The Mark 4 VLBI Correlator: Architecture and Algorithms

A. R. Whitney[1], R. Cappallo[1], W. Aldrich[1], B. Anderson[2], A. Bos[3], J. Casse[4],
J. Goodman[1], S. Parsley[4], S. Pogrebenko[4], R. Schilizzi[4], D. Smythe[1]

[1]MIT Haystack Observatory, Westford, MA, USA
[2]Jodrell Bank Observatory, University of Manchester, UK
[3]ASTRON, Dwingeloo, The Netherlands
[4]Joint Institute for VLBI in Europe, Dwingeloo, The Netherlands

Abstract: The Mark 4 VLBI correlator is a station-based system designed to process data from up to 16 stations at an aggregate data rate of 1 Gbps/station. It is compatible with data recorded on Mark 3A, Mark 4, and VLBA tape-based data-acquisition systems, as well as the new disc-based Mark 5 system which also supports both real-time and quasi-real-time data-transmission over high-speed networks. The system incorporates an XF algorithm implemented in a custom-VLSI that which can be flexibly configured to cross-correlate VLBI data from one to four baselines of a single 2-bit channel at 64 Mbps/channel/station. Each chip also incorporates integral phase generators, rotators and vernier-delay management circuitry. Each of the sixteen Correlator Boards in the system contains 32 correlator chips and can process up to 8192 complex lags that can be flexibly traded off between baselines, lags and channels. Innovative algorithms allow the bandwidth occupied by the station-to-correlator model transmission to create less than a 0.1% overhead while maintaining full capability for all projected terrestrial and space-VLBI applications. The correlator control software allows up to four independent scans to be processed simultaneously to significantly improve processing efficiency.

1. Introduction

As the VLBI technique has developed and evolved since its beginnings more than 30 years ago, it has become ever more important not only to design processors which can keep up with the data flow, but also which implement new capabilities as VLBI pushes to higher-precision geodetic measurements and pursues ever more challenging VLBI science with mm-wavelengths and space-VLBI. The Mark 4 correlator is designed to help address this need by providing the capability to simultaneously process all baselines of a 16-station observation at a playback data rate of up to 1 Gbps/station. In addition, the processing algorithms have been designed to keep abreast of current and future VLBI applications including space-VLBI and the fledgling 'e-VLBI' which uses electronic transmission of raw data from station to correlator.

The Mark 4 correlator is based on the so-called 'XF' algorithm which performs a lag correlation on the raw data. This architecture allows the user to easily trade spectral resolution and the number of baselines or channels processed; for example, the number of processed lags may vary from 32 to 8192 with simple software-controlled reconfiguration. The Mark 4 correlator is 'station-based' in the sense that the data from each station are pre-delayed according to an *a priori* model prior to distribution to the baseline processing modules. A full-custom ASIC correlator chip was designed to do fringe rotation and correlation; new algorithms were devised to make this process efficient and to give extremely low model overhead.

Mark 4 VLBI correlators, with some variants in design and capability, are in operation at MIT Haystack Observatory, U.S. Naval Observatory, Max Planck Institute for Radio Astronomy

(Bonn, Germany) and Joint Institute for VLBI in Europe (Dwingeloo, The Netherlands). The correlator boards are also used in connected-element interferometers at the Westerbork radio telescope array (WRST) at Dwingeloo, The Netherlands and by the Smithsonian Astrophysical Observatory Submillimeter Array (SMA) at Mauna Kea, Hawaii. This paper will concentrate on the VLBI correlator design and algorithms.

The Mark 4 correlator design was developed by an international consortium of institutions, led in the U.S. by MIT Haystack Observatory, and in Europe by the Joint Institute for VLBI in Europe (JIVE) and the Stichting Astronomisch Onderzoek in Nederland (ASTRON).

2. Architectural Overview

The high-level architecture of the Mark 4 correlator is shown in Figure 1[1]. Sixteen channels of data recorded on a Mark IIIA, Mark 4 or VLBA VLBI recording system [Hinteregger, 1991; Napier, 1994] are reproduced on a Playback Unit (PBU). At the maximum aggregate data rate, each of these sixteen channels has a data rate of 64 Mbps and is de-multiplexed over 4 or 8 parallel tracks on the tape. These track-by-track data are transmitted to a Station Unit, which reconstructs channel data from the track-by-track data, extracts phase-calibration information from each channel, gathers sample-statistics data and applies an earth-center-relative delay to each channel according to an *a priori* model. In addition, the Station Unit may optionally apply pulsar gating functions to the data.

The Station Unit segments the data of each channel into Correlator Frames, which may vary in length from 10's of milliseconds to 0.5 seconds (correlator 'wall clock' time), and attaches a 240-bit header to each Correlator Frame containing station model data. The Station Unit then merges the sixteen channels of this formatted data into four high-speed serial transmission lines, each carrying four channels of data, for transmission to the Correlation Unit. The Correlator Frames from all Station Units are synchronized according to a common earth-centered clock distributed to all correlator elements, denoted by the signals SysClk (nominally 32 MHz) and SysTic in Figure 1.

The Correlator Unit consists of four Correlator Segments, each of which consists of an Input Board plus four Correlator Boards. Each Input Board receives 16 high-speed serial transmission lines, one from each of the 16 Station Units, where each line contains four data channels from the corresponding Station Unit . The Input Board resynchronizes and reconstructs the resulting 64 channels of data and applies them in parallel to all four Correlator Boards in the corresponding Correlator Segment so that each Correlator Board in the segment receives the same set of 64 channels. Physically, two Correlator Segments are housed in a single Correlator Crate, which also contains a common Control Board to provide timing and control signals.

As indicated in Figure 1, the correlator is logically divided into four 'slices' of Correlator Boards, where each slice contains one board from each Correlator Segment. As we will discuss later, 'slices' are a natural processing unit and allow the Mark 4 correlator to process up to four different scans simultaneously.

Each Correlator Board contains a full 64x64 crossbar array that allows selection and distribution of any combination of the 64 channels to be distributed to the 32 correlator chips in a highly flexible manner on the board. In this way, the 8192 complex lags available on a single

---

[1] The signal-distribution architecture of the VLBI correlator at Dwingeloo, Netherlands is slightly different, but all correlation processing algorithms are identical.

Correlator Board may be distributed among baselines, channels and lags according to processing requirements. For each baseline, the station model data contained in the Correlator Frame headers of the corresponding stations is sufficient for local on-board calculation of the baseline correlation parameters for each Correlator Frame. The correlation data are read from the correlator chips each Correlator Frame period, accumulated locally on the Correlator Board, and then are buffered through a local single-board crate computer[2] before transmission to a host computer for the addition of necessary bookkeeping information for storage and/or post-correlation processing.

Apart from the primary data interconnections between the Station Units and the Correlator Crates and PBU control, all other control and data paths are either 10Base-T or 100Base-T Ethernet. Mark 4 PBU control is via RS-232 and VLBA Monitor and Control Bus (MCB) [Cappallo, 1999].

3. Playback Units

The Playback Units (PBU's) used with the Mark 4 correlator are based on an enhanced VLBA tape transport [Hinteregger 1991] outfitted with two 32-track headstacks to reproduce 64 tracks at up to 18 Mbps/track (16 Mbps unformatted data rate). The PBU is compatible with tapes written with either the Mark 4 or VLBA data-acquisition system with up to 16 recorded data channels of 64 Mbps each, corresponding to 2-bit sampling at a 32 Msample/sec sample rate. The data from a single channel may be spread to as many as 8 tracks. Data recorded at less than 16 Mbps/trk may be played back up to 16 Mbps/trk provided the reproduced channel rate does not exceed 32 Msample/sec.

Each PBU is equipped with a barcode reader to automatically scan a barcode label affixed to each physical tape. These labels are used in conjunction with experiment logs to verify the identity of each tape before processing.

4. Station Unit

The 64 tracks reproduced by each PBU are transmitted to a corresponding Station Unit. The job of the Station Unit is to reconstruct the channel data and apply a model delay before transmission to the correlator. In practice, the Station Unit performs some additional functions as well. In detail, the Station Unit performs the following functions, as shown in the functional block diagram of Figure 2[3]:

1. Each Station Unit maintains a reconstructed 'center-of-earth' clock appropriate for the scan being processed. All Station Units processing a given scan have their respective 'center-of-earth' clocks synchronized by a correlator-wide reference frequency, SysClk, typically at 32 MHz, and a corresponding tick, called SysTic, with period of 32×10^6 SysClk periods. However, separate Station Units that are concurrently processing different scans maintain their own 'center-of-earth' time appropriate to the scan they are processing. The 'center-of-earth' clock rate is adjusted to correspond to the playback rate of the PBU, which may be faster than the original recording rate in octaves up to a maximum factor of 4.

---

[2] HP744 single-board VME computer running the HP-RT operating system.
[3] The physical architecture of the Station Unit is described by Schilizzi et al, 2001.

3

2. The data from each track of the PBU, which stand independently with synchronization and time information, are recovered in the Track Recovery Module (TRM). Error counts are noted and parity bits are stripped. The recovered tape-time data are compared with the center-of-earth clock plus the desired data delay to compute a position error for adjusting the tape-playback synchronization; the TRM maintains a buffer of 320 kB/track, which allows relatively loose synchronization requirements on the PBU. The data from the TRM are passed to the Channel Recovery Module, along with byte-wise 'data validity' to flag as 'invalid' any data which are known or suspected to be bad due to tape mis-synchronization or tape-playback problems.

3. The Channel Recovery Module reconstructs the original data channels by multiplexing data from the corresponding track(s) as necessary. Up to a maximum of 16 2-bit channels are simultaneously reconstructed at a maximum channel data rate of 32 Msample/sec. The output of the Channel Recovery Module consists of data sign and magnitude, plus an accompanying 'data validity' bit stream formed by combining the validity information from all tracks that contributed to the channel.

4. The reconstructed channel data are transmitted through a Crossbar Switch to the Phase-Calibration module that extracts embedded weak coherent calibration tone(s) that provide instrumental phase calibration information for each channel.

5. The reconstructed channel data following the Crossbar Switch are also transmitted to the Delay Module, which applies a delay to each channel (relative to the center-of-earth clock), then segments the data into Correlator Frames whose boundaries are coincident with sub-multiples of SysTic. A linear delay model is used over the period of each Correlator Frame. This linear model is derived from a least-squares fit to a quintic-spline delay model pre-computed for one-minute scan segments, which in turn is derived from a full-precision model generated by the NASA CALC8 geodynamic/astrometric model package [NASA/GSFC 2001]. At the instant of each 1-sample delay change, a single sample is either duplicated or deleted from the channel data stream according to the sign of the delay rate. Each of the 16 channels may be controlled by an independent delay model within the capability of the buffer limits of the Delay Module. This allows the flexibility to simultaneously process each channel emerging from the CrossBar Switch with a different pointing center. All Station Units processing a given scan divide their data into synchronous Correlator Frames.

6. The Delay Module also places a 240-bit header at the beginning of each Correlator Frame for each channel, replacing the corresponding 240 validity bits that would otherwise be present. The information included in each header includes:

   a. Correlator Frame delay model, which consists of the 'fractional-bit delay'[4] for the first sample following the CF header, plus the linear delay rate used over the CF.

   b. Station quadratic phase model for rf phase with respect to center-of-earth, derived as a least-squares quadratic approximation to a quintic phase spline, analogous to the quintic delay spline discussed above. The phase information inserted into the header represents the phase at baseband converter DC bandedge for the instant corresponding to the first sample following the CF header, while the phase-rate

---

[4] The difference between the continuous model delay and the quantized delay of the data stream.

and acceleration are for mid-band at the same instant. (see Section 7.3 for details).

    c. Oversampling factor, if any (must be same for all stations).

    d. Baseband indicator (net upper or lower sideband; must be same for both stations)

    e. Station and channel ID's, plus a frame serial number and Fletcher checksum, for verification purposes.

This information, combined with the header information from a cross-correlation partner channel from another station, is sufficient to calculate the full set of cross-correlation processing parameters needed by the Mark 4 correlator chip.

In practice, due to the pipelined structure of the Mark 4 correlator, the header data placed by the Delay Module at the beginning of Correlator Frame *n* pertains to Correlator Frame *n+1*. The reason for this will become clear when we discuss the Mark 4 'pipeline' correlation procedure.

7. The data at the output of the Delay Module are sent to a State Counter module which gathers state (sample) statistics necessary for proper cross-correlation normalization. These data are periodically transmitted to the control processor for inclusion with the actual correlation coefficients to support post-correlation processing.

8. The Pulsar Gating Module may be used, along with a pulsar timing model, to appropriately 'gate' the data by manipulating the validity bit stream accompanying each channel. Each channel may have its own pulsar timing model.

9. Finally, the 16 parallel streams of Correlator Frames are multiplexed onto four high-speed serial data links, with 4 arbitrarily-selectable channels on each link, for transmission to the correlator proper. The data for each channel includes sign, magnitude and validity information for each sample. These links utilize the Hewlett-Packard HDMP-1012/14 high-speed serial-link chip set, each link supporting up to ~400 Mbits/sec.

Due to the design of the Station Unit, the maximum length of a CF is limited to 16×10^6 SysClk cycles, which corresponds to a half-second of correlator wall clock time at the nominal SysClk frequency of 32 MHz.

In practice, the correspondence of physical Station Unit modules to those in Figure 2 is not exact. A good description of the physical unit is given elsewhere [Schilizzi, 2001].

4.1 Phase-Calibration Signal Processing

A set of weak phase-coherent tones, produced as the harmonic components of very short periodic pulses, are normally injected into the front-end RF receivers to act as continuous calibration signals for the receiver and local-oscillator chain [Whitney, 1976]. These tones, known as 'phase-calibration tones', are typically spaced at 1 MHz intervals throughout the entire RF passband. The local-oscillator frequency in each baseband converter (BBC) may be adjusted to any multiple of 10 kHz to translate this set of frequency tones to the desired position in the BBC passband. Typically, the tones are placed at 10 kHz, 1.01 MHz, etc. in each upper-sideband BBC channel and, correspondingly, at 990 kHz, 1.99 MHz, etc. in the lower-sideband BBC channel since the USB and LSB channels in each BBC share the same local oscillator. The

amplitude and phase of one or more of these phase-calibration tones may be detected in the data from each BBC to assist in the phase-coherent summing of the data across all BBC channels. The relative phase of multiple tones within a single BBC channel may further be used to estimate the 'single-band' group delay, which can help to resolve the ambiguity that sometimes arises between the single-band group delay and the multi-band group delay determined from a coherent summation of all the BBC channels.

Since the processing of phase-calibration tones is strictly a station-based procedure, this processing may be either undertaken at the observing stations or at the correlator. Since the phase-calibration tones can also act as powerful station-diagnostic signals, at least a rudimentary processing is often done at the station level. The Mark 4 processor provides full capability for multi-tone processing for each channel of each station.

The fact that the BBC-output phase-calibration tones are at fixed frequencies at multiples of 10 kHz can be used to advantage to develop a simple but flexible phase-cal-tone extraction algorithm based on simple table lookup along with a few counters [Rogers 1991; Rogers 1993]. In particular, a lookup table of a length corresponding to 3200 samples is sufficient to define the correlation coefficients for an integral number of phase-cal tone periods for any multiple of 10 kHz.

Figure 3 shows a schematic of the phase-cal extraction algorithm, which consists of a lookup table and nine counters, all initially set to 0. Starting at a sample taken on a station second-tick, each sample number and value are used as an address to retrieve a single byte from the preset lookup table. For every '1' in this retrieved byte, as shown in the example of Figure 3, the associated counter is incremented by one; in addition, a sample counter is incremented for each valid sample processed. After 3200 sample periods, the lookup table is recycled for each subsequent 3200 sample periods. At the end of a phase-cal integration period, typically about ten seconds, the counters are read and cleared and the processed is restarted.

To understand how this process is used to extract phase-cal, consider the case of 2-bit sample data correlated with a reduced 4-level sine/cosine rotation model. Table 1 shows how bits 0-1 of the lookup-table bytes are initialized, starting at the beginning of the table, for the first period of the phase-cal tone corresponding to the 'sin' part of the quadrature processing; for example, the entries corresponding to the phase-range 30-60 degrees are set to 3 for sample value '11', 2 for sample value '10', 1 for sample value '01' and 0 for sample value '00'. The remainder of the 3200 entries are copies of those for the first rotation period. At each sample, the table value corresponding to sample number and value is used to increment the counter corresponding to each '1' in the retrieved byte. At the end of the integration period, the normalized 'sin' part of the phase-cal signal correlation is computed as

$$r_{\sin} = \frac{1.96 \cdot Ctr1 + 0.98 \cdot Ctr0 - 1.47 \cdot N}{N}$$

where N is the sample count. The normalization coefficients were derived by numerically integrating a gaussian probability distribution, taking into account the product table. The harmonic content of the rotator in this case for the 3rd, 5th and 7th harmonics is –13 dB, -21 dB and –26 dB, respectively. Similarly, bits 2-3 of the lookup table are identical to bits 0-1, except they are shifted to the left the equivalent of 90 degrees to represent the cosine. Bits 4-7 may be set to simultaneously extract another phase-cal frequency.

Sample value	Rotator Phase (deg)											
	0 to 30	30 to 60	60 to 90	90 to 120	120 to 150	150 to 180	180 to 210	210 to 240	240 to 270	270 to 300	300 to 330	330 to 360
11	2	3	3	3	3	2	1	0	0	0	0	1
10	2	2	2	2	2	2	1	1	1	1	1	1
01	1	1	1	1	1	1	2	2	2	2	2	2
00	1	0	0	0	0	1	2	3	3	3	3	2

Table 1: Phase-cal lookup table entries for 'sin' component extraction

Other table setups may be used in a similar way with fewer or more bits in each byte to represent a single quadrature component. For example, 4 bits for each component allows the extraction a single tone with a 6-level rotation model to further reduce harmonic content, with a corresponding normalization equation. Or 1 bit may be used for each component to extract four simultaneous tones with a 2-level rotation, again with a corresponding normalization equation, but with higher rotator harmonic content. For multi-bit phase rotation models, the location of rotator state changes may, of course, be adjusted as necessary to minimize rotator harmonic content. Normal operation of the Mark 4 is the simultaneous extraction of two tones using 4 bits of each lookup-table value for each tone.

5. Correlator Board

Each Correlator Board receives 64 channels of data, each of which consists of three parallel bit streams representing sign, magnitude and validity, as shown in Figure 4. In passing through the Input Board, Correlator Frames are resynchronized as necessary to compensate for varying cable lengths from SU's and are made synchronous with a common Beginning of Correlator Frame (BOCF) signal derived from timing signals generated by the associated Control Board. The rising edge of the BOCF signal marks the first bit of each Correlator Frame header while the falling edge marks the first bit of data following the CF header.

The system clock SysClk is at a constant rate[5] of 32 MHz, while the rate of the data stream is SysClk/2^r, where $0 \le r \le 4$, depending on the playback rate through the Station Unit; when $r > 0$, each data sample is duplicated 2^r times by the Station Unit. The phase/delay generators in the correlator chips are driven by SysClk while the shift/accumulate rate of the correlator chips is set to be SysClk/2^s, where $0 \le s \le 4$. For Nyquist sampled data, r is set equal to s, but r and s may be chosen to be different values when processing oversampled data, for example. The selected values of r and s apply to the entire Correlator Board.

Figure 5 shows how the 64 data channels are distributed to the 32 correlator chips on a Correlator Board. Through the use of six 64x64 crossbar switches, each correlator chip can receive any four of the 64 available channels. Depending on the configuration chosen for a particular chip, the CB may process any or all of these signals for auto and/or cross-correlation, or data may be cascaded from chip to chip to increase the number of processed lags, up to a maximum of 8192 complex lags on a single board. This switching arrangement allows flexibility in configuring the correlator board to the requirements of the user.

Again referring to Figure 4, the array of 32 correlator chips is supported by two on-board Texas Instruments TMS320C40 DSP's, one dedicated to data processing and the other to I/O

---

[5] Any change in the SysClk frequency has a global slowdown or speedup effect on the correlator.

management of data to/from the correlator chip array. Two 3MB RAM banks, labeled A and B, are used to accumulate correlated data over multiple CF's, as well as acting as ping-pong buffers for the readout of the correlator results by the VME correlator-crate host computer. Another 3MB downloadable memory bank is available for DSP program code. All of these 3MB memory banks are accessible to both the processing DSP and the VME bus, though any single bank is not available simultaneously to both. A small 1kB dual-port RAM allows real-time high-level control communications between the processing DSP and the VME crate host computer. The processing DSP includes a simple task queue-and-dispatch operating system written especially for this application.

In order to test the correlator-chip array, a 1 MB Test RAM is provided which can be loaded with test data to be processed through the correlator chips. This capability provides a board-local test capability for the health of the correlator-chip array.

Figure 6 shows the actual Correlator Board, which is constructed on a 15-layer PCB of dimensions ~40x50 cm. Board power dissipation at the nominal SysClk frequency of 32 MHz is ~100 watts. With 8 Correlator Boards plus 2 Input Boards and a Control Board in a single crate, total power dissipation per crate approaches ~1000 watts, requiring large cooling fans and proper air flow.

6. Correlator Chip

The Mark 4 correlator chip is a full-custom ASIC CMOS chip designed especially for this application. The general features and capabilities of the VLSI chip are:

- 0.8 μm feature size; ~10^6 transistors; 208-pin ceramic PGA package

- Clock rate to 64 MHz (used only to 32 MHz in VLBI application)

- 512 lags, which are distributed internally into 8 independent *correlator blocks*. Each correlator block consists of two *correlator cells*, a phase/delay generator, and a quadrature 3-level rotator with the following characteristics:

  - phase rate up to full channel bandwidth

  - phase rate and acceleration parameters for accurate phase modeling under worst case (i.e. space VLBI) conditions of delay and phase acceleration

  - internal bit-shift/phase-shift algorithm to manage fractional-bit delay

- 24-bit double-buffered counters on each lag

- 2-way data flow through correlator shift registers to eliminate need for external delay buffer

- Control of shift rate through correlator shift registers is independent of clock for processing oversampled data

- Flexible internal reconfiguration for many combinations of complex/real/auto correlation with varying number of lags

– Signal cascading from chip-to-chip to increase the number of lags by linking multiple chips

For the VLBI application, each chip receives four independent data streams into its X0-X3 inputs (see Figure 5), where each data stream consists of three parallel bit streams representing sign, magnitude and validity. In addition, four additional inputs Y0-Y3 are provided for another application[6]. Internal signal-steering switching allows the chip to be configured into an extremely large ($\sim2^{96}$) number of modes, of which perhaps 100 or so are actually useful; Table 2 lists a sampling of single-chip modes. This flexible architecture allows the chip to satisfy both connected-element[7] and VLBI types of interferometric applications. In Table 2 the configurations which involve complex and/or auto-correlation are for support of VLBI, while those which are only real correlation are primarily used for connected-element applications. In addition, a number of input and output signals are provided to 'cascade' correlation from chip-to-chip to increase the number of lags by using multiple chips; again, many cascade modes are also supported.

Configuration	Correlations		#Lags per correlation	
	Auto	Cross	Auto	Cross
4x4 real cross	-	16	-	32
4x2 real cross	-	8	-	64
4x1 real cross	-	4	-	128
2x2 real cross	-	4	-	128
2x1 real cross	-	2	-	256
2x2 complex cross	-	4	-	64
2x1 complex cross	-	2	-	128
2x2 real 'auto'	4	2	64	128
2x1 real 'auto'	2	1	128	256
2x2 complex 'auto'	4	2	64	64
2x1 complex 'auto'	2	1	128	128
1x1 real cross	-	1	-	512
1x1 complex cross	-	1	-	256
4 Input real 'auto'	4	-	128	-
2 Input real 'auto'	2	-	256	-
1 Input real 'auto'	1	-	512	-

Note: Number of lags divided by 2 if validity-per-lag is used

Table 2: Some Possible Correlator Chip Configurations

6.1 Correlator Block

A simplified schematic diagram of a single correlator block configured for complex correlation is shown in Figure 7. Not shown in this figure are numerous signal selectors and internal block connections that also allow a block to be used for real or auto-correlation, which are simpler operations [Aldrich, 1993; Bos, 1996]. The basic elements of the block are two identical 'cells', dubbed 'real' and 'imaginary', a header-capture buffer, and a common phase/delay generator.

[6] The Y inputs can be used with the X inputs to form a 4x4 array of real correlators. This configuration is used by the Smithsonian Submillimeter Array (SMA) at Mauna Kea, Hawaii to allow the single-baseline real correlation of data sampled at 212 Msample/sec which is de-multiplexed to 4 parallel data streams of 53 MHz each.
[7] The correlator chip and Correlator Board are used, without modification, to support both the Smithsonian Submillimeter Array (SMA) at Mauna Kea, Hawaii, and the Westerbork radio telescope array (WRST) at Dwingeloo, The Netherlands.

The data stream consists of three parallel bit-streams representing sign, magnitude and data-validity. Three primary control signals manage the affairs of the block:

1. SysClk – system clock, nominally a constant 32 MHz reference signal.

2. BOCF (Beginning of Correlator Frame) – Active during each CF header, the BOCF signal controls header capture and the activation of the phase/delay generation at the beginning of the data part of the CF.

3. Shift/Accumulate – Enables shifting of data through the correlation shift registers and accumulation in correlator registers.

Several other control signals not shown in Figure 7 manage more mundane affairs of the block, such as internal switch settings, download of parameters to the phase/delay generator and readout of the header capture buffers, correlator accumulators, etc. Several switchable 1-bit delays allow proper configuration for various cascade modes of auto and cross-correlation.

6.2 Header Capture

The CF header data is carried on the 'validity' bit stream during the time the BOCF signal is active. This 240 bits of station data is captured in a buffer to be read out by a DSP on the Correlator Board and subsequently used, along with CF header data from the correlation-partner channel, to construct baseline correlation processing parameters sent to the correlator chip.

6.3 Correlator Element

Each Correlator Element uses two shift registers shifting in the opposite directions. The registers are 3 bits wide to carry sign, magnitude and validity information. In so-called 'global validity' mode, 32 lags are correlated and accumulated, and a global valid-sample count accumulator is maintained; the shift registers may also be reconfigured to compute 16 lags with a valid-sample count accumulator on each lag. The Shift/Accum signal enables the shifting of data through the shift registers, which may be at a sub-multiple of the SysClk rate to accommodate the processing of oversampled data.

Multiplier/accumulators in each Correlator Element calculate the correlation product of each sample controlled by the Shift/Accum signal and add the result to a 3-bit adder followed by a 24-bit ripple counter; only the 24 bits of the ripple counter are available to the user. The multiplication table is the same as that used in the earlier NFRA correlator chip [Bos, 1991; Cooper, 1970] extended with the handling of the validity data bit, as given by Table 3. Note that the combination $(S,M,V)=(0,0,0)$ is a special case which represents the zero state of the 3-level rotator; any product involving this state adds 3 to the accumulator (which is equivalent to a zero product) and increments the sample count; if $V=0$ for any non-zero state of the rotator, the data are simply ignored (zero is added and the sample count is not incremented). For 2-bit data, this multiplication scheme produces a SNR gain factor of ~0.875 with respect to an analog correlator with a sampling decision level of 0.91 times the rms level of the analog input signal [Bos, 1996]. For 1-bit data, the magnitude bit is fixed to 1, and the corresponding SNR gain factor is ~0.64.

			S	0	0	1	1	0	0	1	1
			M	1	0	0	1	1	0	0	1
			V	0	0	0	0	1	1	1	1
S	M	V									
0	1	0		0	3	0	0	0	0	0	0
0	0	0		3	3	3	3	3	3	3	3
1	1	0		0	3	0	0	0	0	0	0
1	0	0		0	3	0	0	0	0	0	0
0	1	1		0	3	0	0	6	4	2	0
0	0	1		0	3	0	0	4	3	3	2
1	0	1		0	3	0	0	2	3	3	4
1	1	1		0	3	0	0	0	2	4	6

Table 3: Correlation Multiplication Table (S=sign, M=magnitude, V=valid

When the Correlator Element is configured to count valid samples per lag, the sixteen 24-bit counters which maintain the valid-sample count do so with no pre-scaling. Direct counting of the number of valid samples is important for proper normalization, particularly for low-correlation amplitude, since the correlation amplitudes are determined by small differences in large numbers, and any systematic errors in normalization factors can lead to erroneous results. Pre-scaling of the correlation counts, on the other hand, is acceptable, since the noise in the CF correlation coefficients is virtually guaranteed to be noisier than the noise introduced by pre-scaling. Note that in practice, due to the 24-bit limit on valid sample count, the maximum number of data samples in a CF is limited to 16 Msamples, corresponding to a half-second at a playback rate of 32 Msample/sec.

6.4 Phase/Delay Generator

The phase/delay generator is responsible for generating the rotation phase applied to one of the station-data streams and for the management of baseline vernier delay to properly maintain the baseline delay within ±½ sample periods. Proper management of rotator phase and vernier-delay is at the heart of the Mark 4 correlator's ability to properly process VLBI data, which we will examine in some detail in the next section.

7. Phase/Delay Management

The data transmitted from each Station Unit to a Correlator Board, and hence to a Correlator Chip, are pre-delayed according to a model provided to the SU by the host computer, as explained earlier. For short-baseline interferometry it is normally sufficient to change the delay only on CF boundaries, so that each CF (~15 msec to 500 msec) is correlated with a fixed delay. For VLBI, however, this is not normally possible since station delay rates may be up to ~10 km/sec (i.e. ~1 sample/30,000 samples, independent of sample rate) for space-VLBI; even in normal terrestrial-based VLBI, delay rates may be as high as $\sim 3 \times 10^{-6}$ (i.e. 1 sample/300,000 samples). Therefore, it is imperative that delay management be dynamic during each integration period.

The statement of the delay-management problem is straightforward: The delay of the data from each Station Unit is adjusted as needed so that the *station delay-error*, that is, the difference between the smoothly varying model station delay and the quantized delay from the SU, is

always less than $\pm\frac{1}{2}$ sample period. The uncorrected *baseline delay-error* as it reaches the correlator chip is thus within the range ± 1 sample periods; whenever the absolute uncorrected baseline delay exceeds $\frac{1}{2}$ sample, an adjustment of ± 1 sample is necessary to properly correct it. Therefore, an adjustable vernier delay of -1,0,+1 samples is sufficient to always maintain the baseline-delay-error within the necessary range of $\pm\frac{1}{2}$ sample periods.

To best understand the phase/delay generators and their role in the Mark 4 correlation processing, it is useful to construct a simplified schematic diagram of a single Mark 4 complex correlator, as is shown in Figure 8. The X-data are routed directly to quadrature rotators after a 1-sample delay, while the Y-data are passed through a 3-tap vernier delay shift register. The 1-sample delay in the X-data stream compensates for the single-bit delay to tap B of the vernier-delay shift register, so that taps A and C adjust the Y-data-stream delay by ± 1 samples from that applied by the SU. By dynamically selecting the proper tap (A, B or C), the baseline delay error can always be maintained in the range $\pm\frac{1}{2}$ samples, as required.

The phase rotator is also affected by either an X delay change or a baseline delay change. In particular, in the event of an X delay change of one sample, the phase rotator must either be correspondingly stopped or double-incremented for one sample so that the rotator phase applied to the X data stream is continuous with respect to the X-station reference clock. If the baseline delay changes by one sample, the rotator must be simultaneously incremented by $\pm 90/s$ degrees, where s is the oversampling factor (more on this later). With this understanding, we can construct an algorithm that will properly correlate the VLBI data.

For purposes of this discussion, define the station delay $\tau(t)$ such that $\dot{\tau}(t) > 0$ for a station moving away from the source with respect to the center of the earth [in such case, the Station Unit compensates by occasionally dropping a recorded sample from the data stream transmitted to the correlator]. Define ΔX, ΔY and ΔB as the respective delay errors in the sense of [quantized minus model]. At the beginning of a CF, the initial vernier-delay tap position is set according to the rules given by Table 4.

Initial Condition at t_0 (i.e. start of corr frame)	Initial Tap Position	Initial Baseline delay error, ΔB_0
$\Delta Y_0 - \Delta X_0 < -0.5$	A (most upstream)	$\Delta Y_0 - \Delta X_0 + 1$
$-0.5 < \Delta Y_0 - \Delta X_0 < 0.5$	B (middle)	$\Delta Y_0 - \Delta X_0$
$\Delta Y_0 - \Delta X_0 > +0.5$	C (most downstream)	$\Delta Y_0 - \Delta X_0 - 1$

Table 4: Rules for Initial Vernier-Delay Tap Setting

Dynamic delay changes in the X or Y data are compensated by moving the current tap-point so that the relative delay of the data into the correlator is unchanged. It is important that these compensating actions take place *at precisely the instant of the delay-shifts from the station units* so that no data are correlated with an incorrect delay. Table 5 prescribes the actions to be taken to compensate for an X, Y or baseline delay shift, as indicated by a carry from the corresponding delay-error register.

Condition	Actions on delay shift
$\dot{\tau}_x(t) > 0$	*On X delay-shift*: X Station Unit drops one sample; Shift upstream 1 tap position; Double-increment phase-generator
$\dot{\tau}_x(t) < 0$	*On X delay-shift*: X Station Unit duplicates one sample; Shift downstream 1 tap position; Null-increment phase-generator
$\dot{\tau}_y(t) > 0$	*On Y delay-shift*: Y Station Unit drops one sample; Shift downstream 1 tap position
$\dot{\tau}_y(t) < 0$	*On Y delay-shift*: Y Station Unit duplicates one sample; Shift upstream 1 tap position
$\dot{\tau}_b \equiv \dot{\tau}_y(t) - \dot{\tau}_x(t) > 0$	*On baseline delay-shift*: Shift upstream 1 tap position; Apply appropriate baseline phase shift
$\dot{\tau}_b(t) < 0$	*On baseline delay-shift*: Shift downstream 1 tap position; Apply appropriate baseline phase shift

Table 5: Rules for Vernier-Delay Tap Motion

Figure 9 shows a simple example of delay tracking over a period of time in which several delay shifts occur in both the X and Y data streams. The upper two plots show the Y delay and Y-delay error (ΔY). The next two plots show the X delay and the X-delay error (ΔX). The dashed line of the lower plot shows the uncorrected baseline delay, $\Delta Y - \Delta X$, which ranges between the values of -1 and +1. The superimposed solid line shows the baseline-delay-errors after the appropriate tap shifts have been made according to the rules outlined above. Note that the baseline-delay-error stays within $\pm\frac{1}{2}$ sample and tracks smoothly between baseline delay-shifts, as it should. Events labeled 1-21 are indicated below the baseline-delay-error plot and correspond to the actions listed in Table 6.

Event#	Event	Move to Tap
1	Initialize	C
2	Simultaneous X-delay shift, baseline-delay shift	A
3	Y-delay shift	B
4	Y-delay shift	C
5	Baseline-delay shift	B
6	X-delay shift	A
7	Y-delay shift	B
8	Y-delay shift	C
9	Baseline-delay shift	B
10	Simultaneous X-delay shift, Y-delay shift	B
11	Baseline-delay shift	A
12	Y-delay shift	B
13	Y-delay shift	C

14	X-delay-shift	B
15	Baseline-delay shift	A
16	Y-delay shift	B
17	Y-delay shift	C
18	Simultaneous X-delay shift, Baseline-delay shift	A
19	Y-delay shift	B
20	Y-delay shift	C
21	Baseline-delay shift	B

Table 6: Example of Vernier-Delay Tap Motion

7.1 Delay Generator Implementation

A detailed diagram of the phase/delay generator is shown in Figure 10. Each of the three delay-error generators consists of two registers, one representing delay error and the other delay-rate, plus an adder. Each 32-bit delay-error register represents the delay error over the range of $\pm\frac{1}{2}$ samples, and is incremented at each SysClk cycle by the value of the corresponding 18-bit delay rate. When a carry appears from the MSB of the delay-error register, the appropriate tap-shifting/phase-shifting action is commanded to take place. By appropriately adjusting the initially-loaded value into the delay-register, only positive increments need be made, so that only the absolute value of the delay-rate needs to be loaded into the 18-bit delay-rate register. In order to prevent any data already within the correlator shift registers at the time of any delay or phase shift from being correlated with the wrong delay or phase, all data in the shift registers are set invalid at the instant of the delay/phase shift. Correlation resumes as these invalidated data are shifted out.

The maximum delay rate supported by the 18-bit delay-rate register is a one-sample delay-shift per $2^{(32-18)}=16384$ SysClk periods, which corresponds to a delay rate of ~60 microsec/sec and is quite adequate for even worst-case space VLBI observations. The resolution of the delay-rate register is 2^{-32} samples/SysClk, resulting in a maximum accumulated error over n SysClk periods of $\pm 0.5 \cdot n/2^{32}$ samples. For the maximum value of $n = 1.6 \cdot 10^7$ SysClk periods in a single Correlator Frame, this corresponds to a maximum error of ~0.002 samples, which is negligible.

In order to guarantee that ΔX, ΔY and ΔB are tracked properly over each CF, the models for computing ΔX and ΔY should be *exactly* the same as that used by the SU over the same CF. This is guaranteed because each X and Y CF header contains the values for ΔX_0, $\Delta \dot{X}$ and ΔY_0, $\Delta \dot{Y}$, respectively, used by the corresponding Station Units. If ΔB_0 is calculated to be *exactly* $\Delta X_0 - \Delta Y_0$ and $\Delta \dot{B}$ is calculated to be *exactly* $\Delta \dot{X} - \Delta \dot{Y}$ and the initial tap position is set properly at the beginning of CF, there will never be an attempt to shift outside of the three vernier-delay tap positions.

7.2 Phase Generator Implementation

As shown in Figure 10, the phase generator section of the phase/delay generator incorporates a set of three 32-bit registers, one each for phase, phase-rate, and phase-acceleration. Ignoring for the moment the phase increments applied at X-delay or baseline-delay shifts, five parameters are set at the beginning of each Correlator Frame:

14

1. Initial Phase: The 32-bit phase register is loaded with the initial rotator phase and then subsequently updated. The full range of the counter represents a single rotation with a resolution of 2^{-32} rotations. The upper 4 bits are used to define the instantaneous state of a 16-state 3-level quadrature phase rotator in the standard manner [Thompson, 1986].
2. Initial Phase-rate: The 32-bit phase-rate register is loaded with the initial phase-rate value, which is added to the phase register every k SysClk periods.
3. Phase acceleration: The 32-bit register acceleration register is loaded with a constant value, which is added to the phase-rate register every n SysClk periods.
4. Phase update period, k: Controls phase update rate; the phase register is updated every k SysClk periods. The value of k may be set to 1,2,4,8 or 16. Larger values of k increase the effective phase-rate resolution.
5. Phase-rate update period, n: 24-bit register which controls the phase-rate update rate.

In order to analyze the performance of this phase-generator, define the following quantities:

a, the phase-acceleration (Hz/sec)
c, SysClk frequency (SysClk periods/sec)
m, number of bits in the phase register (m=32 for Mark 4 correlator)
f, sample frequency (samples/sec)
k, phase-register update period (SysClks)
n, phase-rate update period. The phase-rate is held constant during each period of n SysClks.
A, the phase-rate increment value placed in the phase-acceleration register
N, integration period (samples)

The phase-rate quantization is given simply by

$$\Delta f = \frac{c}{2^m k} \qquad [\text{Hz}] \qquad (1)$$

which corresponds to ~ 0.47 mHz for m=32, k=16, for example. The maximum phase rate is $\pm f/2^k$, so that extremely high phase-rates are accommodated by a smaller value of k.

There are three types of phase errors that can occur during an integration period, where phase error in this context is defined as the departure of the hardware-generated phase from a smoothly varying second-order phase polynomial:

1. α, the accumulated phase error due to the phase-rate quantization of the rotator.

2. ϕ, the departure of the rotator phase from the model phase due to ignoring acceleration during the fixed-rate segment of the rotator.

3. θ, the accumulated departure of the rotator phase from the model due to quantization of the acceleration register.

We will examine the worst-case values of each of these types of phase errors.

The *maximum* contribution to phase error due to phase-rate quantization over the integration period of N samples is simply

$$\alpha = \frac{\Delta f}{2} \cdot \frac{N}{f} \qquad [\text{rotations}] \qquad (2)$$

Note that α can be minimized by maximizing the values of m (fixed at 32 for Mark 4) and k (maximum value 16).

The second contribution to phase error is due to unmodeled phase-acceleration over the n SysClks during which the phase-rate is held constant. For a single constant-phase-rate segment, this is simply given by

$$\phi = \frac{1}{2} a \cdot \left[\frac{n}{c} \right]^2 \qquad \text{[rotations]} \qquad (3)$$

For the accumulated error over *multiple* segments in the accumulation period we then have

$$\phi = \frac{anN}{2cf} \qquad \text{[rotations]} \qquad (4)$$

The third contribution to the phase error is due to quantization of the phase rate increment. Either A or n can be adjusted to make the long-term acceleration as close as possible to the actual acceleration, a. Since n is generally the larger integer, and the acceleration is proportional to the ratio A/n, we assume that A is exact at an integral value, and n is wrong by the worst case of ½. The resulting phase error is

$$\theta = \frac{AN^2 c^2}{2^{34} kf^2 n^2} \qquad \text{[rotations]} \quad \text{for } cN >> fn \qquad (5)$$

By differentiating the expression for $(\theta + \phi)$, setting the result to zero, and solving for $n,$ we find that

$$n = \sqrt{\frac{cN}{2f}} \qquad (6)$$

This value of n then determines the integer value of $A,$ and then we go back and correct n to be in agreement with A. This procedure does not necessarily give us the optimal values of A and n, as there are fortuitous combinations of A and n that result in very small values of θ. This happens because A is set to the closest integer to

$$\frac{2^{32} nak}{c^2} \qquad (7)$$

so some values of n, a, k and c will result in very small truncation errors. However it was not felt that the slight improvement so achieved was worth the additional algorithmic complexity. Furthermore, since n is used in the correlator-chip phase generator on a baseline basis, the calculation of the value of n must be deferred to the point in the signal path where the baseline acceleration, a, can be formed.

Once the values of n and A have been determined, expressions (3) and (6) can be used to evaluate the *maximum* phase errors of the rotator with respect to an exact second-order polynomial model. These have been calculated for a wide range of f, a, and N, covering all practical cases for which the rotator might be employed. For a sample rate of 32 Msample/sec and 16×10^6 samples per CF, phase errors are below 1.6 degrees for 100 Hz/sec fringe rate acceleration, which represents worst-case terrestrial accelerations at 300 GHz observing

frequency. For lower sample rates, phase errors progressively increase, decreasing the maximum allowable value of *N* for a given acceleration; however, rotator phase error can be significantly reduced by using shorter CF's or by oversampling narrow-bandwidth channels when data are initially recorded.

7.3 Computation of Rotator Phase and Rate

The rotator phase is applied to the X data stream of Figure 8 as a smoothly incrementing function of X-station time, except at baseline-delay shifts, since X delay shifts are compensated for by either a null or double-increment of the rotator phase. Furthermore, the 3-position delay tap ensures that the baseline delay does not exceeds ±1/2 sample as the data enter the correlation shift registers. Therefore, we may analyze the phase function applied to the rotator independently of X-delay shifts and examine the behavior of the rotator as if the delay of the X data stream is held constant and all delay changes occur in the Y data stream. In this conceptual model, the following phase generator rules apply, as illustrated in Figure 11:

1. The long-term rotator phase tracks the model fringe phase at the *DC-edge* of the baseband channel.

2. Within an accumulation period, the phase-rate of the rotator corresponds to the model fringe-phase-rate at *center* of the baseband channel.

3. At every quantized baseline delay shift, there is a corresponding rotator phase shift given by

$$\Delta\phi_{shift} = -sign(f_{rf} \cdot \dot{\tau}_b) \cdot \frac{\pi}{2} \cdot \frac{2 \cdot BW}{f_s} \qquad \text{(radians)} \qquad (8)$$

where f_{rf} is the signed radio-observing frequency (>0 for net USB; <0 for net LSB), $\dot{\tau}_b$ is the baseline delay rate ($\dot{\tau}_b$ >0 for station Y moving away from source faster than station X), *BW* is the channel bandwidth, f_s is the sampling frequency, and the function *sign*(*a*) is +1 if *a* >0 and -1 if *a* <0. For Nyquist sampling, the value of $\Delta\phi_{shift}$ is +/-90 degrees, with corresponding smaller values for oversampled data.

4. At the beginning of each correlator frame, the initial rotator phase, $\phi_{rot}(t_0)$, must be set to the model fringe phase at the DC-edge of the BBC channel, $\phi_0(t_0)$, *plus* a correction corresponding to the initial baseline delay error, $\delta_b(t_0)$, so that the initial rotator phase is set to

$$\phi_{rot}(t_0) = \phi_0(t_0) - sign(f_{rf}) \cdot \frac{\pi}{2} \cdot \frac{2 \cdot BW}{f_s} \cdot \Delta B_0 \qquad \text{(radians)} \qquad (9)$$

where ΔB_0 is calculated according to Table 4.

5. The initial fringe rate at the center of the observing band, $\dot{\phi}_c(t_0)$, is set to

$$\dot{\phi}_c(t_0) = \dot{\phi}_0(t_0) + \dot{\tau}_b \cdot \frac{BW}{2} \qquad (10)$$

For each Correlator Frame of each channel, each Station Unit transmits the model fringe phase at the DC-edge of the BBC channel plus a linear model fringe rate at BBC band-center. The

models from the stations at the baseline ends are directly differenced to compute the baseline fringe rate and acceleration, while the initial rotator phase is computed according to Equation 9.

When used at the Nyquist sample rate, this algorithm contributes a continuum SNR loss, averaged over the full channel bandwidth of ~3.4%. The contribution to spectral-line loss is 0% at center-band, increasing to 10% at bandedge [Thompson, 1986]. Note that these losses are significantly reduced if oversampled data are processed, as is often the case with narrow-band spectral-line observations.

8. Complex Cross-Correlation Process

8.1 Correlator Chain Configuration

For VLBI complex correlation, two or more correlator blocks are interconnected to cross-correlate a single channel from two stations, as shown in Figure 12. By convention, the 'X' data stream is presented to the 'head block' and the 'Y' data stream is presented to the 'tail block'; the X and Y data streams each consist of three parallel bit-streams representing sign, magnitude and data-validity. An even number of 'mid' blocks may be optionally connected between the head and tail blocks to increase the number of lags to the limit of 8192 available on a Correlator Board. The phase/delay generator in *every* block in the Correlator Chain is operated in parallel with *exactly the same set of parameters*, but different functions are allowed to operate in the head, tail and mid blocks:

- Header capture is active in head and tail blocks so that the head block captures the station-based parameters periodically embedded in the X data stream and the tail block captures the corresponding parameters for the Y data stream.

- The 'head' block is configured to fix the position of the vernier-delay tap; the quadrature rotators are active.

- The 'tail' block is configured to bypass the rotators, but the vernier-delay tap is active.

- All 'mid' blocks have rotators bypassed and vernier-delay tap positions fixed.

- Every block is configured to invalidate data in its shift registers whenever an X, Y or baseline delay-shift event is generated to prevent the correlation of any data with incorrect delay or phase.

Because the phase/delay generators in the 'head' and 'tail' blocks are operated exactly in parallel, they may be considered as a single phase/delay generator for purposes of creating the functional block diagram of a VLBI complex correlator as shown in Figure 8.

8.2 Delay and Phase 'Smearing'

Correlation with a very large number of lags can create a 'delay smearing' loss at very high delay rates and accelerations. Delay smearing occurs when the baseline delay changes a significant fraction of a sample period over a time interval corresponding to the number of lags. For example, in the worst-case space-VLBI case of the delay changing by as much as 1 sample/30,000 samples, a direct 8192-lag correlation would cause delay smearing of ±0.136 sample periods around the center lag, with a corresponding SNR loss of ~0.25% [Thompson, 1986]. One method of overcoming this loss is to transmit dynamic station phase in parallel with each station data stream and dynamically compute rotator phase on a lag basis [Carlson, 1999], but this requires extremely heavy overhead in model transmission from station unit to the

correlator. In the Mark 4 correlator, this effect may be overcome by breaking the correlation into several shorter Correlator Chains, each filling a contiguous portion of the total lag space and each with its own phase/delay model, to create the total number of desired lags. This is possible because the Station Unit can duplicate a single channel to multiple SU outputs, each with its own delay and phase model to be processed by each of the shorter Correlator Chains. For example, executing the same 8192-lag correlation using eight 1024-lag Correlator Chains would reduce the delay smearing to ±0.017 sample periods, with a corresponding negligible loss of ~0.004%.

8.3 Correlation Process

As explained in the Station Unit discussion, the data from each channel are formulated into Correlator Frames (CF) by the Station Unit, each of which may be up to a maximum of 1.6×10^7 samples in length. Each CF from each Station Unit includes a 240-bit 'header' which contains *a priori* station model data for the *next* CF, followed by the data of the *current* CF. The correlation operation is 'pipelined' such that, during the correlation of CF *n*, the correlation data from CF *n-1* are read from the correlator-chip and the processing parameters for CF *n+1* are computed and written to the chip in preparation for correlation of CF *n+1*.

To understand the operation in detail, it is useful to separately discuss the actions that take place during the header part and the data part of Correlator Frame *n*. During the time that the 240-bit station header is being transmitted, the following actions take place within each block of a correlator chip:

1. Correlation is suspended.

2. The contents of all lag-accumulation counters in the chip, as well the final state of all counters, registers and status flags in the phase/delay generator[8], all of which correspond to just-completed CF *n-1*, are transferred to a set of on-chip static buffers.

3. The lag-accumulation counters are cleared in preparation for correlating CF *n*.

4. The phase/delay generators are initialized from on-chip buffers which have been previously loaded with parameters for processing CF *n*.

5. The X and Y 'validity' bits are set globally to 'invalid' at each lag, ensuring that no residual data from the previous frame are correlated when processing of CF *n* starts.

6. Each 'head' block captures the 240-bit CF header for station X while each 'tail' block captures the CF header for station Y. This captured station-header data contains the *a apriori* station model information (initial phase and fractional delay, phase rate, phase acceleration, etc) pertinent to CF *n+1*.

of a CF *n*, which may be in lengths as long as a half-second, the following actions take place under control of a local DSP on the Correlator Board:

1. From each correlator block on a correlator chip, the buffered correlation counts from CF *n-1,* along with the final state of the phase/delay generator and status indications, are read and accumulated into either RAM Bank A or B on the Correlator Board[9].

---

[8] The terminal values of counters and registers in the phase/delay generator can be used as a powerful diagnostic tool for proper operation.

[9] Normally, chip correlation data are further accumulated on the Correlator Board for periods of up to several seconds, before being transferred to the host VME crate computer.

2. The captured CF header data is read by the DSP from all 'head' and 'tail' blocks of the correlator chip; the DSP then combines the *a priori* station-model data from corresponding pairs of 'head' and 'tail' blocks to compute the difference delay, delay rate, phase and phase rate parameters for correlation of CF *n+1*; these processing parameters are written back to temporary holding buffers on the chip, poised to be made active for CF *n+1*.

Note that the embedded DSP has nearly an entire correlator frame to perform these actions, which are asynchronous with respect to the data stream and which must only be completed before the beginning of CF *n+1*.

The process outlined above repeats continuously as the processing progresses. It is important to note that the station-model information transmitted in the station 240-bit data-frame headers is completely adequate to locally create a difference model which can sustain the correlation processing with no additional information. In this way, dependence on real-time intervention or synchronism with any device or process outside of the Correlator Board is removed. Furthermore, because the station model data and ID are embedded in the each data stream, there is no possibility for the model and corresponding data to be separated. In a large correlator system such as the Mark 4 with many hundreds of channels of data having to be routed to their correct destinations, this provides some level of comfort.

9. Correlator Configurations and Capabilities

As shown in Figures 1 and 5, the signal distribution to Correlator Boards and Correlator Chips is highly flexible. The four Correlator Boards that form each 'slice' in Figure 1 receive all signals from all stations, so that a slice is a useful processing element and is the minimum hardware allocation normally used to process a single scan. The controlling software is organized such that the processing of a single scan may occupy from one to four slices. Furthermore, within the limits of the available correlator resources, up to *four independent scans may be processed simultaneously*. This 'multi-streaming' capability is particularly useful for geodetic-VLBI processing where a set of observing stations may divide into two or more simultaneous independent subsets of stations observing different sources, and subsequently coalesce into a single array.

9.1 Continuum Processing

Continuum processing is usually conducted with the complex Correlator Chains formed of only a 'head' and 'tail' block (see Figure 12) for a total of 32 lags (with validity per lag) per channel per baseline. Typically, a single Correlator Board is configured to process all baselines of cross-correlation and the corresponding station auto-correlations for one or more channels. Table 7 shows the 32-lag continuum processing capabilities for modes using 1, 2 or 4 correlator 'slices' for both single and dual-polarization processing; channel data rates may be up to 32 Msample/sec (64 Mb/sec) in each case. Integration periods may be specified from ~100 msec to ~10 seconds depending on the field of view desired and, in some cases, Ethernet data capacity between Correlator Crate VME computers and the Control Computer.

#Slices	Capability				Comments
	Polar-ization	#stns	#baselines	#chns/station	
4	Single	16	120	16	Complete 16-station processing w/auto-corr
2	Single	11	55	16	Complete 11-station processing w/auto-corr
1	Single	8	28	16	Complete 8-station processing w/auto-corr
4	Dual	16	120	4	Full dual-pol processing with w/auto-corr
2	Dual	11	55	4	Full dual-pol processing with w/auto-corr
1	Dual	8	28	4	Full dual-pol processing with w/auto-corr

Table 7: Correlator capability for 32-lag (with sample-count/lag) continuum processing

Note that with a full complement of 16 PBU's and Station Units, for example, two independent 8-station continuum experiments may be processed in simultaneous processing 'streams'. Data recorded at sample rates <16 Msample/sec may, of course, be reproduced at rates up to 32 Msample/sec, depending on the capabilities of the PBU.

9.2 Spectral-line Processing

Up to 8192 complex lags (w/global validity) may be processed on each of the sixteen Correlator Boards. These available lags may be flexibility traded among channels and baselines; normally auto-correlations are done simultaneously with cross-correlations. For example, data from an 8-station 1-channel experiment scan can be processed with 4096 lags along with full station auto-correlations. Or an 8-station dual-polarization experiment can be processed with 1024 lags with all polarization combinations processed.

The Mark 4 correlator installed at JIVE [Schilizzi, 2001] includes a Data Distributor module between the Station Units and correlator proper which allows more flexibility in data distribution; it also allow data re-circulation to the correlator, increasing the number of available lags for data replayed at less than the maximum 32 Msample/sec rate, and improving processing efficiency for low-data-rate spectral-line observations. The correlators at Haystack Observatory, USNO and MPIfR do not include a Data Distributor module since the majority of data processed at these installations is continuum in nature.

9.3 Fringe Finding

Of course, the large lag space available on the Mark 4 correlator is useful in fringe finding for observations complicated by malfunctioning clocks or equipment. Several experiments have been recovered which would have been hopelessly compromised by the small search-lag-space available in the previous generation of Mark 3A correlator facilities. The ability to dedicate one of the processing 'streams' to fringe finding while performing routine processing with another 'stream' also contributes significantly to the overall throughput capability of the Mark 4 correlator.

9.4 e-VLBI Capabilities

The new Mark 5 disc-based VLBI data system [Haystack Observatory 2002] has been designed to add new capabilities when used with the Mark 4 correlator. In particular, the ability will be added to process VLBI data in both real-time and near-real-time from data transmitted from stations over high-speed wide-area networks. This is possible because the Mark 5 system is based on a standard PC platform which allows raw station data to be flexibly moved between discs, a network connection and the VLBI data source/sink. For example, in cases where data must be recorded at one or more stations and then transmitted at a slower rate, the Mark 5 system can be used at the correlator as a temporary buffer until all data have arrived for correlation. If available network connections allow for real-time data transmission, the Mark 5 can transmit incoming data directly to the Mark 4 correlator for processing. Even operating at 1 Gbps, the Mark 5 provides two full seconds of solid-state buffering to accommodate any differences in transmission latency. First tests of e-VLBI of the Mark 4 correlator with the Mark 5 data system are expected in 2002.

10. Software Architecture

The control software for the Mark 4 correlator system is necessarily quite complex, though considerable effort has been expended to enable both flexibility and efficiency of operation. A simplified high-level block diagram of the correlator control software architecture is shown in Figure 13. The software itself is written mostly in standard C, running on an HP C340 computer under HP-UX or a HP744 VME single-board computer under HP-RT.

10.1 Correlation

The primary control of the correlator is managed by the software module 'conductor' which, as the name implies, orchestrates the many actions that need to be coordinated to properly correlate a data set. Conductor is implemented as a finite-state machine, with transitions between states caused by events. These events are comprised of messages passed from other modules in the system as well as commands from ASCII script files called Task Stream Files, each of which contains a list of scans and stations to be processed from a specific experiment, along with ASCII 'keys' specifying correlator configuration (number of correlator slices to be used, number of lags to be processed, etc). Up to four independent 'streams' may be processed concurrently, with a single Task Stream File specifying up to two simultaneous streams; multiple Task Stream Files may be active at the same time so long as the number of streams does not exceed four. 'Conductor' is supported in its role by four major subordinate modules which each execute specific tasks:

- 'genaroot' creates a root data file for each scan which contains all *a priori* model and configuration information, as well as creating a file containing quintic-spline approximations to the high-precision CALC8 model.

- 'corr_man' is responsible for direct management of the correlator subsystem, including configuration and data readout to an output file.

- 'su_man' is responsible for management of all Station Unit actions.

- 'opera' manages the operator interface, allowing the operator to control and monitor the processing of the scans specified in the Task Stream Files; it is implemented by a mixture of C and Tcl/Tk.

The specification of an experiment schedule to be processed and all of the possible configurations of the correlator are specified in a set of ASCII text files based on the internationally-agreed VEX syntax and rules [VEX, 1998]. These files rely heavily on the use of ASCII keys to extract desired information. This set of files includes:

- ovex – 'observe' file (one per experiment). This file is a superset of the observing 'vex' file used at the stations to drive data collection. It specifies the complete setup and configuration of the data-acquisition system for each scan, as well as a complete list of scans and participating stations, plus *a priori* station locations, source positions, EOP, etc.

- lvex – 'log' file (one per experiment). This file is created at the correlator from the information contained in the log files from the stations participating in an experiment. Information of particular use to the correlator are such things as tape positions and tape-drive headstack-position offsets for each scan, which allow the PBU to operate more efficiently.

- evex – 'experiment' file (global). This file contains a set of reference parameters associated with each experiment, such as associated experiment-specific control file names, processing speed-up factors and correlator operating modes.

- ivex – 'initialization' file (global). This file contains a list of several selectable global initialization states of the correlator, for normal processing as well as for tests and special processing. Each initialization state includes Station Unit and tape-transport parameters, data and communications routing, etc.

- svex – 'SU' file (global). Contains detailed configurations for Station Units, such as channel output mapping and algorithms for phase-calibration processing.

- cvex – 'correlator configuration' file (global). The cvex file specifies in detail the configuration of the correlator for many different applications. Each configuration is referenced by a key within the Task Control File. Correlator configurations of different combination of lags, baselines and channel, plus any special configurations, may be specified in the cvex file.

Each processed scan leads to the creation of a suite of output files, including the root file, a correlation output file for each baseline, and individual station-based files for playback error statistics, sampler statistics and phase-calibration. In addition to the processed output data, the file suite for each scan includes complete information for model reconstruction and correlator-configuration reconstruction.

10.2 Post-Correlation

The first stage of post-correlation processing analyzes the data from each baseline of a scan individually to find fringes and estimate observable parameters, including the correlation coefficient, phase, group delay and phase-delay rate. This processing is always fully coherent over all processed channels. The data are automatically monitored for quality according to the inter-relationship of several observed parameters, such as RMS phase variation vs. SNR, phase-scatter between channels and phase-calibration stability. The algorithms used in this processing are similar to the algorithms used by the Mark IIIA correlator [Whitney, 1976] and will not be detailed here.

A typical example of fringe-finding output is shown in Figure 14. This particular scan is from a 16-channel S/X-band geodetic VLBI experiment between antennas in Gilmore Creek, Alaska

and Matera, Italy, with ten 8-MHz-wide channels devoted to X-band and six to S-band. Shown in Figure 14 are the X-band results only, which are composed of 10 channels spaced in frequency from 8213 MHz to 8933 MHz for high-precision group-delay determination; the two 'outrigger' frequency bands (8213 MHz and 8933 MHz) are actually each composed of adjacent upper/lower sideband channel-pairs spanning 16 MHz each; although these four sideband channels are recorded and correlated separately, each upper/lower sideband pair is combined in the fringe-search procedure to appear as a single frequency band since they share a common BBC local oscillator and are consequently inherently coherent. The six inner frequency bands each consist of a single 8-MHz-wide USB channel. The set of six S-band 8-MHz-wide USB-only channels were correlated simultaneously, but are not shown here. Data were recorded at 16 Msample/sec/channel, 1 bit/sample, for a total combined 16-channel data rate of 256 Mbits/sec. Correlation was done with 32 lags with an integration period of 5 seconds over a total observation length of ~30 seconds. As can be seen from the data printed along the right upper side, the correlation amplitude is ~40 (in units of 10^{-3}, with all known loss factors accounted for) with a signal-to-noise ratio computed to be ~153. The upper plot shows an overlay of the delay rate spectra (correlation amplitude vs. residual phase delay rate) and the so-called 'delay-resolution function' (correlation amplitude vs. residual multi-band group delay). The second row of plots shows the averaged single-channel correlation amplitude vs. residual delay on the left and the averaged cross-power amplitude and phase on the right. Below these plots are a set of plots of correlation amplitude and residual phase vs. time over the scan for each of the 8 frequency bands plus the vector-summed amplitude and residual phase of all 8 bands (labeled 'All'); a slight rolloff in correlation amplitude can be seen in the higher-numbered frequency bands corresponding to RF-receiver rolloff at higher frequencies. Immediately below these plots are plots showing the valid-data percentage (linear scale, 0-100%, USB/LSB shown separately) accepted in each integration period of each frequency band. Below this are plots of tape error statistics ('Tperr', log scale, range 10^{-6} to 10^{-2} parity error rate) and phase-calibration phase (-180 to +180 degrees) vs. time for each station in each frequency band.

Tabulated below the plots in Figure 14 are the detailed numerical results of the scan, including both *a priori* and measured values of many parameters, including channel-by-channel correlation amplitude and phase, multi-band group delay, single-band group delay, phase-delay rates, etc. Also shown are observed and theoretical statistical quantities, such as integration-period by integration-period amplitude and phase scatter, as well as similar channel-by-channel scatter statistics. This collection of statistics, along with the SNR, allows an automated 'fringe-quality code' to be assigned to this scan for use in further downstream processing. The scan in Figure 14 is assigned a fringe-quality code 9 (upper RH corner), which is assigned to scans which meet all assigned judgment criteria.

An important aspect of the Mark 4 processing philosophy is that the *a priori* model information used for the correlation processing is fully preserved in the correlation output files, as well as all residual observables derived from the fringe-fitting process. The availability of this *a priori* model information provides several capabilities that are difficult without it. For example, total observables are easily calculated, which are not only essential for most geodetic work, but also provide powerful diagnostics for correct operation of the correlator; for example, it becomes easy to test that total observable delay, delay-rate and phase are invariant to small changes in *a priori* model parameters, as they should be. In addition, software may be easily written to modify the observing epoch of a particular scan, which is sometimes necessary, or even to modify the *a priori* model (within the limits of the original field-of-view, of course) and recompute residuals

without re-correlation; the latter capability is sometimes important to re-process previously correlated data with more accurate models to reduce potential systematic errors arising from the original correlation model.

For geodetic VLBI, the high-precision group delay observable made possible with multi-band bandwidth synthesis is the fundamental observable for follow-on processing. For astronomical data, Mark 4 channel-by-channel correlation amplitude and phase data may be transferred to any of several astronomical image-processing packages such as AIPS [Greisen 1998], HOPS [Haystack Observatory 2002] or the 'Caltech' imaging package [Pearson 1991].

11. Summary

The Mark 4 correlator was designed as the first large-scale VLBI correlator to process data at 1 Gbps/station, and is the first to be developed as a major shared international effort. Targeted at both astronomy and geodetic interests, the system design includes the flexibility to accommodate a wide range of operating conditions and configurations. Among the new and innovative aspects of its design include:

- full-custom ASIC VLBI correlator chip with embedded rotators and delay management
- multi-application correlator board for diverse projects
- new efficient algorithms for handling vernier-delay management and fringe rotation
- station model embedded in data stream with extremely low overhead (<0.1% worst case)
- multiple phase-calibration-tone extraction as standard part of correlation processing
- support for simultaneous processing of multiple independent scans
- ready to support real-time/near-real-time VLBI observations with new Mark 5 data system

Four Mark 4 VLBI processors are now installed and operating in the world with two largely independent sets of operating software to help ensure correct operation. As more and more stations worldwide expand their observing schedules and data rates, the Mark 4 correlators stand ready for the processing challenge.

Acknowledgments

References

Aldrich, W.H. & Whitney, A.R., "The Haystack VLSI Correlator Chip", Haystack Mark 4 memo #226 (1993).

Bos, A., "A high-speed 2-bit correlator chip for radioastronomy", *IEEE Transactions on Instrumentation and Measurement*, 40-3, 591-595 (1991).

Bos, A. Aldrich, W.H. & Whitney, A.R., "The Haystack Correlator Chip", EVN doc #237,(1996).

Cappallo, R.J., "Recorder Controller Commands", VLBA Acquisition memo 238 (1999).

Carlson, B.R., Dewdney, P.E., Burgess, T.A. & Casorso, R.V., "The S2 VLBI Correlator: A Correlator for Space VLBI and Geodetic Signal Processing" in *Publications of the Astronomical Society of the Pacific*, **111**, 1025-1047 (1999).

Cooper, B.F.D., "Correlator with two-bit quantization", *Australian J. Physics*, **23**, 2-19 (1970).

Greisen, E.W., "Recent Developments in Experimental AIPS", *Astronomical Data Analysis Software and Systems VII*, **145**, 204 (1998).

Haystack Observatory, 'Haystack Observatory Postprocessing System' [Online document], Available http://web.haystack.mit.edu/haystack/hops.html (2002).

Haystack Observatory, 'Mark 5 VLBI Data System' [Online document], Available http://web.haystack.mit.edu/haystack/vlbisystems.html (2002).

Hinteregger, H.F., Rogers, A.E.E., Cappallo, R.J., Webber, J.C., Petrachenko, W.T. & Allen, H. "A High Data Rate Recorder for Radio Astronomy", *IEEE Transactions on Magnetics*, B, 3455 (1991).

Napier, P.J., Bagri, D.S., Clark, B.G., Rogers, A.E.E., Romney, J.D., Thompson, A.R. & Walker, R.C., "The Very Long Baseline Array", *Proc. IEEE*, **82**, 658 (1994).

NASA/GSFC, 'Mark-4 VLBI Analysis Software Calc/Solve' [Online document], Available http://gemini.gsfc.nasa.gov/solve (2001).

Pearson, T.J., "Caltech VLBI Analysis Programs", *Bull. Amer. Astron. Soc.*, **23**, 991-992 (1991).

Rogers, A.E.E, "Proposed phase-cal extractor for upgraded formatter A/D board", VLBA Acquisition memo 248/Mark 4 memo 176 (1991).

Rogers, A.E.E., "Digital tone extractor normalization", VLBA Acquisition memo 347/Mark 4 memo 133 (1993).

Schilizzi, R.T. et al, "The EVN-Mark IV VLBI Data Processor', Experimental Astronomy, **12**, 49-67 (2001).

Thompson, A.R., Moran, J.M., & Swenson, G.W. *Interferometry and Synthesis in Radio Astronomy*, Wiley (1993).

VEX, 'VEX File Definition/Example, Rev 1.5a' [Online document], Available http://lupus.gsfc.nasa.gov/vex/vex.html (1998).

Whitney, A.R, "A very-long-baseline interferometer system for geodetic applications", *Radio Science*, **11-5**, 421-432 (1976).

Whitney, A.R., "The Mark IV VLBI Data-Acquisition and Correlation System" in *Developments in Astrometry and Their Impact on Astrophysics and Geodynamics*, eds Mueller, I.I & Kolaczek, B., 151-157 (1993).
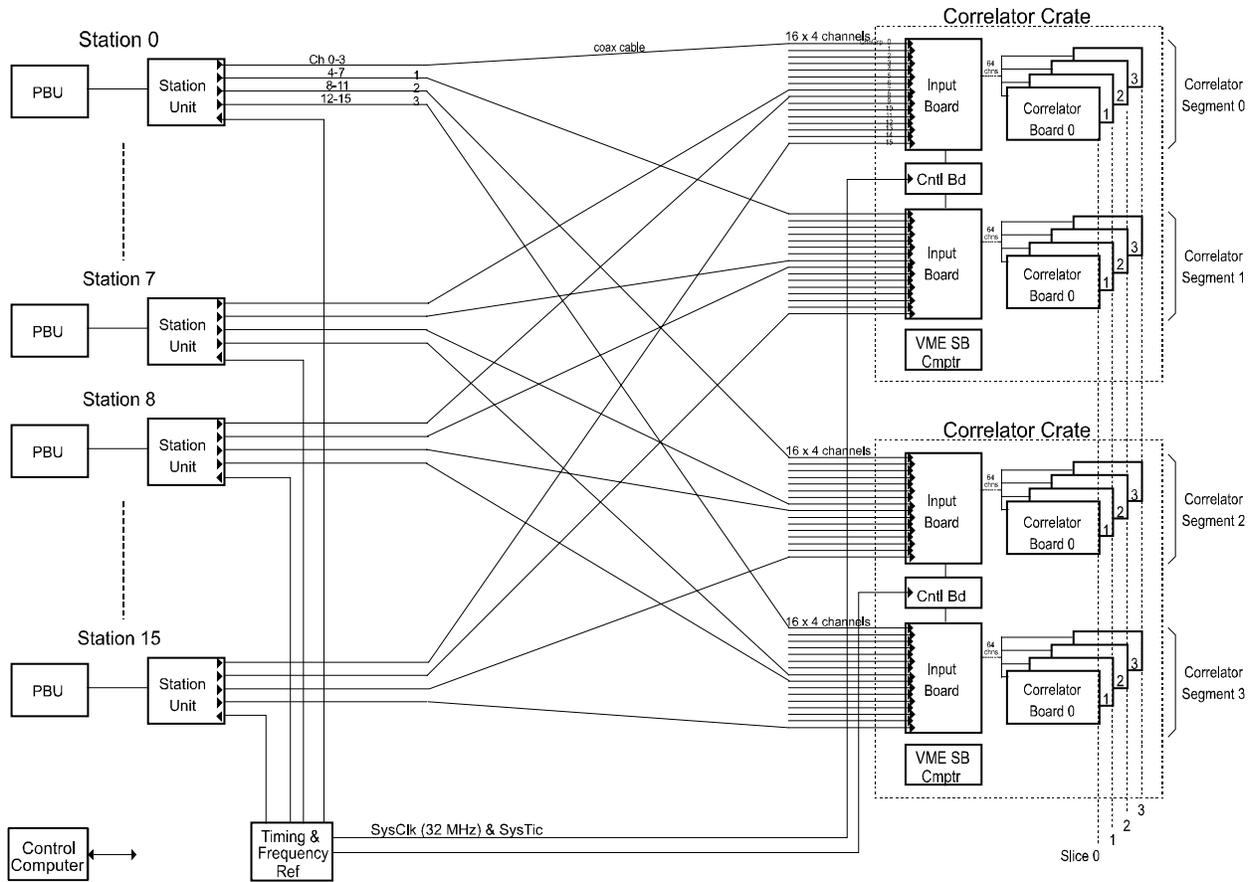
Figure 1: Correlator Block Diagram
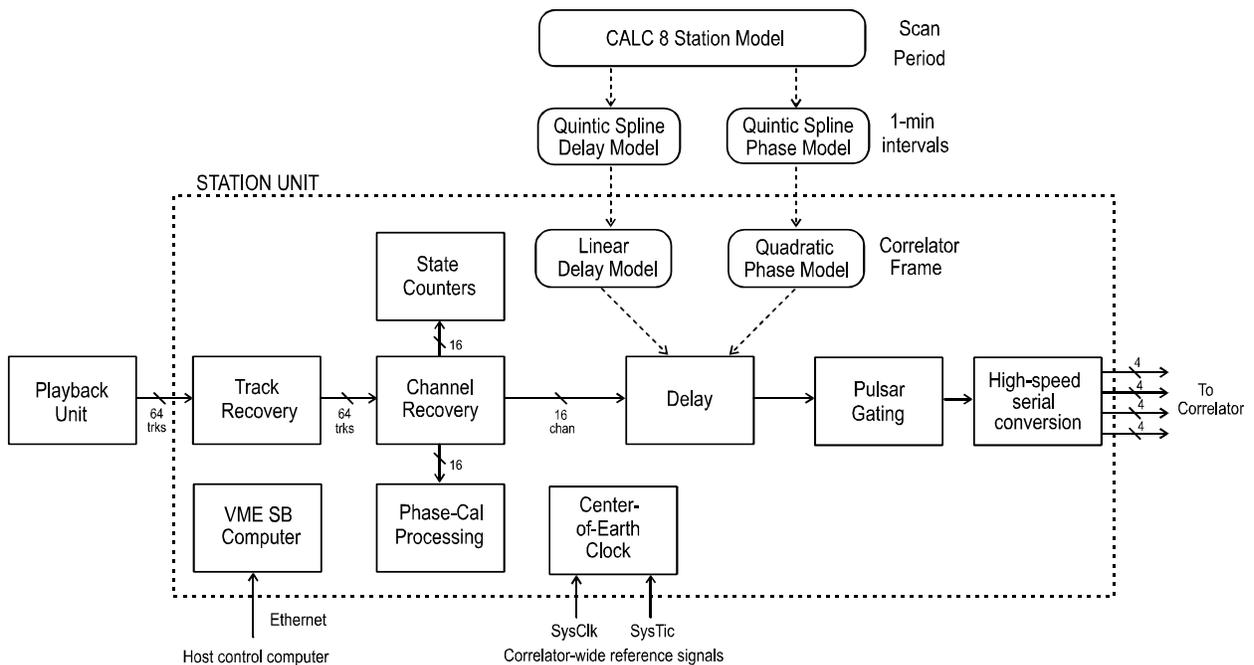
Figure 2: Station Unit Functional Block Diagram
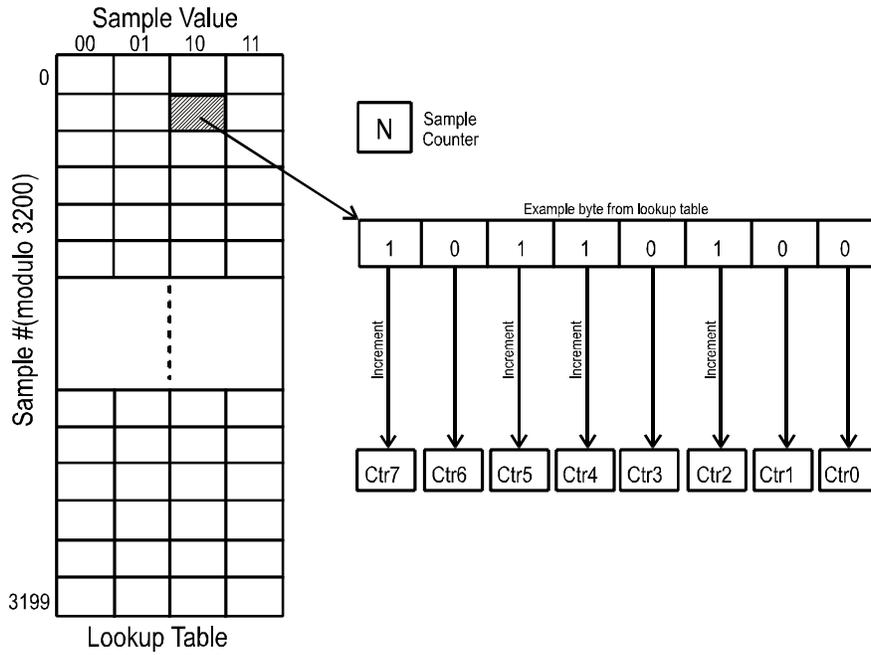
Figure 3: Schematic of Phase-Cal Tone Processing Algorithm
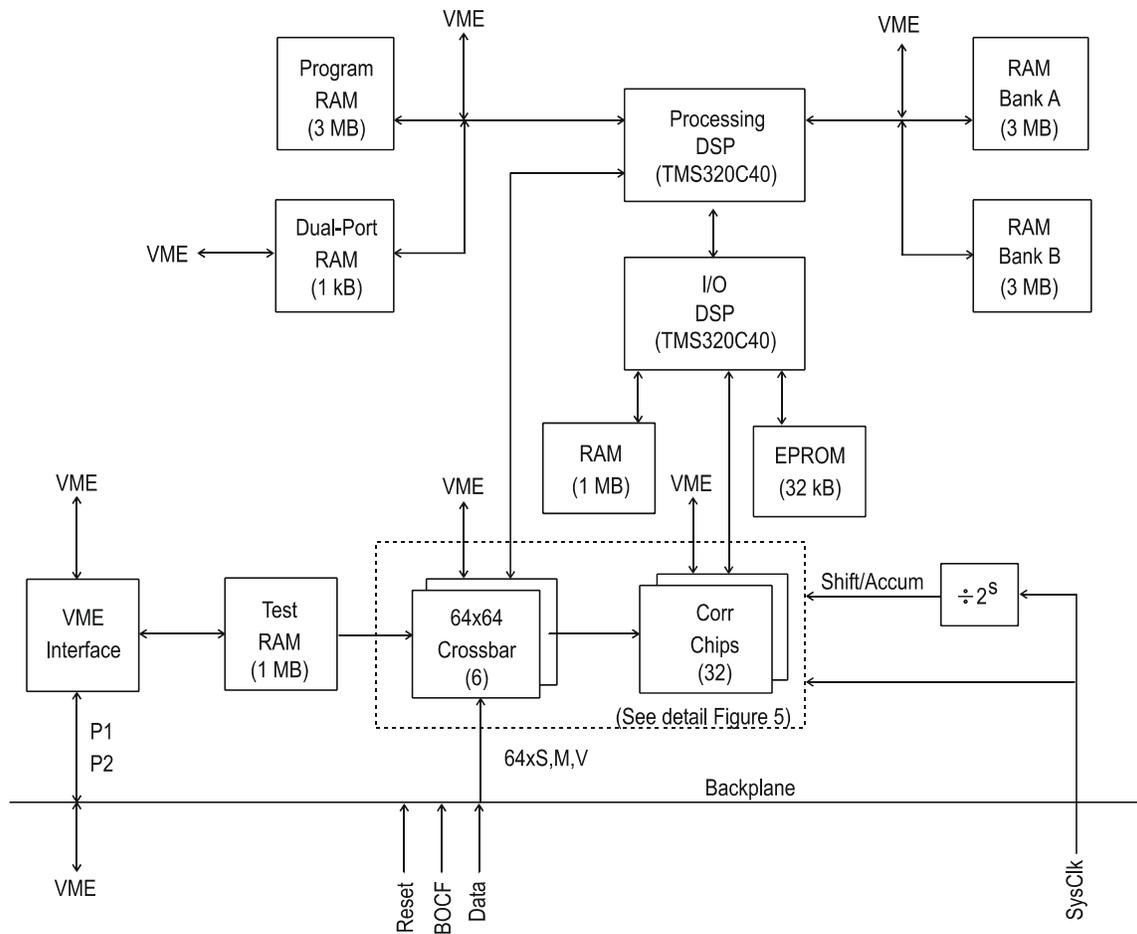
Figure 4: Simplified Block Diagram of Correlator Board
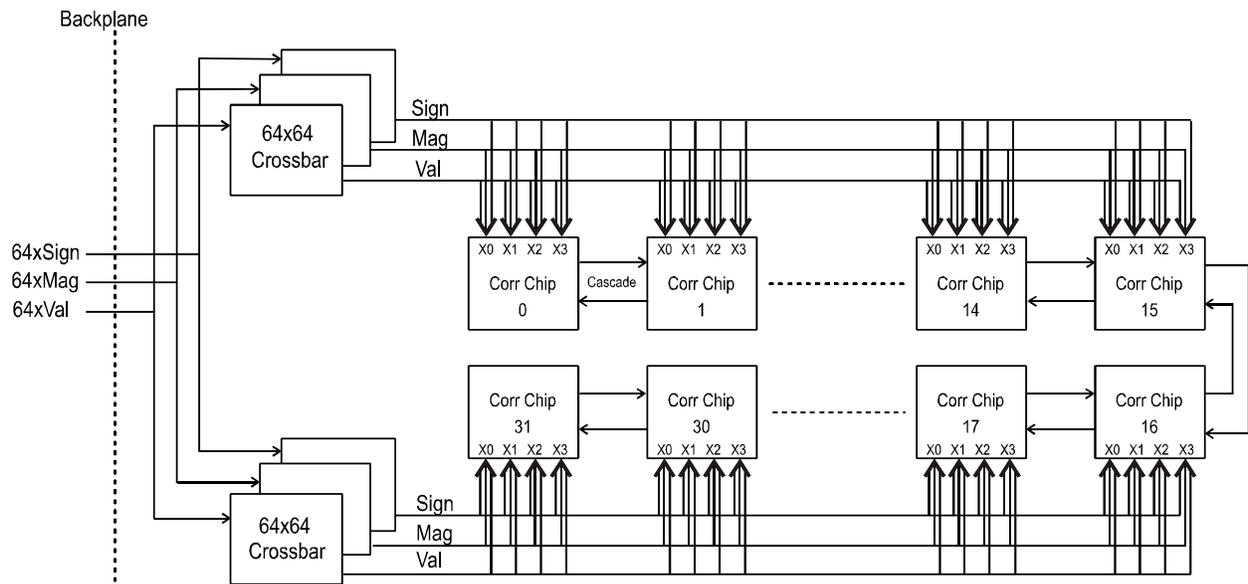
Figure 5: Correlator Board Signal Distribution
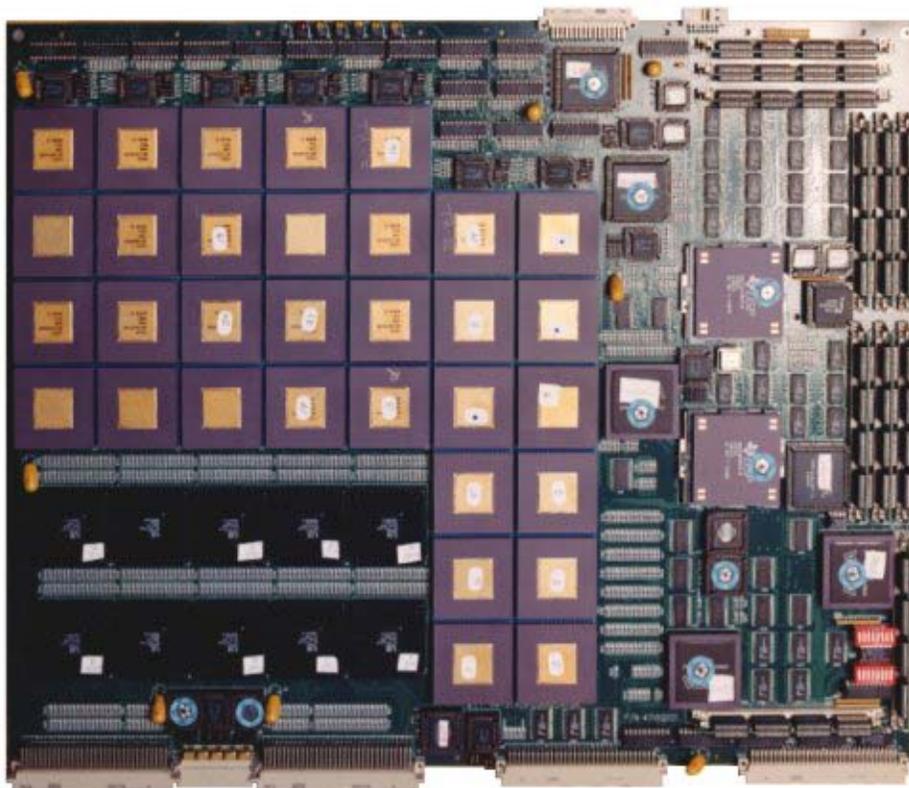
Figure 6: Mark 4 Correlator Board

29
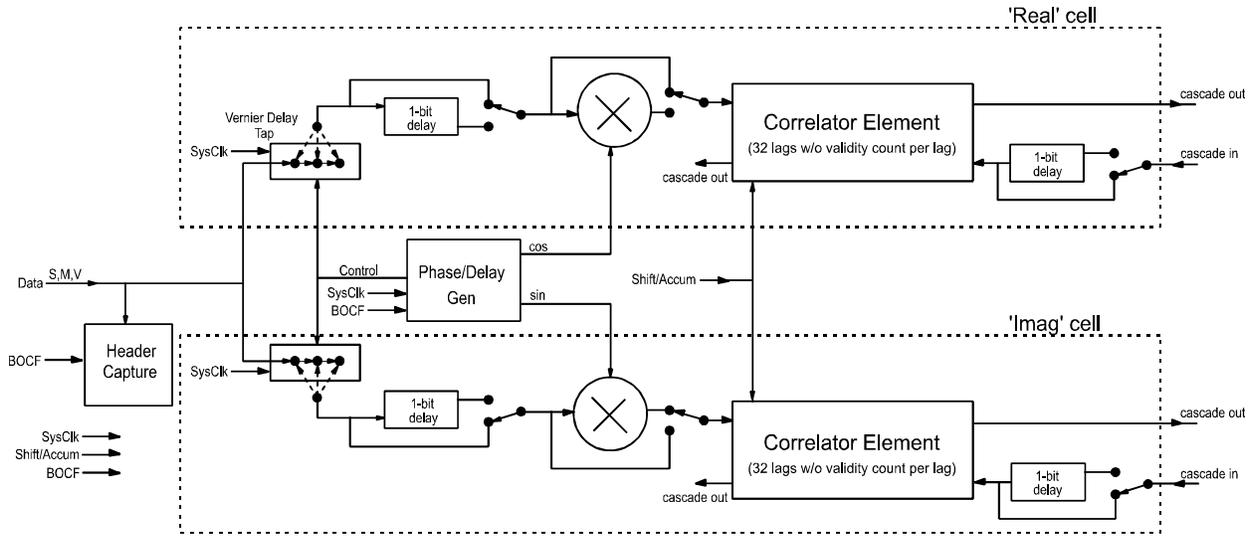
Figure 7: Correlator Block (1 of 8 on chip)
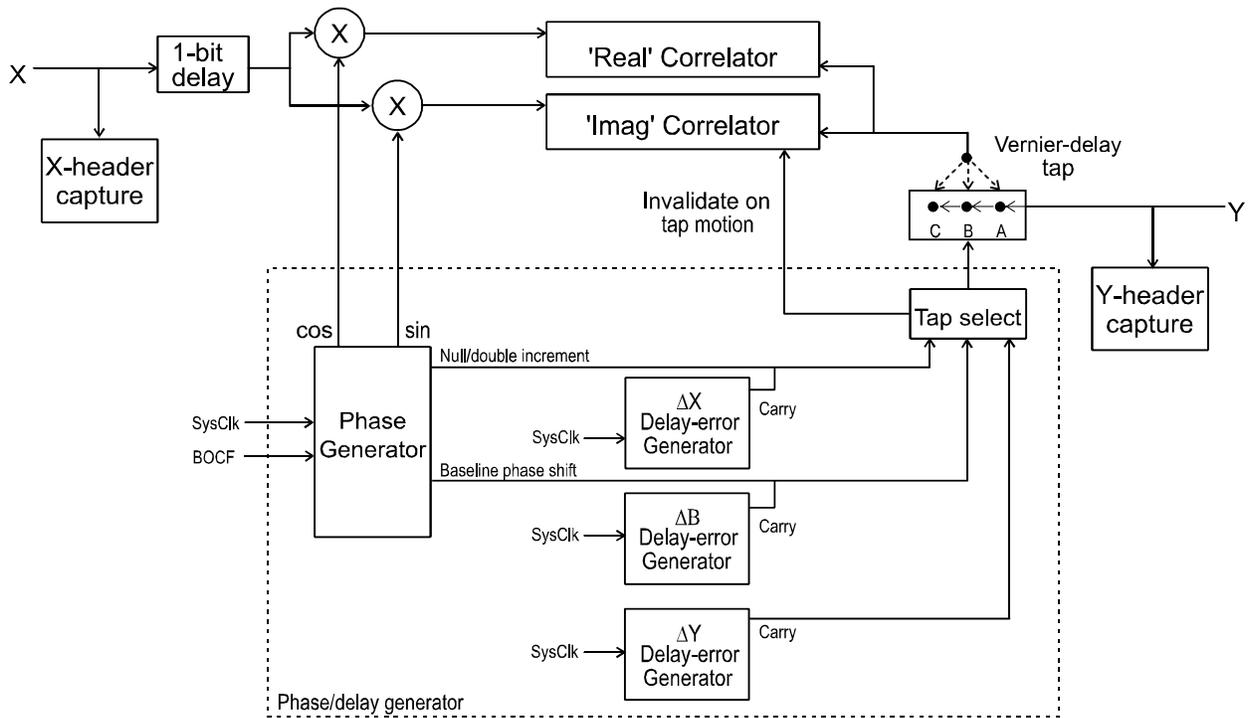
Figure 8: Complex Correlator Block Diagram
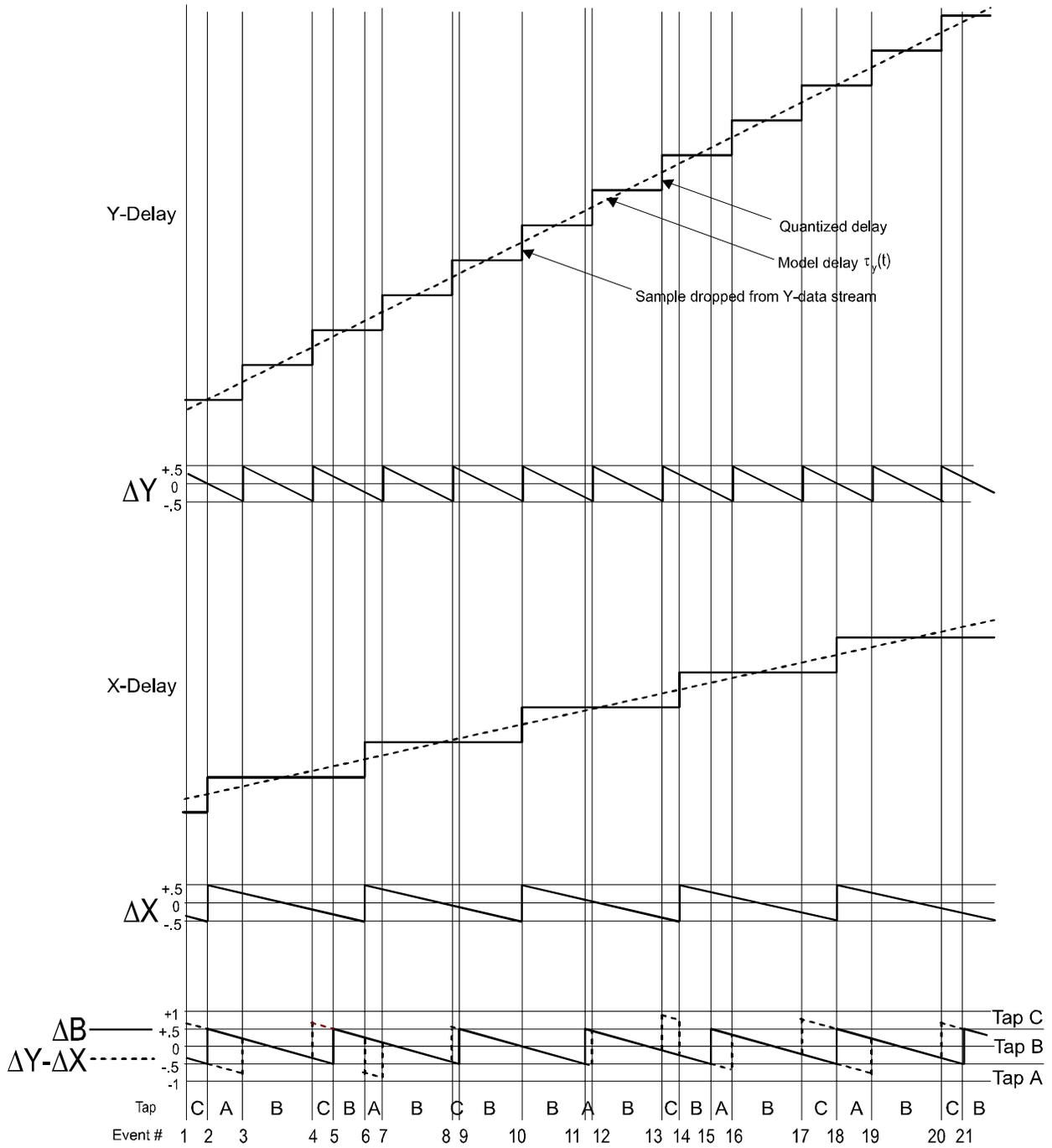
Figure 9: Vernier-delay tracking example
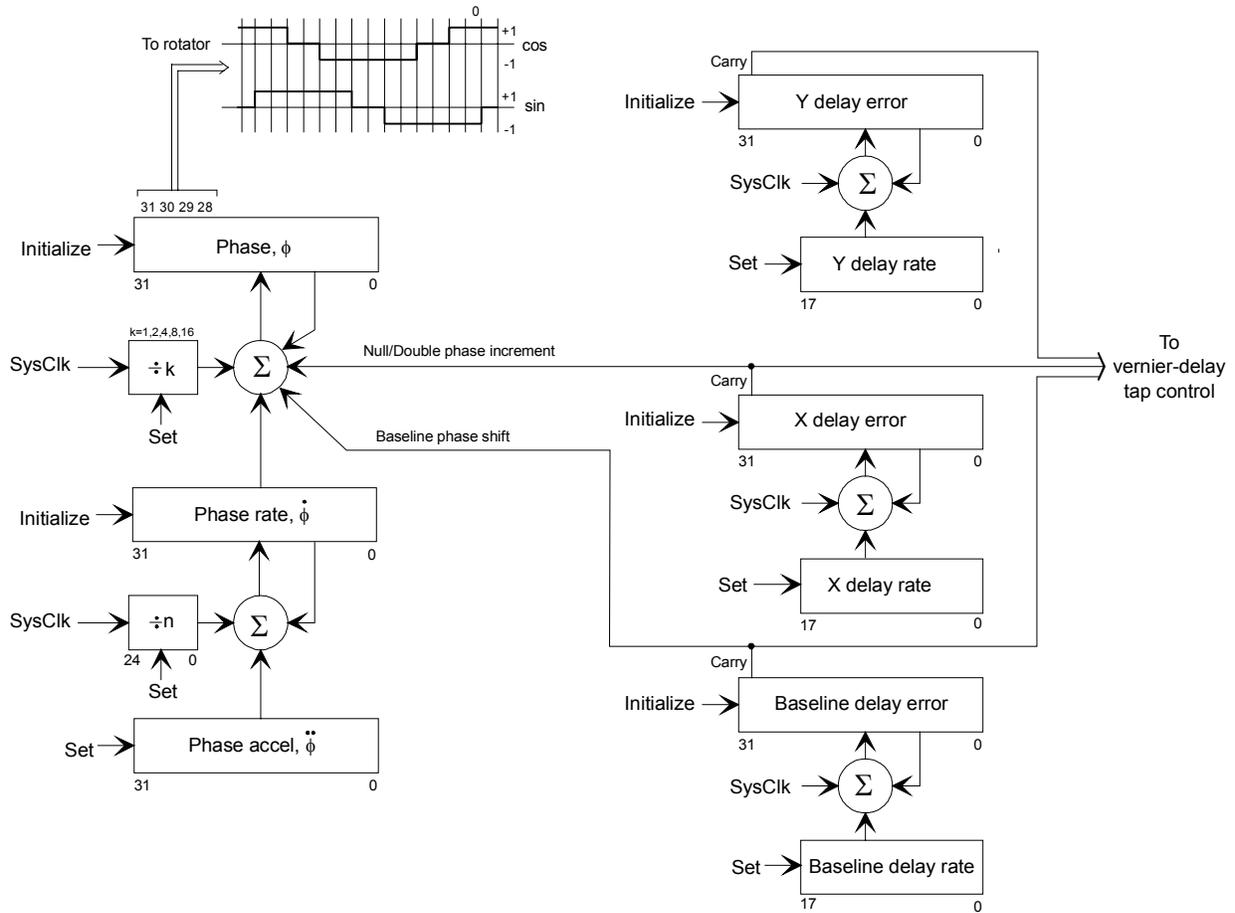
Figure 10: Phase/Delay Generator Block Diagram
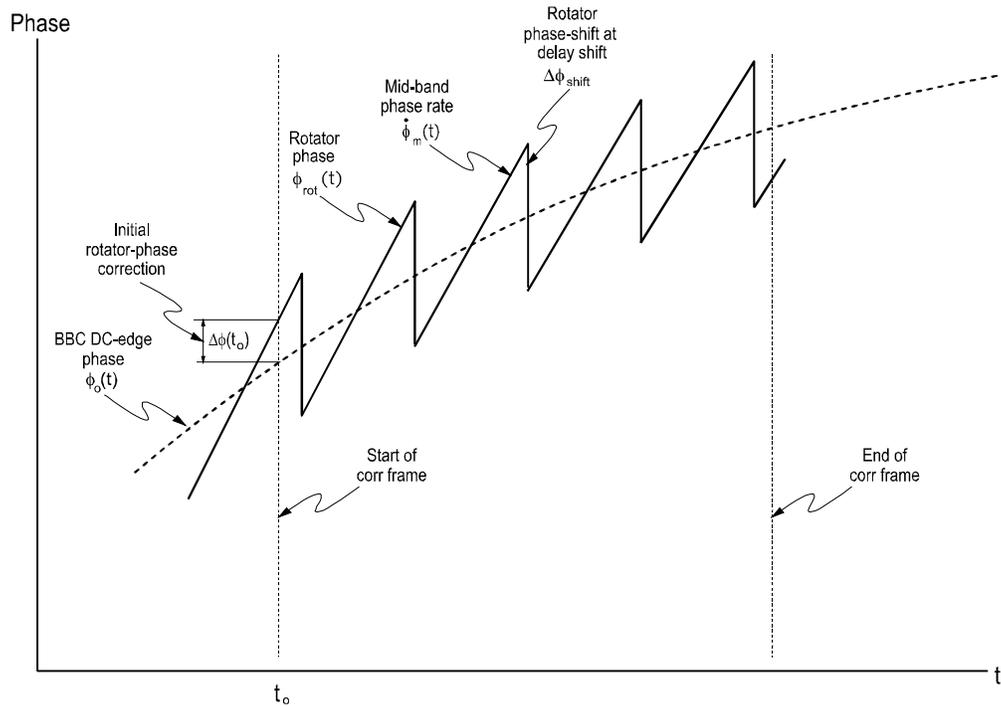
Figure 11: Rotator Phase During Correlator Frame

32
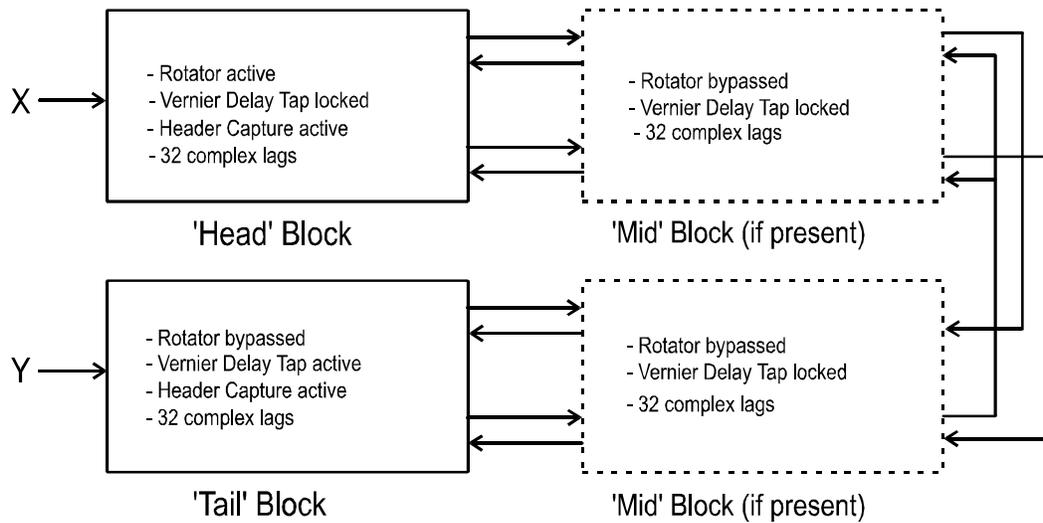
Figure 12: Block Interconnections for Complex Correlation
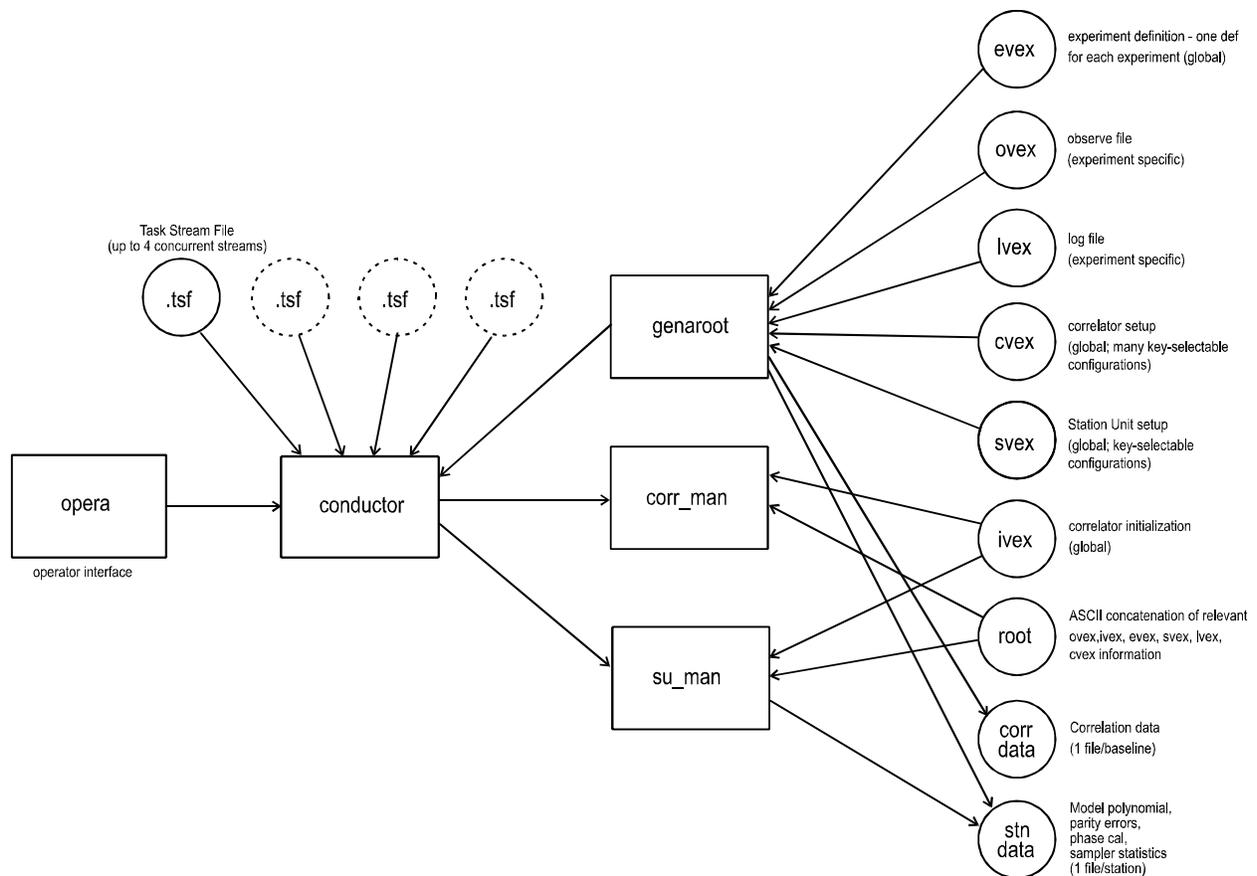
Figure 13: Simplified Mark 4 Correlator Software Block Diagram

Mk4 Fringe Plot

4C39_25.pehrto, 254-0927, Al
GILCREEK - MATERA, fgroup X, pol RR

Delay-rate spectra

Multi-band 'delay-resolution' function

multiband delay (µs)

delay rate (ns/s)

singleband delay (µs)

Avgd. Xpower Spectrum (MHz)

Fringe quality 9
SNR   153.1
PFD   0.0e+00
Intg.time 29.753
Amp   39.729
Phase   -42.3
Sbdelay (us)
           -0.010973
Mbdelay (us)
           0.007076
Fr. rate (Hz)
           0.002452
Ref freq (MHz)
           8212.9900
AP (sec)   3.000

Exp.   C1016
Exper #   2912
Yr:day   2001:254
Start   092742.00
Stop   092812.00
FRT   092757.00
Corr. date:
2001:307:010004
Fourfit date:
2001:320:211502
Position (J2000)
09h27m 3.0139s
+39Ê02'20.852"

Amp. and Phase vs. time for each freq., 10 segments, 1 APs / segment (3.00 sec / seg.), time ticks 2 sec

"g"   "h"   "i"   "j"   "k"   "l"   "m"   "n"   All

Phase
Amp

USB
LSB
Tperr   I (SU 1)   A (SU 5)
pcal θ

|								Freq (MHz)	All
8212.99	8252.99	8352.99	8512.99	8732.99	8852.99	8912.99	8932.99	Freq (MHz)	All
-42.9	-41.2	-41.4	-44.6	-41.1	-40.1	-43.3	-43.1	Phase	-42.3
42.5	48.0	49.8	42.9	40.6	38.2	30.2	31.3	Ampl.	39.7
32.4	32.5	32.7	32.7	32.9	32.7	32.6	32.8	Sbd box	32.6
10/10	10/0	10/0	10/0	10/0	10/0	10/0	10/10	APs used	
4010:4010	4010:4010	4010:4010	4010:4010	4010:4010	4010:4010	4010:4010	4010:4010	PC freqs	
112:90	-68:153	116:55	-172:113	-56:-147	-150:98	136:81	-79:-112	PC phase	
0:-2	0:5	0:-3	0:-1	0:1	0:1	0:-4	0:1	Manl PC	
37:51	38:50	36:50	34:47	41:49	41:47	37:47	38:45	PC amp	
X1U,X1L	X2U	X3U	X4U	X5U	X6U	X7U	X8U,X8L	Chan ids	A
2 4 6 8	10 12	14 16	18 20	22 24	26 28	30 32	35 7 9	Tracks	
X1U,X1L	X2U	X3U	X4U	X5U	X6U	X7U	X8U,X8L	Chan ids	I
24 6 8	10 12	14 16	18 20	22 24	26 28	30 32	35 7 9	Tracks	

	RMS	Theor.						
Group delay (usec)		-1.54819453774E+04	Apriori delay (usec)	-1.54819524538E+04		Resid mbdelay (usec)	7.07636E-03	+/- 3.7E-06
Sband delay (usec)		-1.54819634265E+04	Apriori clock (usec)	4.2336473E+00		Resid sbdelay (usec)	-1.09727E-02	+/- 2.3E-04
Phase delay (usec)		-1.54819524681E+04	Apriori clockrate (us/s)	-3.8999999E-07		Resid phdelay (usec)	-1.43009E-05	+/- 1.3E-07
Delay rate (us/s)		3.18842278626E-01	Apriori rate (us/s)	3.18841968447E-01		Resid rate (us/s)	3.10179E-07	+/- 1.5E-08
Total phase (deg)		-288.2	Apriori accel (us/s/s)	1.00484887849E-04		Resid phase (deg)	-42.3	+/- 0.4
	RMS	Theor.	Amplitude	39.729 +/- 0.259		Pcal mode: NORMAL, NORMAL		
ph/seg (deg)	2.9	1.1	Search (32X256)	37.971		Pcal rate: -7.037E-09, 4.651E-09 (us/s)		
amp/seg (%)	2.1	2.0	Interp.	37.971		Bits/sample: 1		
ph/frq (deg)	1.4	1.0	Inc. seg. avg.	39.772		Sample rate(MSamp/s): 16		
amp/frq (%)	16.7	1.7	Inc. frq. avg.	39.732				

Control file: ../cf_2912   Input file: /data1/2912/254-0927/Al..pehrto   Output file: Suppressed by test mode

Figure 14: Example fringe-search output