Submillimeter Array Technical Memorandum

Number: 56
Date: December 19, 1991
From: Colin Masson

Fringe rotation, Delay stepping, and Phase switching

Summary

In the operation of the interferometer, many parts of the system must cooperate to ensure that the signals are synchronized at the correlator, and that the phase variations of the interference fringes are compensated. In addition, regular patterns of phase-switching must be imposed on the signals and subsequently synchronously detected in order to separate the sidebands and to suppress crosstalk and offsets in the system. These requirements all overlap with one another so the whole pattern of operation must be considered at once. The operation of this system has effects on the minimum allowable sample time for the SMA, as well as driving a number of technical requirements. In this memo, I first discuss the various operations which the hardware must perform and subsequently relate these to the specifications of the system. I describe two possible arrangements of phase switching which appear to meet the requirements for the SMA. The first has nested switching cycles, while the second has interleaved cycles. The interleaved scheme appears to be preferable, although it entails a slightly faster dump rate for the correlator. The choices of scheme and switching speed seem reasonable, but they may need to be adjusted later after more detailed analyses of the computing load.

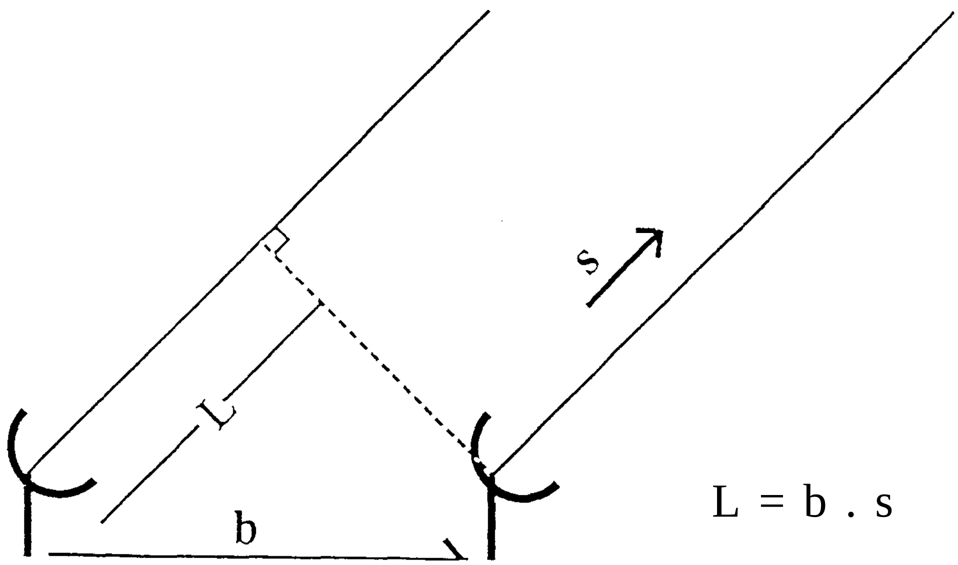

Figure 1. Geometry of an interferometer.

I. Interferometer Basics

The basic geometry of an interferometer observing a radio source is shown in Figure 1. The path difference, L, between two elements of the interferometer is given by the dot product of the baseline vector, b, and the unit vector, s, in the direction of the source.

$$L = b \cdot s$$

As the interferometer tracks the source across the sky, the path difference between the elements changes. For a pure E-W baseline, the path difference is given by

$$L = |b| \sin H,$$

where H is the hour angle of the source in radians, and the rate of change of path difference, dL/dt, is given by

$$\frac{dL}{dt} = |b| \cos H \frac{dH}{dt}.$$

For more general baselines, $|b|$ is replaced by the component of the baseline parallel to the earth's equator, and H is replaced by H- θ, where θ is the angle between the equatorial component of b, and the E-W direction. On any baseline of the interferometer, the path differences may vary over a range from $-|b|$ to $+|b|$, and the rate of change of path varies between $\pm |b|$ dH/dt. Note that there are times when the rate of change can be zero.

To achieve maximum correlation between broadband signals, there must be a delay line in the system, which compensates for the varying geometric delay $\tau = L/c$, where c is the speed of light, as well as any static delay due to unequal cable lengths. The static delay can be minimized by a careful design of the cabling system, and will not be discussed further in this memo, as any small offsets can be handled by small adjustments to the variable delay lines. The most natural way to organize compensation for the geometric delay is to pick the center of the array as a reference point and then to provide each antenna with a variable delay line of maximum length $\tau_{max} = |b_{max}|/c$. When the array is observing a source directly overhead, all the delay lines are set to $\tau_{max}/2$. As the source is tracked across the sky, the delay lines are all varied to maintain a constant geometric delay relative to the center of the array. In this way, the delays are always correct for baselines between all pairs of antennas. The lengths of the delay lines must be large enough to cope with the longest baselines we envisage or, at least, there must be provision for extending them.

In addition to providing the correct delay, a correction must be made for the geometric phase, or fringe phase $\varnothing = 2\pi f \tau$, where f is the frequency of the incoming signal. In a very simple interferometer with no frequency conversion between the antennas and the correlator, this would automatically be taken care of by the delay line, since the ratio between phase and delay in free space is the same as that in the delay line. This cancellation would also be exact over an arbitrarily large range of input frequencies f. In our case, the delay will take place at a very low frequency in the IF, where there is only a small phase change for each increment of delay. Therefore an

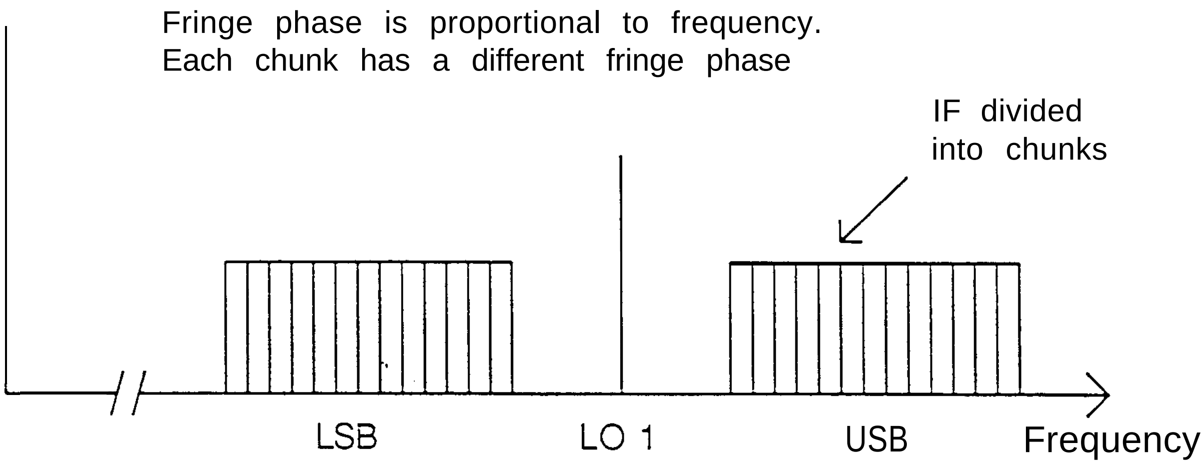

Figure 2. Frequency ranges in the two sidebands of the interferometer.

additional phase must be introduced into the local oscillators in the system to cancel the uncorrected part of the fringe phase. This is discussed in more detail below. The fringe rate, F, is given by

$$F = \frac{1}{2\pi}\frac{d\varnothing}{dt} = f\,\frac{d\tau}{dt}$$

and the maximum fringe rate F_{max} is given by

$$F_{max} = \frac{1}{2\pi}\left(\frac{d\varnothing}{dt}\right)_{max} = f_{max}\ \tau_{max}\frac{dH}{dt}$$

where f_{max} is the maximum operating frequency of the interferometer.

The range of input frequencies to the interferometer is sketched in Figure 2. In general, with double-sideband receivers, two different fringe rates will be present simultaneously, and the system must be capable of dealing with this. Furthermore, the fringe rate varies across the observing band. A further complication, in principle, is that sources in different parts of the primary beam have different fringe rates. Even when the fringes from a source at the beam center are compensated exactly, there are small residual rates which depend on the distribution of sources in the beam. These small residual rates are calculated below.

In principle, the fringe rates could be removed in software after sampling, but relatively fast sampling would be required. Also, when the fringe rates became zero, there would be no rejection of drifts and cross talk, and no possibility of separating sidebands. In practice it is more convenient to adjust the phases of the local oscillators to cancel the fringe phases and then to substitute a fixed pattern of phase shifts, which can be demodulated synchronously in the correlator. The adjustment of the LO phases is called phase-rotation or lobe rotation, and the fixed pattern of phase shifts is called phase switching. The phase switching serves two major purposes : 1) to cancel the effects of drifts and crosstalk in the system, and 2) to measure the upper and lower

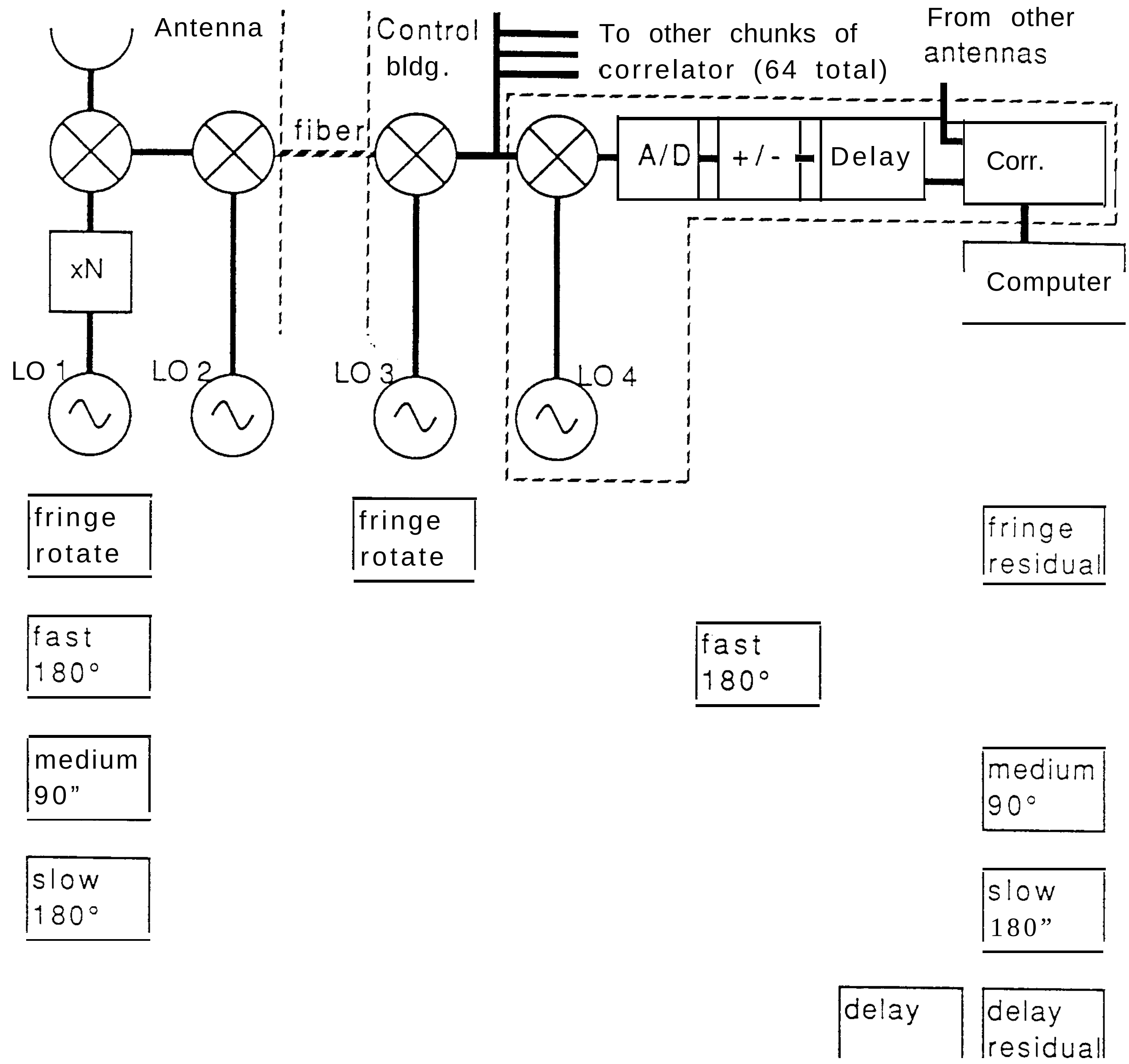

Figure 3. Frequency conversion scheme of the SMA, showing the locations of phase and delay switching elements. Only one receiver is shown. The other is similar.

sideband signals independently. Just as in the case of delay lines, the simplest scheme for phase rotation is to pick a reference position at the center of the array and correct the phase of each antenna with respect to that position. The reference position does not necessarily have to be the same as that used for the delay, although reference positions far from the center of the array will produce larger maximum fringe rates.

II. Application to the SMA.

Figure 3 is a diagram showing the layout and frequency conversion scheme in one receiver band of one antenna of the SMA. The other antennas and receiver bands are identical. Below the diagram are labels indicating at which points various fringe rotating and phase-switching operations are performed. The first local oscillator in general consists of an oscillator and a multiplier. In the text, it is referrred to as if it is a single oscillator. In practice, however, any phase changes must be divided by the multiplication factor N before being applied to the oscillator. Two frequency conversions are shown before the correlator, although we may use only one in practice. The cot-relator is divided into a large number of chunks, each of which covers only a small bandwidth. The dashed box shows the contents of one chunk, and is repeated 64 times per receiver.

Phase Rotation

A mathematical description of the propagation of phases through the system is presented in the Appendix. In the present text, I give a verbal description to show how we arrive at the several phase functions at the various oscillators. The phase rotation is required to correct for the geometric phase of the signal, since the delay line operates at a much lower frequency than that of the received signal. As noted above, since the phase is proportional to frequency, the necessary phase differs between the upper and lower sidebands, and also varies across the finite bandwidths. Since the delay line automatically provides exactly the correct phase shift across the band for both sidebands at the final IF where it operates, the necessary phase correction at each oscillator is that phase which would be required for a signal at the frequency of the local oscillator. There is a significant difference between the first and subsequent local oscillators, since the sidebands are unmixed before the first conversion and mixed afterwards. Thus the phase rotations required for any number of local oscillators after the first can be performed at only one of the oscillators, but the first LO must be rotated separately.

After the first local oscillator has been corrected for the appropriate phase, the upper and lower sidebands have equal and opposite residual fringe rates. Phase rotation of the subsequent oscillators will change the fringe rates, but they will still be equal and opposite for the two sidebands. For convenience we will probably choose to phase rotate L03. LO4 will be difficult to rotate because there are many oscillators, one for each chunk of the IF band.

To decide what phase to apply to LO3, consider first a single 32 MHz chunk of the correlator. The delay line operates on the baseband signal (O-32 MHz). Thus, if the phase rotation of LO3 sets the fringe phase to 0” at the DC end of the chunk, the delay line will cancel out any residual fringe phase across the band. However, the fringe phase will not be zero in any of the other chunks of the spectrum, since the initial fringe phase varies from chunk to chunk, while the applied correction is the same for all chunks, since it is applied at LOs 1 and 3, which are common to all chunks. Considering the whole 2 GHz IF band, therefore, we can set the fringe phase to zero at some point in the band, but there is a residual fringe phase which varies linearly across the band. After passing through the delay lines which correct for the local phase slope across each chunk, the residual fringe phase has a staircase pattern across the band, as sketched in Fig. 4. With a whole array of phase rotators for the oscillators L04, this staircase pattern could be removed, but it is more cost-effective to make a correction in software.

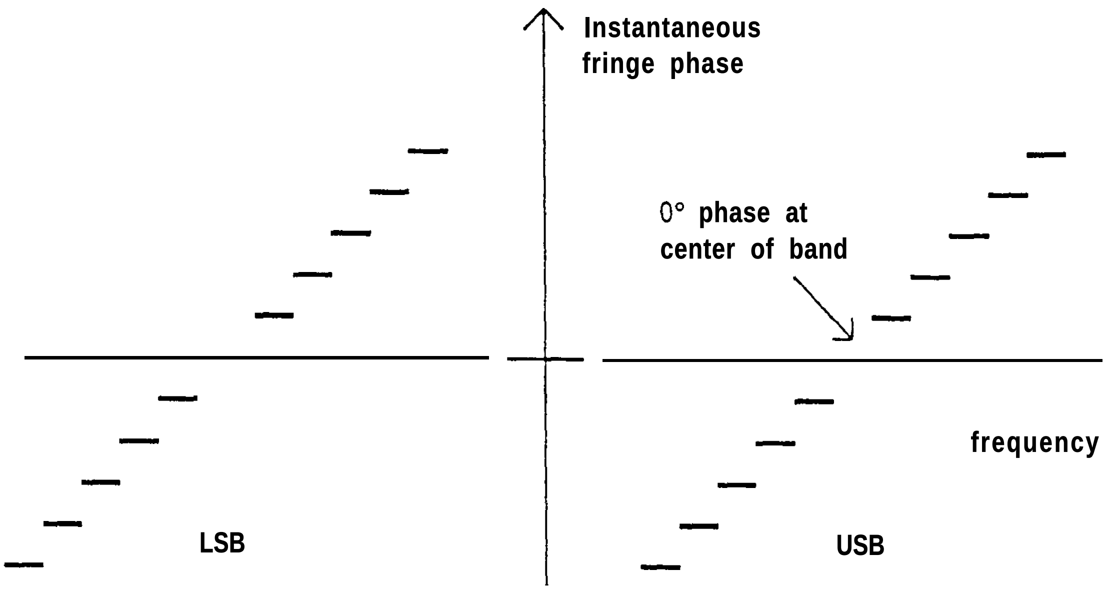

Figure 4. Sketch of fringe phase across the IF bands, when the phase rotators are adjusted to set the phase to 0" at the center of the band. There is a linear slope across the band, although the phase is flat across each chunk, due to the delay lines.

To make such a correction in software without losing sensitivity, it is necessary to sample the correlated signal sufficiently fast that the fringe phase changes by a negligible amount between samples. For 0.4 % loss, boxcar integrations should last less than 1/20 of a cycle. The lowest rate is given by setting the fringe phase to zero in the center of the whole IF band. The maximum residual fringe rate, R_{max}, is at the edge of the band and is given by the formula

$$R_{max} = \frac{\Delta f_I}{2} \tau_{max} \frac{dH}{dt}$$

where Δf_I is the IF bandwidth. The corresponding maximum sample time, t_{fringe}, is given by

$$t_{fringe} = \frac{1}{20} \frac{2}{\Delta f_I} \frac{1}{\tau_{max}} \frac{dt}{dH}$$

$$= 412.5 \frac{1}{\Delta f_I \, b_{max}} \text{ sec}$$

which yields a minimum time of 412 msec for our expected parameters of 500 m maximum baseline and 2 GHz IF band. Since the phase residuals are different for the two sidebands, the slow 90° cycle which separates the two sidebands must be completed in a time shorter than t_{fringe}.

Delay Rates

The delay line is digital and is made up of memory chips. Ideally the delay would be much finer than the sampling rate, so that the delay could be adjusted perfectly at all times. Once again this is not economic and we rely on the computer to compensate for the residual delay errors. If there is a small delay error, the correlation peak does not fall exactly on the central lag, but is shifted slightly.

There is also, in general, a phase shift associated with the delay error. The delay associated with each stage of the delay line is $\Delta\tau$, which is equal to the time between A/D samples and is therefore inversely proportional to the bandwidth of each chunk of the correlator. When the delay error reaches $\Delta\tau/2$, the delay line is increased by one element and the delay becomes $-\Delta\tau/2$. Since the lag spectrum is adequately sampled, the situation can be rectified by interpolating the lag function and resampling it, centered on the peak. This could be done by an explicit convolution with a sinc-shaped convolving function in the lag domain or, more conveniently, by a multiplication of the spectrum by a phase gradient.

Just as in the compensation of the residual fringe rates, this imposes a requirement for frequent sampling so that the cross correlation function is not smeared out. The minimum sampling time, t_{delay}, is given by

$$t_{delay} = \frac{1}{10}\frac{\Delta\tau}{\tau_{max}}\frac{dt}{dH}$$

$$= \frac{1}{\Delta f_C\, \tau_{max}}\frac{dt}{dH}$$

$$= 2.063 \times 10^5 \left(\frac{1\ \mathrm{MHz}}{\Delta f_C}\right)\frac{1}{b_{max}}\ \mathrm{sec}$$

where Δf_C is the bandwidth of each chunk. This gives a limit of 12.89 sec for 32 MHz chunks and a baseline of 500 m. Comparing this with equation for t_{fringe} above, we can see that this is a much less stringent limit provided that the IF band is divided up into more than 2 chunks, which will certainly be true for any correlator architecture which we might employ.

Beam Smearing

There is one final limit to the sampling time, t_{beam}, which is set by the presence of sources in different parts of the beam, which have different fringe rates. If we use the same criterion as before, that the averaging period should be less than 1/20 of a fringe period, then the limit for sources at the edge of the beam is

$$t_{beam} = \frac{1}{10}\frac{D}{c}\frac{1}{\tau_{max}}\frac{dt}{dH}$$

$$= 8.251 \times 10^3 \frac{1}{b_{max}}\ \mathrm{sec}$$

Phase switching - fast 180”

The purpose of this fast 180° phase switching is to cancel out the effects of any zero level errors in the A/D converters, as well as any crosstalk or spurious signals injected between the point where the phase switch is injected and the point where it is demodulated. Since the design of the VLA, the standard way to perform this phase switching is with a set of orthogonal Walsh functions, one per antenna. At the cot-relator, the products of these pairs of Walsh functions are themselves Walsh functions. There is a finite period over which these Walsh functions complete an integral number of cycles and are orthogonal, in the sense that the integral of the product of any two such functions is exactly zero. The sampling of the SMA data system must then be arranged in exact multiples of this basic period.

For N antennas, with a phase-switch clock running at a maximum frequency of p Hz, the basic period is $P_2\{ N / 2p \}$ seconds, where P_2 is the next largest number which is an integral power of 2. The factor of 2 in the denominator arises because each half cycle (chip) has an independent value. Thus the basic period has $P_2\{ N \}$ chips in it and can support $P_2\{ N \}$ independent switching functions. If we choose to use different Walsh functions for the two receivers in each antenna, then the cycle length is doubled. Taking this doubling into account, in the initial array the cycle length is 8 cycles of the clock (16 chips), which would have to be increased to 16 (32 chips) if we expand to 12+2 antennas. For p = 160 Hz, the initial cycle period is 50 msec.

The switching rate, p, is constrained by a number of different factors. Since it determines the minimum integration time of the array, it is desirable to have it be as high as possible. A high rate is also best for rejecting crosstalk effects which drift in time. On the other hand, a high rate is likely to lead to signal loss if there is any misalignment between the modulation at the antennas, and the demodulation at the correlator. If we allow for a 2 us timing error (comparable with the cable delay across the array), there will be a 0.4% loss from this misalignment when p = 1000 Hz. To avoid losses, the switching must be done in a wideband lock loop. Since the final loop will have a bandwidth of about 1 MHz, a reasonable limit to p is about 1 kHz if the losses are to be kept below 1%.

The degree of rejection of spurious signals depends on the value of p. If the spurious signal is exactly an even harmonic of p, it is rejected perfectly, but an odd harmonic is only rejected by a (voltage) factor of s/p, where s is the frequency of the spurious signal. This analysis applies only to a perfect square wave and exact harmonics, but a similar analysis applies to Walsh functions in general, so we can only expect a typical rejection of the order of s/p. However, the residual signal is aliased to a frequency in the range of 0 - p, and will tend to average to zero over a sufficiently long integration time. Thus p = 1000 Hz will give about 60 dB rejection of a 1 MHz signal in each integration period, and further rejection will come from longer averaging and perhaps a slower 180 phase switching cycle. If possible, we should retain the flexibility to adjust p to avoid possible beats between spurious signals and harmonics of the switching frequency, and to adjust our sampling times.

To achieve the expected 60 dB rejection, the demodulating signal must be perfect at the level of 0.0001% in amplitude and timing. This should be achievable with digital signals, but the rejection may be limited by some other imperfection in the system. The demodulation is particularly easy to achieve in a digital system, since the sign of the digital signal can be inverted immediately after the sampler. This placing of the demodulation has the further advantage of avoiding timing errors due to the lags in the correlator.

Phase switching - medium 90”

The 90” degree phase-switching cycle is used to separate the upper and lower sidebands of the receiver. The mathematical analysis of this switching cycle is included in the Appendix. When the phase of LO1 is changed by 90°, one IF sideband is advanced by 90°, while the other is delayed by 90”. This is easier to see if you think of a frequency shift of, say, 1 Hz. Then the USB IF is decreased in frequency by 1 Hz, and the lower is increased by 1 Hz.

After correlation, the signals are dumped into two buffers, for periods when the phase is 0” and 90°, the 90” sample is shifted in phase computationally by a further 90”. The USB signal then has 0” phase shift, while the LSB has a shift of 180°. The sum of the two samples gives a pure USB signal, while the difference gives a pure LSB signal.

As for the fast 180° switching, the 90” switching is done with Walsh functions. Conceptually, the easiest way to arrange this is to nest the 180” and 90” cycles so that the full period of the fast 180° cycle is the basic period of the 90” cycle. Then, we will have a full cycle which is 8 times longer than the full period of the fast 180° cycle, or 400 msec. At the end of this period we will have a full set of USB and LSB measurements for all baselines, and this is the minimum integration time for the SMA. The fast 180° switching is demodulated entirely in hardware so it imposes no restrictions on the dumping rate of the correlator or the speed of the associated processor. Up to the end of the 90° cycle, all the correlator outputs can be handled by simple additions and subtractions in the lag domain. After the 90” cycle is complete, the Fourier transforms can be performed, followed by the phase shifts to compensate for fractional delay shifts, and the phase-shifting and summations required for the sideband separation.

Phase switching - slow 180°

It is an easy matter in software to superpose slower switching cycles if required. If it turns out to be useful, then we could construct a slower cycle using the 400 msec full period as the basic element. This need not have any impact on the sample rate of the system, since we have complete information at the end of each 400 msec.

Conclusions

The conclusion is that the SMA can be constructed with a relatively straightforward phase switching scheme. The first LO is rotated to compensate for the fringe rate at the frequency of LO1. The third LO is rotated to correct for all subsequent conversions. Since the oscillators LO4 have a range of different frequencies, this scheme does not correct for the fringe rates all across the IF band. The residual rate at the edges of the band sets the most important limit on the sampling rate. All switching cycles up to and including the 90” cycle which separates the sidebands must be completed within this time limit. There are two basic schemes for organizing the phase switching, *nested* cycles, described above, in which the fast 180° cycle completes within one chip of the 90° cycle, and *interleaved* cycles, in which both 180° and 90" cycles use Walsh functions chosen from the same basic set. The interleaved scheme has a shorter time to complete (or a slower clock rate, p), but requires the correlator to dump faster.

Nested Cycles

In the nested scheme, the phase switching is separated into a fast and a slow part to minimise the dump rate of the correlator, since the fast 180° switching can be demodulated in hardware. The correlator must be dumped after an integral number of complete fast 180" cycles. For the fastest operation of the interferometer, it should be dumped after each cycle. The correlator dump time is then used as the basic element of a slower 90" cycle which permits separation of the sidebands.

The fast 180° cycle is driven by an oscillator running at approximately 160 Hz. With one Walsh function for each receiver on each antenna, and a maximum number of antennas of 16, the 32 Walsh functions complete in 50 msec. A set of 8 Walsh functions used for the 90" phase switching will complete in a period of 50 x 8 = 400 msec. This produces negligible signal loss with a 500m baseline and a 2 GHz IF bandwidth.

The 90" cycle elements can be accumulated in the lag domain so there is little load on the cot-relator computer until the end of the 90" cycle, when all the Fourier transforms must be performed. We should aim to be able to cope with computations at this rate.

Interleaved Cycles

In the interleaved scheme, the length of the basic Walsh cycle is expanded by a factor of 2 to provide separate functions for both 90" and 180" switching. The total period of the cycle can be the same as before (400 msec), but the basic clock can be much slower, reducing the required speed for the phase synthesizer, and associated equipment. The disadvantage is that the conelator must be dumped faster than in the nested scheme, since the 90" switching occurs at the same rate as the 180°. For the initial size of the SMA, this should not be a serious problem, since the dump time can be as long as 12.5 ms, but if we expand our baselines and IF bandwidth, we may want to perform some extra phase rotation to avoid the faster dump rates required by the basic system described here.

Summary

The basic parameters of the SMA are presented in Table 1, for both schemes. For some of the parameters, two values are given, one of which is the initial value, and the second of which is the assumed limit of expansion. The derived parameters, such as maximum fringe rates, are shown below this, first as formulae showing their dependence on the basic parameters. Finally the suggested choices for switching rates are listed, with their implications for correlator dump rates and interferometer sample times.

The interleaved switching scheme appears to be the better choice for the SMA, since it minimizes the timing and speed requirements across the array, provided that the correlator can handle the faster dumping. In this scheme, we would probably choose the basic chip time of 12.5 msec as the fundamental update cycle for the system. For each 12.5 msec period we would calculate an initial phase and phase rate for the synthesizers so that all of the first LO phase rotation and switching would be controlled directly by the antenna computers, with no special switching hardware.

From the numbers given above it is clear that the minimum integration time for the SMA is set by the length of the Walsh function cycle and the phase switch frequency. At a 160 Hz phase switch rate (nested cycles) or 40 Hz (interleaved cycles), the minimum integration time is 400 msec. This may be useful for solar observations. For most observations there is no astronomical need to sample any faster than the limit set by beam smearing, or 16.5 sec. The most convenient way to handle the integrations is to set 400 msec as the basic sample time, and to add together a series of 400 msec samples to make up the total integration time.

The computational load on the correlator computer is probably the chief cost factor limiting the sample rate, although there may also be some impact in the correlator hardware involved in the dumping every 12.5 msec. The cot-relator must perform all of the Fourier transforms, followed by phase corrections, at the end of every 90” cycle. This task, and the dump rates from the correlator chips, must be analyzed before we settle on our final choice of phase-switching system.

The expansion to 12 antennas and a larger IF bandwidth can be handled by the phase switching hardware, but the correlator dump rates and computational load will go up by a large factor if we maintain this simple arrangement. Of course, computers will be cheaper by then, but it may be worthwhile to change the hardware configuration to provide at least partial compensation of the phase shift across the IF band and permit longer cycles.

Topics which must still be analyzed are:

1) Update rates and interpoiation scheme for the phase rotators
2) Timing accuracy between stations
3) Response of the phase-lock loops to the phase switching
3) Correlator processing requirements - do we need DSP chips, for example, in a special purpose processor
4) Science requirements for fast sampling
5) Possibilities of more complex architectures in the LO system to relax constraints on the sampling times
6) Rejection of birdies by the phase-switching

Table I

Phase and Delay Parameters of SMA

Baseline (b)		500m	1000 m
Min Frequency (f_{min})		200 GHz	
Max Frequency (f_{max})		1000 GHz	
IF Bandwidth (Δf_I)		2 GHz	4 GHz
Number of antennas (N)		6+2	12+2
IF chunk width (Δf_c)		32 MHz	128 MHz
Number of chunks	$= \Delta f_I / \Delta f_c$	64	40
Delay step	$= 0.5 / \Delta f_c$	15.62 ns	3.906 ns
Max delay	$= b / c$	1.667 μs	3.333 μs
Max delay rate	$= 2.424 \times 10^{-13}\, b$	121.2 ps/s	242.4 ps/s
Max delay step rate	$= 4.828 \times 10^{-13}\, \Delta f_c\, b$	7.725×10^{-3}	6.180×10^{-2} steps/s
Max fringe rate	$= 2.424 \times 10^{-13}\, f_{max}\, b$	121.2 Hz	242.2 Hz
Max residual fringe rate	$= 1.212 \times 10^{-13}\, \Delta f_I\, b$	0.1212 Hz	0.2424 Hz
Dump rate limits:			
Delay smearing	$= 2.063 \times 10^{11} / (\Delta f_c\, b)$	**12.89** sec	1.611 sec
Beam smearing	= 8.251 x 103 / b	16.50 sec	8.251 sec
Fringe smearing	$= 4.125 \times 10^{11} / (\Delta f_I\, b)$	0.4125 sec	0.1031 sec
Nested Cycles			
Walsh function cycle length (fast 180”)		16	32 half clock cycles
Walsh function cycle length (medium 90°)		8	16 full fast 180° cycles
Phase switch clock frequency		160 Hz	2560 Hz
fast 180° cycle period		50 msec	6.25 msec
90” cycle period		400 msec	100 msec
correlator dump time		50 msec	6.25 msec
Fourier transform time		400 msec	100 msec
Interleaved Cycles			
Walsh function cycle length (90” and fast 180”)		**32**	64 half clock cycles
Phase switch clock frequency		40 Hz	320 Hz
Walsh cycle period		400 msec	100 msec
correlator dump time		12.5 msec	1.5625 msec
Fourier transform time		400 msec	100 msec

The signals in one arm of the interferometer are considered below as they progress through the system. Before the first mixer, the signal, S_0, is a mixture of the upper and lower sidebands, both delayed by the geometric delay τ:

$$S_0 = a_u \cos[\,\omega_u(t-\tau) + \phi_u] + a_l \cos[\,\omega_l(t-\tau) + \phi_l]\,.$$

In the first mixing, it is mixed with $\cos[\omega_1 t + \phi_1]$ and then put through a filter which selects the difference frequency which is in the IF band, and rejects the sum, which is at a very high frequency. Ignoring factors of order unity, the accepted component, St, is given by

$$S_1 = a_u \cos[\,-\omega_u\tau + (\omega_u - \omega_1)t + \phi_u - \phi_1\,] + a_l \cos[\,-\omega_l\tau + (\omega_l - \omega_1)t + \phi_l - \phi_1\,]\,.$$

Note that the lower sideband formally has a negative frequency. In this convention, the IF filter accepts a band which has two ranges, one at negative and one at positive frequencies. The signal is now mixed with one or more local oscillators, whose sum is ω_N, and the sum of whose phases is ϕ_N, to give a signal S_N. At this point there is an important subtlety. Because the IF filtering after each mixer accepts a given band of ***absolute*** frequency, the mixing products selected in the upper sideband involve the term $-\omega_N$, while the products selected in the lower sideband involve the term $+\omega_N$. I have ignored the possibility of lower sideband conversions in this part of the system, but the extension of the equations is trivial, and the sign is in any case opposite between the signals from the upper and lower sidebands of the first mixer. S_N is given by

$$\begin{aligned} S_N = \;& a_u \cos[\,-\omega_u\tau + (\omega_u - \omega_1 - \omega_N)t + \phi_u - \phi_1 - \phi_N] \;+ \\ & a_l \cos[\,-\omega_l\tau + (\omega_u - \omega_1 + \omega_N)t + \phi_u - \phi_1 + \phi_N]\,. \end{aligned}$$

The final step is to pass the signal through our delay line which has a delay of τ_D, giving a signal S_D, where

$$\begin{aligned} S_D = \;& a_u \cos[\,-\omega_u\tau - (\omega_u - \omega_1 - \omega_N)\tau_D + (\omega_u - \omega_1 - \omega_N)t + \phi_u - \phi_1 - \phi_N] \;+ \\ & a_l \cos[\,-\omega_l\tau - (\omega_u - \omega_1 + \omega_N)\tau_D + (\omega_u - \omega_1 + \omega_N)t + \phi_u - \phi_1 + \phi_N]\,. \end{aligned}$$

The other arm of the interferometer has a similar signal, except that τ_D, ϕ_1, and ϕ_N are set to zero. In general, each arm has phase and delay offsets but what counts are the differences between them, so there is no loss of generality in assuming all the offsets to be lumped into one arm. The two signals S_D from the two arms are put into the correlator, which is another mixer, after which the terms near DC selected out. There is an in-phase (real) and a quadrature (imaginary) part to the correlator output, I, and Q, where the quadrature part is formed by a separate mixer where one of the two signals has had its phase shifted by $+\pi/2$. Thus,

$$\begin{aligned} I \;=\;& a^2_u \cos[\,-\omega_u\tau - (\omega_u - \omega_1 - \omega_N)\tau_D + \phi_u - \phi_1 - \phi_N] \;+ \\ & a^2_l \cos[\,-\omega_l\tau - (\omega_u - \omega_1 + \omega_N)\tau_D + \phi_u - \phi_1 + \phi_N] \end{aligned}$$

$$Q = a^2_u \sin[-\omega_u\tau - (\omega_u - \omega_1 - \omega_N)\tau_D + \phi_u - \phi_1 - \phi_N] -$$
$$a^2_l \sin[-\omega_l\tau - (\omega_u - \omega_1 + \omega_N)\tau_D + \phi_u - \phi_1 + \phi_N] .$$

Now we consider the tracking of the delay line, which sets $\tau = -\tau_D$, and phase rotation of the first LO, which sets $\phi_1 = -\omega_1\tau$, giving

$$I = a^2_u \cos[-\omega_N\tau + \phi_u - \phi_N] +$$
$$a^2_l \cos[+\omega_N\tau + \phi_u + \phi_N]$$

$$Q = a^2_u \sin[-\omega_N\tau + \phi_u - \phi_N] -$$
$$a^2_l \sin[+\omega_N\tau + \phi_u + \phi_N] .$$

Ideally, we would then set $\phi_N = -\omega_N\tau$, leaving only the phase of the astronomical signal itself. Since the oscillators LO4 have different frequencies, we cannot do this unless we adjust each one of them individually, and there will always be some residual phase, which is different for the two sidebands. The sidebands must be separated before this phase residual changes significantly if the phase residual is to be corrected in software.

Sideband Separation

Sideband separation is straightforward mathematically. Consider the new correlator outputs I', and Q' formed by adding $+\pi/2$ to ϕ_1.

$$I' = a^2_u \sin[-\omega_N\tau + \phi_u - \phi_N] +$$
$$a^2_l \sin[+\omega_N\tau + \phi_u + \phi_N]$$

$$Q' = -a^2_u \cos[-\omega_N\tau + \phi_u - \phi_N] +$$
$$a^2_l \cos[+\omega_N\tau + \phi_u + \phi_N] .$$

The upper and lower sideband signals may now be calculated from sums and differences

$$I_U = I - Q'$$
$$= a^2_u \cos[-\omega_N\tau + \phi_u - \phi_N]$$

$$Q_U = Q + I'$$
$$= a^2_u \sin[-\omega_N\tau + \phi_u - \phi_N]$$

$$IL = I + Q'$$
$$= a^2_l \cos[+\omega_N\tau + \phi_u + \phi_N]$$

$$Q_L = -Q + I'$$
$$= a^2_l \sin[+\omega_N\tau + \phi_u + \phi_N].$$

Note that the imaginary correlator adds +x/2 in the IF, which has a different effect from the addition of $+\pi/2$ to the first LO. Sideband separation requires a combination of the two.

To make this a little more concrete, consider a series of sample periods in which the phase changes regularly between 0” and 90°, such that in periods 1, 3, 5, . . . the phase is 0” and in periods 2, 4, 6, ... it is 90”. Let the outputs of the I and Q correlators in period i be I_i and Q_i. Then,

$$I_U = I_1 - Q_2 + I_3 - Q_4 -t \ldots$$
$$Qu = Q_1 + I_2 + Q_3 + I_4 + \ldots$$

$$I_L = I_1 + Q_2 + I_3 + Q_4 + \ldots$$
$$Q_L = -Q_1 + I_2 - Q_3 + 14 + \ldots$$

In practice our digital conelator will not have an explicit I and Q channel, but the two will be formed computationally from the symmetric and antisymmetric parts of the cross-correlation function.