ALMA Memo No. 287

Fringe Tracking, Sideband Separation, and Phase Switching In The ALMA Telescope

Larry R. D'Addario
2000-Feb-15

Introduction

This report describes the conceptual design of certain parts of the local oscillator subsystem of the ALMA telescope, namely those parts concerned with fine control of the phase of the LO signals, along with related considerations for the correlator and the real-time software. Precise control of the LO phase is needed for fringe tracking, which is simply a matter of causing the LO to follow the variation in delay from the source to the receiver as the earth rotates. The same hardware can be used to effect suppression of the interferometer response from the unwanted sideband of the first mixer, which is expected to be incompletely suppressed in the front end; and to suppress certain spurious signals that might get introduced into the signal path after the first mixer. It can also be used to allow a small amount of fine adjustment of the receiver's center frequency.

Quantitative parameters mentioned here are consistent with the ALMA system design as of this writing. Some of these may be changed for practical reasons during the course of development, but the overall design is expected to remain as described here.

Hardware

Figure 1 shows a simplified block diagram of the signal path for a single baseband channel of one antenna of the telescope. There are two frequency conversions before digitization, followed by digital transmission to the central correlator. Delay tracking is accomplished digitally within the correlator. Each of the LOs is synthesized at the antenna using reference signals distributed from the center. (In an alternate scheme now under study, the first LO would be synthesized centrally and transmitted on optical fiber. Some implications of this are discussed at the end of this report.) Although the first LO frequency is as large as 938 GHz and the second LO is tunable over 6-14 GHz, each of them is based on a phase locked loop that includes a small offset frequency generated by a direct digital synthesizer (DDS). The DDSs allows very fine frequency and phase control, and they form the basis of the features discussed here. For a discussion of the principles of operation of DDSs, see e.g. [1].

The DDSs are required to have these properties: 125 MHz clock frequency; 32b frequency and phase registers; phase offset register of 2b or more; ability to initialize all registers to a known state (typically zero) on a known clock cycle; ability to load the frequency register and/or the phase offset register on a known clock cycle.

The DDSs are under the control of a local microprocessor, which is capable of updating the DDS registers several thousand times per second (although such frequent updates are not expected to be needed in foreseeable ALMA modes). The local microprocessor in turn communicates with the Monitor/Control (MC) subsystem via the ALMA Monitor/Control Bus (AMB, based on CAN). Parameters specifying the desired phase vs. time function are sent over the AMB as needed, but no more often than every 50 msec. Precise timing is achieved via a 20 Hz pulse which is distributed throughout the array and is available to each DDS assembly; commands over the AMB are effective on the next leading edge of this pulse. The pulse rise time is sufficient to allow resolution of a particular cycle of the 125 MHz DDS clock. Together, the 125 MHz and 20 Hz reference signals allow the timing of all events within the DDS assemblies to be known with a resolution of $< 8\,$nsec and an ambiguity interval of 50 msec. (If necessary, an additional reference at 25 MHz can be made available to facilitate implementation of this timing precision.)

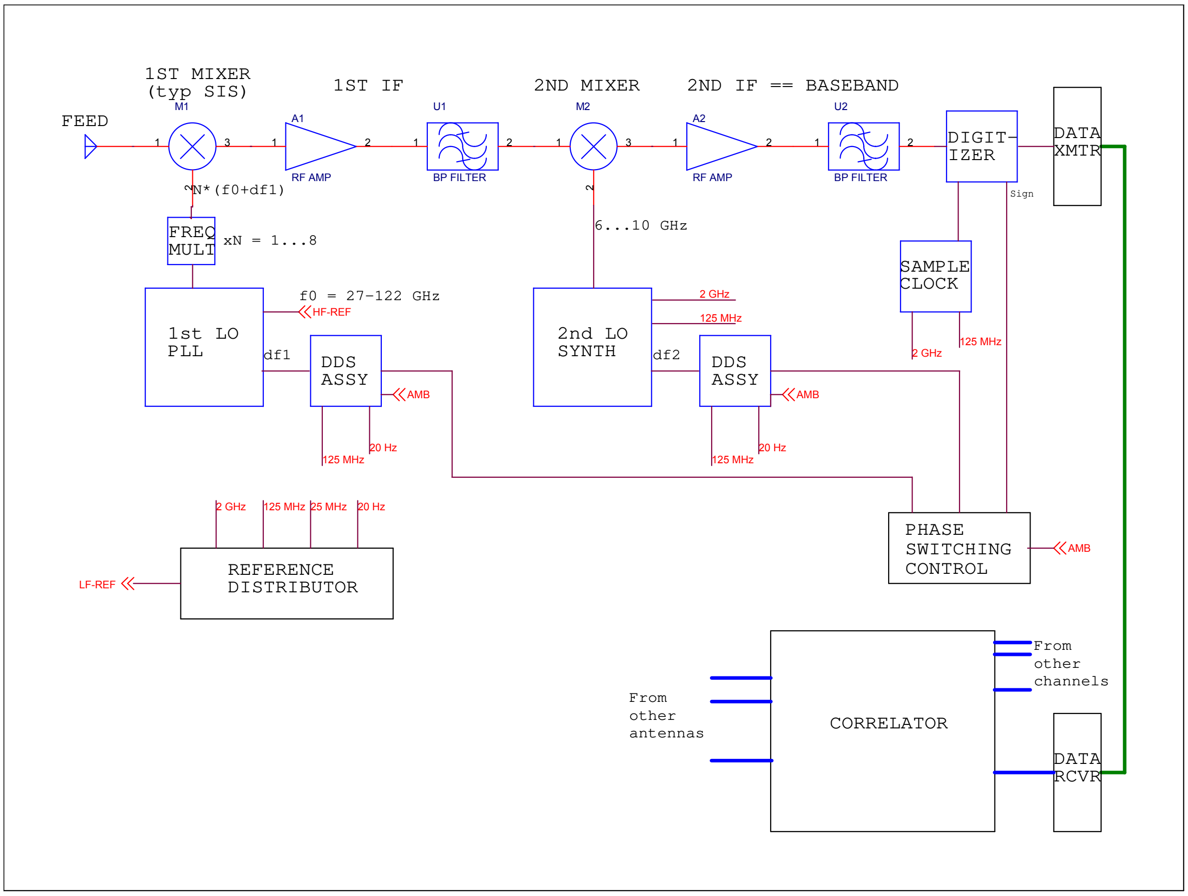

Figure 1: Simplified block diagram for one channel of one antenna.

Frequency Tuning and Phase Ambiguity Avoidance

Generally, the receiver tuning is fixed by the LO subsystem's synthesizers, with the DDSs departing from their nominal frequencies (31.25 MHz in the present design) only to the extent necessary to implement fringe rotation and other small offsets as discussed below. The synthesizer resolution is much less than the processed bandwidth, and should be sufficient for the anticipated scientific purposes. However, offsets up to about ± 20 MHz should be possible if aneed for them should arise.

Whenever the receiver tuning is changed (e.g., between receiving bands), whether or not there is a special frequency offset, it is required that the phase of all LOs be unambiguous. That is, it should be predictable and repeatable, not dependent on the previous state of the system. Specifically, if we change to a new frequency and then back to the previous one, the LOs should have the same phases that they would have had if we had remained continuously at the same frequency. All other parts of the LO synthesizers are designed to achieve this; they operate on fixed references without any frequency dividers. But each DDS is, in effect, dividing down its clock to produce the commanded output frequency, so that the output phase depends on the initialization of its internal registers. When a frequency change is commanded, the phase depends on the previous state. To avoid this, each DDS can be commanded to re-initialize its registers at a specific 20 Hz timing event, and thus on a specific 125 MHz clock cycle. From there, the phase can be adjusted to any desired value by an appropriate series of frequency register updates.

To achieve the phase ambiguity avoidance just described, it is important that the interface from the MC to LO describe the desired *phase* vs. time, not just the desired *frequency*. If the tracking of a source is interrupted by a switch to a different source (and hence a different fringe frequency), a switch to a different observing frequency, or some other significant mode change, then upon return to the previous setup the MC system should calculate and command the same LO phase that it would have commanded at that time if there had been no interruption. The DDS control firmware (in the local microprocessor) must be capable of correctly implementing such commands. (See Appendix A for further discussion of how this can be done.)

Fringe Tracking

The principle of fringe tracking is to delay all LO signals by the same amount as the RF signal is delayed on its path to the antenna, with respect to arrival of the same RF signal at a reference location (such as at the center of the array). But a delay in a sinusoidal LO signal is the same as a phase shift, and since the delay changes as the earth rotates so does this phase shift. The slope of the phase with respect to time is a frequency offset, known as the fringe frequency. (For a more thorough discussion, see [2].)

The fringe frequency is given by $f_L \, dD(t)/dt$, where f_L is the total LO frequency and $D(t)$ is the RF delay. The maximum fringe frequency corresponds to the maximum rate of delay; for sidereal source tracking, this occurs when the source is on the equator and meridian and the antenna is at the maximum east-west displacement from the reference position. Taking f_L =950 GHz and an antenna displacement of 5 km then gives a maximum fringe frequency of 1.15 kHz.

In the ALMA system, there are two stages of frequency conversion. The fringe tracking phase function can be applied to either of these, or part of it can be applied to each. However, in computing the total LO frequency, each separate LO frequency must be taken as positive for LSB conversion and negative for USB conversion, so the result is different for each sideband. The present ALMA design assumes suppression of one sideband by the first mixer, and contemplates processing only the other sideband. In that case, for simplicity, we can do all fringe tracking in one LO with no loss of functionality or performance. Whereas some modes use the same first IF to drive several baseband channels, each with a separate second LO, it is the second LO that must handle the fringe tracking. On the other hand, to support the possibility that some front ends may have double sideband mixers and to allow for processing both sidebands simultaneously (with separation after correlation), we should also make it possible to do fringe tracking in the first LO. By applying the appropriate phase shift to each LO separately, the equivalent delay is correctly applied to each and this is independent of sideband. It is easy and convenient to implement the DDS assemblies (including firmware) identically for the first and second LOs, allowing complete flexibility and relegating control of the fringe tracking method to MC software.

A DDS with a 32b frequency register and 125 MHz clock has a frequency resolution of .0291 Hz. An error of 1 least-significant bit in frequency integrates to a phase error of .01 radian (0.6d) in 55 msec. Therefore an updating period for the DDS frequency of 10 to 100 msec will be adequate, provided that the frequency is calculated to <1 bit. It is also necessary that the updates occur at precisely controlled times. The values written to the DDS frequency register should include corrections to prevent long-term phase drift due to accumulation of roundoff errors, and to cause the DDS phase register to follow the desired phase-vs.-time function. (See Appendix A for further details.) All of this is the internal responsibility of the DDS assembly.

At the interface to MC, data that convey the desired phase-vs.-time function to sufficient accuracy are required. There are many ways to do this; the details will not be specified here, but some design considerations follow. In general, a (small) set of numerical parameters will be delivered periodically; from these, the DDS assembly will calculate the phase function over the next period, using a fixed formula. Parameterizations that prevent discontinuities in the function and at least two derivatives are probably best. For any simple parameterization, it is easy to calculate the update period that will achieve a given accuracy for sidereal source tracking; but it is not so easy to do this other modes, such as on-the-fly (OTF) imaging. There is a tradeoff between the number of parameters (and hence the complexity of the evaluation formula) and the update period. There is not much point in making the update period longer than an elementary observation interval (where the telescope stays in a stable state and does not switch sources nor change scanning direction nor make any other change that would require switching to a new phase function). The interface must use the 20 Hz timing events for synchronization, so the minimum update period is 50 msec. It can be concluded that an update period between about 100 msec and 2000 msec is appropriate, with the actual value TBD.

Sideband Suppression

A strong astronomical signal in the undesired sideband of the first mixer may cause significant confusion in the scientific interpretation of the data. If fringe tracking is implemented in only one LO and programmed for the desired sideband, then the undesired sideband is incorrectly tracked. Upon cross-correlation, the visibility phase for a signal in the undesired sideband will vary at the residual fringe frequency. If we integrate for many cycles of this frequency, the undesired response is strongly suppressed. However, there are always places in the sky where the fringe frequency is zero so that no suppression occurs, or where the residual fringe frequency is too low to produce significant suppression in a suitable integrating time. With a large array, this happens on some baseline quite often. As mentioned above, fringe tracking can also be implemented in both LOs in such a way that both sidebands are correctly tracked, in which case suppression by this mechanism never occurs.

A related method allows suppression of the undesired sideband regardless of fringe frequency [5]. Let the frequencies of the first and second LOs both be offset from nominal by the same amount, either with the same sign or with opposite signs. Then at the second IF all signal frequencies originating in one sideband are shifted from nominal by twice the offset, and those originating in the other sideband are unshifted. Now let the offset frequencies at all antennas be different, such that at the kth antenna the offset is kf_0, $k = 0, ..., N-1$. Upon cross-correlation, the visibility phase of a signal in the undesired sideband varies at the difference of the offsets, producing zero on all baselines after an integral number of cycles of f_0. To support a minimum integrating time of 16.0 msec (as in the current correlator design), the minimum value of f_0 is 62.5 Hz, and the maximum offset is 3937.5 Hz for $N = 64$ antennas.

The sideband suppression is not effective in single dish observations, including cross-correlation of oppositely polarized channels of the same antenna. This is because the synthesizers for both first and second LOs are common to the two polarizations; we cannot provide different offsets for oppositely polarized channels. However, it is possible to perform sideband identification by observing twice, with and without offsets; the baseband spectrum due to one sideband will remain fixed, and that from the other will shift by twice the offset.

Sign Switching, or Phase Switching By 180 Degrees

Suppose that at some point in the signal path we multiply the signal by waveform that is -1 for half the time and $+1$ for the other half, in a periodic pattern, and at some other point later in the signal path we again multiply it by the same waveform, where the two instances of the waveform are exactly in phase. The net effect on the desired signal is then null. However, if some other signal (undesired) is added to the original signal at a point anywhere between the two multipliers, then the extra signal retains the sign modulation. In this way, the desired signal can be distinguished from certain undesired signals. Included in the latter are spurious couplings, low-level oscillations in amplifiers, and d.c. offsets.

By inserting the sign modulation as early as possible and removing it as late as possible, the maximum opportunity to separate undesired signals is taken. One way to introduce a sign change is by an LO phase shift of 180d. This can be done in the first LO. Once the signal is digitized, a sign change can be introduced by bit inversion. This can be done immediately after digitization, since addition of undesired signals in the subsequent digital processing is far less likely than in the preceding analog processing. This is the implementation intended for ALMA: apply sign modulation via the first LO phase, and apply demodulation within the digitizer, all within the antenna so that the two waveforms can be accurately synchronized.

Implementation of the 180d phase reversals in the LO should use the phase-offset feature of typical DDS chips, rather than attempting to alter the phase of the basic numerical oscillator or to effect the phase change in separate digital or analog circuitry.

The sign modulation waveform must be different at each antenna and all must be mutually orthogonal. This ensures that, when the signals from different antennas are cross-correlated, undesired signals that are coherent among antennas will have zero net correlation in any integration time that contains an integral number of complete switching cycles. To allow for cross-correlation of oppositely polarized channels of the same antenna, we should also use different and orthogonal waveforms for those channels; but this is not possible in the present ALMA design since the LO synthesizers are common between polarization channels. Thus, for a 64 antenna array, we need a total of 64 mutually orthogonal binary sequences. For a given number of sequences, the shortest mutually orthogonal set is the set of Walsh functions [3,4]; N orthogonal functions requires each to have length N times the smallest interval between transitions, provided that N is a power of two. Another is the set of Rademacher functions, which are square waves whose frequency ratios are powers of two; N of them requires a length of $2^{N/2+1}$. So for our case (N=64) we need Walsh functions of length 64 or Rademacher functions of length 8.6×10^9 (8570 sec with 1 μsec steps).

If Rademacher functions could be used, they would have the advantage of retaining orthogonality when one is time shifted, in which case we could avoid the need for accurate synchronization of the switching times among the antennas. Walsh functions do not have this property, so we must provide for this synchronization. The switching times must be synchronized when the signals are cross correlated, not when the sign changes are applied at the antenna. In between, they suffer different amounts of delay, and these delays vary with time as a source is tracked. Neglecting common-mode delay, the delay for each antenna is exactly the geometric path delay of the signal relative to the reference location, which is easily calculated. This is at most $(10\text{km})/c = 33.3\,\mu$sec if the maximum east-west extent is 10 km. The switching times at the antenna should be offset from some nominal times by the negative of the antenna's delay.

For some observing modes (OTF mapping, especially), short correlator integrations are desired, and a full switching cycle should be completed in one such integration. With this in mind, the current correlator design supports a minimum integration time of 16 msec. To complete a length-64 Walsh function cycle in this time requires a switching interval of 250 μsec (4 kHz rate). The circuit of Figure 2 shows one way to implement a switch-signal generator, including the ability to set the offset time with 40 nsec resolution.

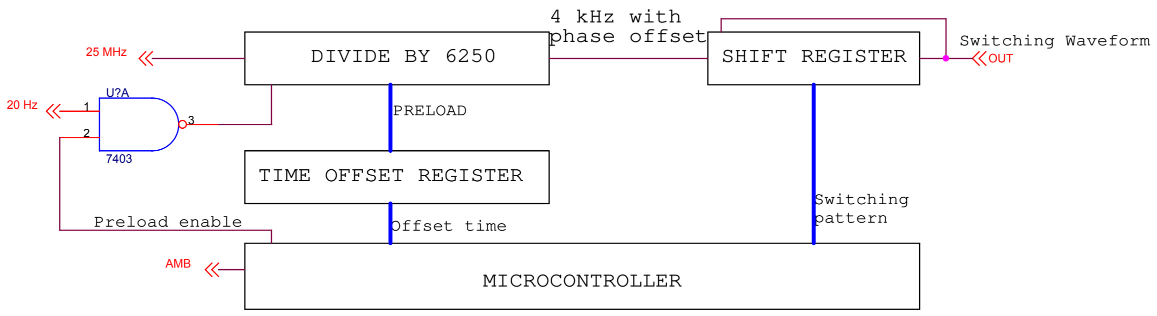

Figure 2: Sample design for switching waveform generator.

It would be helpful to have a set of switching functions which maintain orthogonality after an arbitrary time shift (unlike Walsh functions) and which are of reasonably short duration (unlike Rademaker functions). This is true of the set of square waves whose frequencies are 0, 1, 2, 3, ..., $N-1$. They are mutually orthogonal over the unit interval in spite of any time offsets. But these functions, unlike Walsh and Rademaker functions, are very difficult to synthesize exactly by traditional digital means from a common clock. To do so, the clock frequency would have to be the twice the product of the union of the prime factors of 1, ..., $N-1$; for $N = 64$, this is about 1.2×10^{27} (cf. 1.6×10^{10} for Rademaker functions). On the other hand, suppose that we do not insist on exact synthesis. Since the shortest period among the functions is $1/(N-1)$, allowing an error in orthogonality of ϵ permits an error in transition times of $\epsilon/(N-1)$, which can be achieved with a clock frequency of $(N-1)/\epsilon$. Taking $\epsilon = 10^{-4}$ (40 dB) and $N = 64$, we get a clock frequency of 640 kHz over an interval of 1 sec or 64 MHz over 10 msec.

Phase Switching By 90 Degrees

General

In addition to the 180d phase switching considered above, it can be useful to impose a phase change of 90d on the first LO for half of the time. We do not intend to include this feature in ALMA, but it is discussed here for completeness and to cover the possibility that it may be added in the future. When both 90d and 180d switching are implemented, the LO has a 4-state switching cycle (0d, 90d, 180d, 270d) with equal time in each state. Nevertheless, we begin by discussing the 90d switching separately, as a 2-state cycle.

Sideband Separation

One feature that would be added to the telescope by 90d phase switching is sideband separation. If the first mixer operates in double-sideband mode (so that it is nearly equally sensitive to signals in either sideband), then it is possible to observe in the two sidebands separately and simultaneously without additional loss of sensitivity. To do this, 90d phase switching is imposed on the first LO of all antennas using orthogonal switching waveforms (e.g., Walsh functions). During cross-correlation, two separate integration register banks are provided, and the correlator switches between them in synchronism with the phase switching. For each baseline, we integrate into one bank if the product of the two switching waveforms is positive and into the other if negative. After one or more complete switching cycles, the result will be that one bank contains the sum of the visibilities due to signals in the two sidebands, and the other contains the (phase shifted) difference. In the next stage of processing, these can be added and subtracted to provide the visibilities for each sideband separately. For additional discussion, see [2].

Gibbs Phenomenon

Another feature is obtained with exactly the same setup, namely regularization of the Gibbs phenomenon ringing [6,7]. It is fundamental to a lag-type cross correlator that the complex cross power spectrum is subject to ringing at the band edges due to truncation of the cross correlation function before the Fourier transform. Difficulties arise because the amount of error depends non-linearly on the signal, being worse for the imaginary part than the real part of the cross power spectrum. When a 90d phase shift is introduced in one antenna, the real and imaginary parts are interchanged. If the shifted and unshifted measurements are then averaged, the dependence on signal phase is removed and the error has the same form for all measurements; the problem is reduced to a variation of interferometer gain with baseband frequency, and this can be calibrated out by the usual astronomical methods.

Implementation Issues

To implement the LO phase switching, we can again use the phase-offset feature of most DDS chips, which typically provides for at least 4 offset values equally spaced in phase. The correlator must provide the bin-switching feature, which is not otherwise required. In the present correlator design, the shortest interval between switches is 16 msec, so that the minimum time for a complete length-64 Walsh function is 1.024 sec. The correlator and LO switching must be synchronized. For 180d switching only, the correlator integrating cycle can be out of phase with the phase switching, but the antennas must still be synchronized with each other at the point where the signals are combined in the correlator; we discussed earlier a mechanism to achieve this. The same mechanism (Figure 2) can be used to synchronize the antennas to the correlator, except that now we must know the total transmission delay from antennas to correlator, including both the common and differential components, rather than just the differential part. (The differential part is easily calculated from the interferometer geometry alone, but the common part must be measured.)

The 4-state phase switching can be achieved in several ways. With two nested Walsh function cycles applied to the first LO, the total length is N^2=4096. Alternatively, 2-state 90d phase switching could be applied to the first LO using one Walsh function, and 180d phase switching could be applied to the second LO (and digitizer sign) using an orthogonal Walsh function. Then twice as many orthogonal functions are needed, but this merely doubles the total length (to 128). This method fails to suppress any spurious signals added between the first and second mixers. Another alternative is to apply 4-state switching to the first LO, but to use a shifted-m sequence (with m=4 states) of length N=64 [8] rather than nested Walsh functions. Although the second and third possibilities allow many fewer steps in the switching sequences than the first, they do not shorten the total cycle time because the limiting consideration is the minimum correlator integrating interval. With nested Walsh functions, the inner cycle can be used for 180d switching and can be completed in each 16 msec correlator integration; the outer cycle can then be completed in 64 such integrations (1.024 sec). With shifted-m sequences, the total time is also 64 integrations but the correlator must provide 4 integration register banks rather than 2.

Implications For The Direct Photonic LO Scheme

If the first LO is completely synthesized at the center, then there is no PLL at the antenna and no opportunity to make fine adjustments there to the frequency and phase by using a DDS. Instead, the central synthesizer will include a separate DDS-supplied offset for each antenna. The main implication of this for all of the features considered in this report is that the propagation delay of the LO signal from the center to each antenna must be taken into account.

The photonic LO transmission will include an automatic servo that maintains the propagation delay constant to very high precision, and it will be necessary to make this work across large (but rapid) changes in observing frequency. However, the delay cannot be assumed constant across power outages, certain maintenance activities, or periods of disuse. Furthermore, the servo does not produce measurements of the total delay; it only works to keep the delay constant.

The maximum delay, assuming 25 km of fiber, is about 120 μsec. For fringe rotation at the maximum sidereal fringe rate of 1.15 kHz, ignoring the propagation delay produces a phase error of 0.14 cycle. This is quite a lot, but it is slowly varying and might be tolerated with frequent astronomical calibration. If the delay is known to 1 μsec accuracy and it is taken into account in commanding the central DDS, this error is reduced to .0012 cycle. For non-sidereal sources or for OTF scanning, fringe frequencies can be much higher with proportionally larger errors.

Synchronization of phase switching is made more difficult when the LO and digitizer are separated. Synchronization among antennas was discussed earlier, but now we are concerned about synchronization within one antenna's devices. An error in cross-antenna synchronization reduces the amount by which undesired signals are suppressed, but it has no effect on the desired signal. An error in synchronization within one antenna, between the phase modulation and de-modulation of one signal path, affects the desired signal as well. With 250 μsec switching intervals (for length-64 switching within 16 msec), an error of 1 μsec produces a loss of 0.25% in the worst case. This problem is avoided if phase switching is done only at the second LO, but this fails to reject any undesired signals that are introduced between the first and second mixers.

These considerations imply that we must maintain knowledge of the total delay on the transmission path of each antenna's first LO signal to $< 1\,\mu$sec, and we must account for these delays in calculating both the fringe phase function and the phase switching offset time. It is suggested that we set $0.1\,\mu$sec delay knowledge as a design goal, since fiber length stability 50–500 times better than this can be expected on time scales of a week or more. We will need a switching waveform generator at both the center (for the 1st LO) and at the antenna, and these must be each be programmed with different offset times. One way to adjust them would be to insert similar test tones into the IFs of two antennas, disable phase switching in one antenna, and then trim the phase switch timing on the other antenna for minimum cross correlation.

References

[1] Analog Devices, 1999, Model AD9850 data sheet, Rev E, or similar publications of other manufacturers. `<http://www.analog.com/pdf/AD9850_e.pdf>`

[2] A. Thompson, J. Moran, G. Swenson, 1986, *Interferometry and Synthesis in Radio Astronomy,* New York: Wiley. Sec 6.1, "Frequency conversion, fringe rotation, and complex correlators," pp 143–155.

[3] Thompson *et al.*, ibid., Sec 7.5, "Phase switching," pp 202–206.

[4] H. F. Harmuth, 1972, *Transmission of Information by Orthogonal Functions,* 2nd ed., Berlin: Springer-Verlag.

[5] The method was suggested by Barry Clark in discussions during early planning for the ALMA telescope, then called the Millimeter Array.

[6] R. Bracewell, 1965, *The Fourier Transform and Its Applications,* New York: McGraw-Hill, pp 209–211.

[7] L. D'Addario, 1989, "Cross correlators," Lecture 4 in *Synthesis Imaging In Radio Astronomy,* Astron. Soc. Pac. Conf. Ser., vol 6.

[8] E. Keto, 1999, "Shifted m-sequences for phase switching." Submillimeter Array Project Tech. Memo No. 134, Smithsonian Astrophysical Observatory, 1999-Mar-24.
`<http://sma2.harvard.edu/private/memos>`

Appendix A: DDS Phase Control Methods

Some detailed design considerations for the DDS assemblies are given here.

DDS chips generally provide write-only access to the frequency register (or delta-phase register), and they provide no external access at all to the phase register of the underlying numerically controlled oscillator. Nevertheless, some of them provide a reset input that will cause all internal registers to be cleared (including the phase register) on a known clock transition. We need to choose a chip that has this feature. Just after a reset, the phase register value is known, and the controlling microprocessor can know the exact content of the phase register at any desired time after that, in spite of the fact that the register is not readable. This is because the microprocessor knows the exact value of every number that it writes to the frequency register, and the exact clock cycle on which that number started to be used. Using only integer arithmetic, it can calculate the phase register content at any future time:

$$P = \sum_i f_i N_i \bmod 2^L$$

where f_i is the ith frequency register value written since a reset, N_i is the number of clock cycles between the effective time of f_i and f_{i+1}, and L is the length of the phase register (32b). This calculation is exact, not being subject to roundoff errors. It is convenient if $N_i \equiv N$ is constant; that is, updates should be strictly periodic. Typically a new frequency update becomes effective a fixed number of clock cycles after the chip receives a strobe signal. The strobe signal should be generated in hardware so that it is synchronous with the DDS clock and so that N is exactly determined.

Recall that the interface from MC specifies the desired phase vs. time. It does so parametrically, requiring the microprocessor to evaluate a formula in order to know the desired value at any specific time. Then the microprocessor knows both the desired phase and the actual phase (phase register contents). It should proceed to set the DDS frequency register in such a way that the actual phase will equal the desired phase at some point in the near future, such as at the next scheduled frequency update. Note that this is quite different from the more obvious procedure in which the microprocessor differentiates the desired phase to obtain the desired frequency, and then writes this frequency to the DDS. That procedure would not allow the phase to be unambiguous after a mode change (see discussion under "Frequency Tuning" above). Furthermore, the suggested procedure prevents the long term phase drift that would otherwise occur due to roundoff errors in the computations and the fact that the frequency resolution is non-zero. (If the phase were initialized correctly and thereafter only the frequency was tracked, roundoff errors would tend to cause a random walk away from the desired phase.)

Another design technique that is very helpful (though not essential) in achieving precise timing is to drive the microprocessor from a system-synchronous clock rather than from a free-running crystal. In this way, single task or interrupt-driven code can know the exact time at which an I/O signal is generated or received. The processor can also use internal timers to generate system-synchronous events at arbitrary times, when hardware timing signals may not be readily available.