ns}}{16} = 0.015\,625 \text{ rev.} = 5.625°.$$

Well, let's see. If the phase-correction algorithm were modified to phase correctly at band center, depending on the delay actually used, then only half of this phase error would appear at the d-c and 50 MHz band edges. But there's no way to meet the phase-error specification at L-band!

Because of the doubt raised by the foregoing calculation, I have calculated -- not for the first time, I trust -- the phase errors caused by the long runs of unequalized mainline waveguide. My calculations give the phase differences between a linear approximation, involving the correct center frequency phase shift and group delay, and the exact phase shifts at the lower and upper band edges of channel A, as transmitted in the waveguide. It was a good decision to use the higher-frequency waveguide channels, that exhibit less dispersion, for the longer runs. I have supposed that waveguide channels 1 and 11 are reserved as spares. My results are tabulated below.

Waveguide Channel	Carrier Frequency (MHz)	Waveguide Length (m)	$\Delta\phi(f_1)$ (degrees)	$\Delta\phi(f_2)$ (degrees)
2	28790	484.00	0.263	0.263
3	31210	1589.92	0.680	0.678
4	33590	3188.09	1.094	1.092
5	36010	5222.90	1.458	1.455
6	38390	7659.48	1.768	1.765
7	40810	10472.87	2.015	2.013
8	43910	13643.92	2.111	2.109
9	45610	17157.23	2.373	2.369
10	47990	21000.00	2.499	2.497

Well, let's see again. We are dealing with a phase error that is a concave-up function of frequency; both the error and its first derivative vanish at the center frequency of the channel A band. If the center-frequency phase shift were modified, the peak phase error due to the unequalized 21 km run could be kept down to 1.25°, but again, there's no way to meet the phase-error specification at L-band.

It seems to me that the phase error specification of 1° rms per GHz of observing frequency can barely be met at C-band, and that the phase error at L-band will actually be worse because of the parametric up-converters that will be added to the signal paths. If this is our intention, let's admit it to ourselves. If not, let's produce the hardware to provide finer delay steps and to phase-equalize the long waveguide runs.

(3) The 8-track chart recordings noted above are records of a test in which a C-band signal generator is coupled into the feed horn of antenna 5 so as to excite both channels A and C. The signal generator is switched alternately on for 5 minutes and off for 10. This allows measurements of gain and phase difference between the two channels to be made from the feed horn to various points in the system. One track of the first recording carried gain of one of the channels through the front end, mixer, and vertex room I.F. amplifier; it showed 2 dB peak-to-peak smooth changes -- with periods in the order of several hours. The second recording carried a track of gain difference between the two channels; it showed 0.5 dB peak-to-peak changes, also smooth. Both recordings had tracks of phase difference between the portions of the two channels in the vertex room. These also showed smooth changes -- 10° peak-to-peak on the first recording and 8° on the second.

Both the gain and the phase changes recorded here are excessive; the phase changes, which add directly to correlation coefficient phase, need to be reduced by an order of magnitude if the system specification is to be met at L-band. Because it is regenerative, I suspect the parametric front end amplifier, but I recommend that further tests to isolate the trouble should be run before concentrating on the paramps.

There is much more to be said about the VLA phase stability problem, but rather than waiting to create it all here, I elect to cut this memo short so that it will still be more-or-less current when it reaches those who may find it helpful.