

Digital Signal Processing for the SKA: A Strawman Design

Larry R. D'Addario, NRAO¹ (email: ldaddario@nrao.edu)
Constantin Timoc, Spaceborne Inc. (email: ctimoc@aol.com)
2002 August 15

Abstract

It now appears that current (2002) technology allows construction of a cross-correlator large enough to compute visibilities on all baselines of an array containing 2000 to 5000 stations, and to do this with a substantial bandwidth (several GHz). It also appears that the present cost of such a correlator is affordable for the SKA, and it is reasonable to assume that the future cost (at the time of actual construction) will be far less.

This paper presents a strawman design that illustrates how such performance can be achieved. It also discusses other aspects of digital signal processing for a large array, including digitization, delay tracking, phase tracking, and signal transmission.

I. Introduction

Until recently, it appeared that the construction of a radiotelescope with thousands of antennas would be limited by the need to cross-correlate the signals on the resulting millions of baselines. This follows from extrapolating the costs of existing correlators or those under development, where tens to hundreds of antennas are correlated, using the facts that the computation required grows as N^2 for N antennas and that some elements of the structure, particularly interconnections, grow in number as N^3 . However, over the last two years several developments have occurred to change this conclusion. It now seems feasible, at a cost far less than that of the antennas, to construct a correlator for all baselines of several thousand antennas and for a bandwidth of several GHz, including spectral analysis of the visibilities to several thousand channels. This prediction of feasibility and cost is based on today's technology and prices, without use of Moore's law or other projections.

The factors that make this possible include:

1. Continuing advances in silicon processing now allow higher densities. This is the somewhat predictable investment-driven improvement that is the basis of Moore's law. Currently available 0.18 micron CMOS processes are the basis of our strawman design. Process improvements have also resulted in lower defect rates, allowing chips as large as 30x30 mm to be produced at reasonable yields. Our estimates use 20x20 mm chips. Note that the most recent correlator ASIC in astronomy, for the ALMA telescope, uses a 0.25 micron process to produce 8.3x8.8 mm chips [1].
2. For a project as large as SKA, use of full-custom ASIC design is easily justified. This means that the design is optimized at the transistor level. Full custom design takes longer and is more expensive than gate array or standard cell design, but it is necessary for achieving state-of-the-art circuit density. With few exceptions, previous ASICs for astronomical correlators have used the less costly techniques.
3. The system level architecture of a large cross-correlator has benefited from several new ideas and improved understanding of old ideas. This has happened in connection with development of the ATA and Australia telescopes.
 - a. The large *computational* advantage of the FX architecture over XF has been known for a long time [2], but there is no advantage in the number of accumulation registers needed, which is the product of baselines and spectral channels. By re-ordering the data, it is possible to move most accumulators to inexpensive RAM, enabling the ASICs to handle far more baselines [3]. In this way, the computational advantage of FX allows practical savings.

¹ The NRAO is operated by Associated Universities, Inc. under a cooperative agreement with the National Science Foundation.

- b. Heretofore, astronomical FX correlators have used segmented data FFTs to implement the spectral filter banks (F part of the correlator). This has led to inefficiency because of the need for overlapping of segments or zero padding so as to achieve good channel isolation. This can be avoided by using more sophisticated filter bank designs (sometimes called polyphase filter banks) that require only slightly more computation than the FFT alone [4][5]. In addition, there are FFT implementations that require only $K - 1$ samples of internal storage for a length- K transform [6][7], rather than the usual $K \log K$ samples; this produces a large saving in chip area per filter bank.
- c. An FX correlator need not use fine quantization of the signals during cross correlation in order to achieve high spectral dynamic range, contrary to earlier ideas. After the filter banks, the signals may be coarsely re-quantized without loss of accuracy, provided that the self-correlation of each channel is measured using the more finely quantized data. At large N , relatively few self-correlations are required. Our design uses 4-level quantization for the real and imaginary parts at the cross-correlator inputs. (It is also possible to use dynamic re-quantization, where the thresholds for each spectral channel are separately adjusted according to the signal power [5]. But this extra complexity produces no better accuracy and only a small increase in SNR.)
- d. The number of logically separate connections between the antenna-based part (F part, for FX) and the baseline-based part (X part) of a correlator goes as N^3 [8], so these interconnections can easily dominate the cost. To minimize this, the two parts should each be packaged into the minimum number of units, and no sample should be needed at more than one X unit. This requires that the organization of the samples be changed from having all spectral channels together and antennas separate, to all antennas together and channels separate. A large buffer memory is necessary to achieve this, but it is inexpensive.

Here we investigate the design of an FX correlator to form all cross correlations among $N = 2560$ antennas at a bandwidth of $B = 3.2$ GHz (in 4 channels of 800 MHz each) and a spectral resolution of 0.2 MHz. The design follows from the feasibility of two ASICs: the first implements 8 filter banks of 4096 channels each at 800 MHz sample rate, and the second implements an array of 80x80 complex multiplier-accumulators (CMACs) at 400 MHz sample rate. If these ASICs can be properly arranged and interconnected, the correlator will require $BN/(20 \times 400\text{MHz}) = 1280$ filter bank ICs and $BN^2/(2 \times 80 \times 80 \times 400\text{MHz}) = 4096$ CMAC ICs.

Sections II–IV present a strawman design for this correlator. Section V discusses related signal processing issues, including digitization, tracking, and signal transmission from the antennas. Section VII gives a cost estimate for all the signal processing electronics.

II. Strawman Correlator Design Overview

Figure 1 shows the overall signal flow of the digital processing, from digitization at the antennas through output of the visibility data to a storage archive. Here we concentrate on the central processing. This processing is organized as an FX correlator, where each signal channel is first analyzed into 4096 subchannels using a filter bank (F portion) and then each subchannel is cross-correlated against the corresponding subchannel of all other antennas (X portion). Signals from an antenna are in 800 MHz bandwidth channels, with two such channels coming from orthogonal polarizations. Polarization pairs are kept together so as to enable cross-polarization correlation, as discussed later. The central processor is then constructed in large sections, each of which handles one polarization pair of channels. The sections are independent of each other, and two of them are needed to process the 3.2 GHz of total bandwidth (four channels or two pairs) proposed here. It is straightforward to add more sections so as to expand the total bandwidth.

As shown in Figure 1, one processing section has three parts: a set of “F units,” a set of “X units,” and the interconnections from F to X. In the F units, every channel (two from each of 2560 antennas) is passed through a separate filter bank to produce 4096 subchannels; this is done with no change in aggregate sampling rate. In each X unit, cross correlations are computed among all antennas for a fraction of the subchannels. In both cases, “unit” is a package of electronics small enough so that its internal interconnections are short and straightforward. To minimize the F to X interconnections, the number of units should be minimized and thus each should accomplish as much processing as possible. In practice this means that a unit should fit on a single printed circuit board, or at most a few adjacent boards. The amount of processing that fits on a single board is obviously a function of technology. It is very important that the X units be organized as stated, so that each correlates all antennas; this avoids any signal connections among X units and allows each F unit

output to drive only one X unit. Each X unit then processes as many subchannels as its technology permits. As we will show, the design achieves 128 subchannels per unit so that 32 units are needed to process all 4096 subchannels.

Similarly, each F unit processes as many antenna channels as its technology permits. We will show that the design achieves 64 antennas (128 channels) per unit. Data from each of the 64 pairs of filter banks must then be re-organized for transmission to the 32 X units. On each cable, we send the data from all 64 channel pairs but from only 128 subchannels.

III. Cross Correlators (X Units)

We discuss the X units first because their architecture determines many requirements for the interconnections and for the F units.

A. Architecture

The design is essentially bottom-up, in that it follows from the feasibility of the chip-level implementation. The ASIC’s internal design is described in the next subsection (III.B); it has the following characteristics.

- 6400 CMACs arranged in a 80x80 array
- 4-level quantized signals at 2b+2b (real+imaginary) per sample
- 16b accumulators with readout of most significant 12b
- 400 MHz clock rate
- 80 bi-directional 4b I/O signals on each of 4 sides of the array
- I/O multiplexed by 16: 6.4 Gb/s per differential pair of pins
- four modes of operation, to support arrays of chips

To form one X unit, 64 of these ASICs are combined into a 8x8 array, allowing 640 signals per side. Each signal is distributed vertically or horizontally to 640 CMACs. The array is configured as shown in Figure 2. It is capable of correlating 640 “row” signals by 640 “column” signals at 400 MHz. By re-using the array 8 times, it can form all cross correlations among $640 \times 4 = 2560$ antennas. By re-using it an additional 8 times, it can do this for a second channel (opposite polarization) of the same 2560 antennas. The 16 times re-use means that the sampling rate for any one correlation is $400\text{MHz}/16=25$ MHz. Since the Nyquist rate is equal to the bandwidth for complex signals, this unit processes 1/32 of the 800 MHz channel bandwidth or 128 of the 4096 subchannels. Thus 32 units are needed to process all 800 MHz (dual polarization).

The 8 cycles of computation for each channel are listed in Figure 2a. When the row and column signals are both from the same set of 640 antennas, only half of the array is needed. To use it efficiently, the other half simultaneously processes data from a different set of 640 antennas. Two of the cycles use this mode, where the array is split along a diagonal and signals are input to all four sides. In the other 6 cycles, the row and column signals are from disjoint sets of 640 antennas, and signals are input to one vertical and one horizontal side. To minimize high speed interconnections and switching at the board level, the array supports three more modes with inputs on different pairs of sides.

To support re-use of the array so as to correlate all 2560 antennas, it is necessary to buffer the data. Figure 2a shows four buffers (Q1, Q2, Q3, Q4), one for each set of 640 antennas. The buffers hold 8192 samples from each of the two channels of each antenna, so the size of each is 40 Mb. (The reason for this number of samples will be discussed shortly.) As shown in the cycle table, each buffer is re-read 4 times during the 8 cycles. Since the two channels (x and y in Figure 2a) are processed separately, each can be written with new data while the other is being read.

One cycle consists of correlating 8192 samples from the selected buffers and reading the 640x640 complex results out of the array. The number of samples is determined by the readout time. For each ASIC separately, there are 80x80 complex results, and we assume that each can be read out as a parallel data word in one clock cycle. Actually, it is necessary not only to read out each result but also to add it into an accumulator that supports longer integration. The longer accumulator is implemented in a (commercial) RAM associated with each ASIC. On each clock cycle, one CMAC’s accumulation buffer on each ASIC is driven onto a chip-wide bus and added to a (complex value) word that has been read from the RAM; the sum is then written back to the RAM at the same location. RAM devices that can do this at the required speed are now available from several manufacturers (so-called “QDR” architecture). It thus takes at least 6400 clock cycles to read out all CMACs of each ASIC; the input buffer size is chosen to ensure that 8192 cycles are available. The

accumulation results for the next 8192 samples are later added to the corresponding RAM locations, building up longer integrations; each individual CMAC's accumulator can be copied and cleared in one cycle, so no sample is lost and the long integrations are continuous.

The order of the incoming data must be such that all samples of each 8192-sample block are from the same subchannel. (This has a strong affect on the architecture of the F units, as we shall see later.) The next block is then from the next subchannel, continuing until all 128 subchannels of this unit have been processed, and then repeating from the first subchannel. The results from each subchannel are accumulated into a separate location in the RAMs. The required RAM size is then 819,200 complex words. The word size depends on the maximum integration time before data is dumped to off-board memory. Using 32b+32b allows about 5 seconds before overflow. In that case, the size of each RAM is 52.4 Mb, or 105 Mb if double-buffering is needed. There are 64 of these RAMs per unit, one for each ASIC.

Cross Polarization Products. In addition to correlating all baselines for each polarization channel separately, it is sometimes necessary to form the cross-polarized correlations: xy and yx . This can be done at half the bandwidth by re-using the same hardware and the same data. To support this, each buffer holds the x polarization channel of one set of 640 antennas and the y polarization of a different set. During successive sets of 8 cycles, the products xx , yy , xy , and yx are each computed. In this case each unit supports 12.5 MHz of bandwidth in 64 subchannels, and the aggregate for all 32 units is 400 MHz in 2048 subchannels.

Quantization. The X units operate with signals quantized to 4 levels. In the past it has been thought that this provides insufficient dynamic range, since one subchannel may carry a total power substantially different from that which provides optimum SNR, even when the average among subchannels is near optimum. It is now understood that this need not be the case. First, small departures from optimum (a few dB, such as might be caused by non-flat receiver gain or gain setting tolerances) cause very small loss of SNR and can be neglected. Second, if the no-signal level is set near optimum, then large departures for one subchannel are generally in the high power direction, being caused by the presence of a strong astronomical line or of interference; then the SNR is very high, so the fact that it is less than optimum is unimportant. The real issue is not SNR, but rather accuracy. The true cross-power is a non-linear function of three measurable quantities: R_{ab} , R_{aa} , and R_{bb} , which are respectively the cross-power and the two self-powers in the quantized signals. When the quantization is coarse and the signal level departs substantially from optimum, the quantized self-powers R_{aa} and R_{bb} become weak functions of the true self-powers, and this is what limits dynamic range. This can be avoided by measuring R_{aa} and R_{bb} using more finely quantized signals, even though the measurement of R_{ab} uses signals that are rounded to much coarser quantization. In a large array, the cost of this is small because the self-powers are measured only once for each antenna channel, while the cross-powers are needed for each baseline.

B. ASIC Implementation

We begin by assuming the following ASIC characteristics, based on current (year 2002) technology:

- silicon CMOS process with 0.18 micron resolution
- die size 20x20 mm (limited by yield)
- clock rate 400 MHz (limited by power dissipation)
- maximum pads around 1600; maximum I/O pads around 1300.

The elementary computational unit is the CMAC, for which a block diagram is given in Figure 3. It requires 4 real multipliers and two accumulators. Each multiplier operates on 4-level (2 bit) samples; complexity increases rapidly for finer quantization, and (as explained above) 4 levels are sufficient for cross-correlation if each signal's self correlation is computed from more finely quantized samples. Following Cooper [9], we use the weighting and encoding of Figure 4, resulting in the minimized logic of Figure 5. Each product is a 3-bit number in offset binary. Two such multipliers and a 4-bit accumulator are combined in a pipelined structure as shown in Figure 6, and further accumulation is accomplished by counting overflows from the MSB. Each counter (not shown) is 12b long; it can be copied to a 12b readout buffer and cleared in one clock cycle. Two of these dual-multiplier/accumulators are needed for each CMAC, one for the real part and one for the imaginary part of the result.

It is our estimate that as many as 10^4 CMACs of this type can be implemented on a 20x20 mm chip, along with I/O circuitry to be described shortly. This comes from the fact that one CMAC requires about 1200 transistors or 300 gates, leading to a chip with 3 million gates, which is about the size of a Pentium 4, or 20x20 mm. Since this estimate is very preliminary, we base the design on only 6400 CMACs per chip

and we choose to arrange them as an 80x80 array in order to support the board level architecture described above. The topology is shown in Figure 7. In most modes, all CMACs in a row or column are driven by the same signal, and that signal is also passed to an output on the opposite side of the chip so as to drive the same row or column of the next chip in the array. All signal ports are bi-directional, so the horizontal and vertical signal flows can be in either direction. For those chips that are on the diagonal of the larger array, another mode is needed in which the chip's array is split along its diagonal and all sides are inputs.

The chip's CMAC array requires 80 4-bit I/O signals along each chip edge, or 1280 bitstreams at 400 Mb/sec. If each of these were connected through a separate pin, the number of CMACs would have to be reduced in order to keep the pin count feasible — the chip would be pin limited. To avoid this and to make maximum use of the chip area, we choose to multiplex the signals to a higher bit rate. Multiplexers reaching 10 Gb/s have been shown to be feasible in this technology [10-11]. Our design uses 16x multiplexing for an I/O rate of 6.4 Gb/s. (We considered another approach to limiting the pin count. Whereas each sample will be re-used 4 times, the pin count could be reduced by that factor if the input buffers were on-chip. But the memories would consume far more chip area than the multiplexers, so the multiplexing approach was selected.) The chip requires both multiplexers and demultiplexers along each side in order to support all of the input modes. When data is input along only two sides, each signal is copied to an output on the opposite side for passing to the next chip of the array.

Figures 8 and 9 show block diagrams of the 16-way multiplexers and demultiplexers, which are made up of the 2-way multiplexers and demultiplexers shown in Figures 10 and 11, respectively.

C. Packaging and Power Dissipation

The core of the ASIC contains 6,400 CMACs or a total of 1.92 million gates (7.68 million transistors). It is estimated that this core dissipates ~30 W based on the fact that a microprocessor with about 8 million transistors operating at 400 MHz (e.g., Pentium III) dissipates this amount. In addition to the CMACs, the chip contains 80 multiplexer/demultiplexer circuits. Measurements show that the power consumption for one multiplexer is 130 mW at 10 GHz [10], which should become about 80 mW at 6.4 GHz. For the demultiplexer, power dissipation has not been reported, but it should also be about 80 mW at 6.4 GHz because it contains about the same number of transistors as the multiplexer. Total dissipation of the 80 multiplexer/demultiplexer blocks is thus about 12.8 W. Therefore, the total power dissipation of the CMAC chip is approximately 43 W. Although air cooling at this level of power dissipation is possible, liquid cooling is preferred since it results in higher packaging density at the board level.

The cross-correlator board will contain an array of 8x8 ASIC chips plus accumulation RAMs on an area of about 50x50 cm² and it will dissipate ~2.76 kW. Although it appears difficult to maintain the junction temperatures below 85 C on a board with such a high power dissipation, it is practical to do so with liquid cooling. A power dissipation of 2 kW has been reported on an area of only 13x13 cm² [14]. The current necessary for the cross-correlator board is about 1,500 A and this will require a 20-layer printed circuit board of which ~15 layers are allocated to power distribution and 5 layers to signal routing. This estimate is based on a similar printed circuit board used in IBM computers [15].

The power consumption of the remaining components shown in Fig. 2b is estimated at ~500 W. These support the required input buffering and include 40 optical receivers at 12.8 Gb/s, 3 W each; 40 demultiplexers from 12.8 Gb/s, 3 W each; 640 demultiplexers from 800 Mb/s, 0.2 W each; 640 RAMs (32K x 16b), 0.1 W each; 80 multiplexers to 6.4 Gb/s, 0.3 W each; and 320 2x2 matrix switches at a data rate of 6.4 Gb/s, 0.15 W each.

A preliminary packaging concept is to place half of the input circuitry on each of two separate boards, one on each side of the main (CMAC) board, with the high-speed signals connected by short vertical pins (either differential or coaxial) and not via a backplane. The entire X unit thus becomes a 3-board stack. Each of the input boards would dissipate 250 W, which is manageable with air cooling.

IV. Spectroscopic Filter Banks (F Units)

A. Architecture

A block diagram of one F unit is shown in Figure 12. It accepts 128 input signals (2 polarization channels from each of 64 antennas) at 800 MHz bandwidth and 6b+6b resolution and analyzes each into 4096 subbands using polyphase filter banks (PFBs). To minimize the chance of overflow with signals of high spectral dynamic range, the quantization increases within the filter banks and becomes 16b+16b at

their outputs. For each subchannel, the squared magnitude of each finely-quantized sample is computed and accumulated, forming the self-correlations that will eventually be needed to evaluate the cross-correlation results, as explained earlier. The samples are then re-quantized (rounded) to 4 levels ($2b+2b$), and all further processing uses these coarsely quantized samples.

The PFBs produce no change in aggregate sampling rate, and the output samples appear sequentially by subchannel; i.e., a PFB generates one sample for each of the 4096 subchannels, followed by the next sample for each subchannel, etc. The re-quantized sample stream is then sorted into 32 groups of 128 subchannels each, so that each group can be routed to a different X unit. Next, the corresponding subchannel groups from all 64 antennas and 2 polarization channels are multiplexed together, since they are all destined for the same X unit. This produces a sample stream that varies fastest by antenna, then by polarization channel, then by subchannel, then by sample number. As explained earlier, the X units need to process blocks of 8192 samples from the same subchannel before proceeding to the next subchannel. It is thus necessary to re-arrange the order of the samples. This is done by using a dual-port RAM as a corner turner (CT). The CTs can logically be located anywhere between the PFBs and the X units; in the actual implementation (section III.B below) they are moved to a convenient place inside the F unit.

Figure 12 shows, for each stage of signal processing, the ordering of the samples in one logical sample stream as well as the aggregate bandwidth of that sample stream. The data rate per bitstream is also shown; this depends on the number of bits per sample and the degree of parallelizing, which vary through the processing.

B. Implementation

A difficulty with the topology of Fig. 12 is the large number of low speed interconnections (4096 sample streams) between the DEMUX and MUX stages. This is made feasible in the implementation by splitting the MUX into two parts, implemented in different chips, so that fewer and faster connections are needed between chips. This is shown in Figure 13.

The implementation of Fig. 13 includes two different ASICs, one of which is dominated by the polyphase filter banks (“PFB chip”) and the other of which contains most of the multiplexers (“MUX chip”). It appears that a PFB chip can be built to handle 8 of the 128 antenna channels (limited by logic density), so each F unit requires 16 of them; and that a MUX chip can handle 8 of the 32 output streams (limited by pin count), so 4 are required. Further details of the internal structure of the PFB chip is given in Figure 14.

The corner turners are implemented in RAMs that are placed between the two ASICs (see Fig. 13), where the data rate is convenient for use of commercial RAM devices. Out of each PFB chip we have 32 parallel sample streams of 4 bits each, where each stream carries data sequentially from 128 subchannels of 4 antennas and 2 polarizations; we call this “subchannel:antenna:polarization” order. We write 8192 of these data blocks to RAM ($8192 \times 128 \times 8 = 8,388,608$ samples) and read them out re-ordered as polarization:subchannel:sample:antenna. Groups of 8 sample streams are sent to one MUX chip as 32 bitstreams. The reordering of one such group is handled by a RAM of size $32b \times 8,388,608$ words (64 megabytes).

The MUX chip then combines the corresponding streams from the 16 PFB chips to form sample streams containing 16 antennas for sending to the X units. Finally, 4 groups of 16 antennas and the 4 bits of each sample are multiplexed together for transmission on single optical fiber at 12.8 Gb/s. In current technology, this last multiplexer probably needs to be a separate (commercial) IC, but in the near future it should be possible to include it in the MUX chip. The MUX chip’s logic is very simple but it is pin-limited. Whereas we require only 320 of them in the full system, it may be reasonable to implement them using FPGAs rather than ASICs.

C. Polyphase Filter Banks

The conclusion that PFBs for 8 channels can be implemented on a 20x20mm ASIC is based on preliminary PFB designs that have been done for the AT [12] and the ATA [7], where it was found that one 4096-subchannel PFB for 200 MHz bandwidth can be implemented in a Xilinx XC2V3000 FPGA at 200 MHz clock. This FPGA is about the same size as our proposed ASIC, but the ASIC is expected to achieve much higher logic density by virtue of transistor-level design and lack of general purpose routing. We estimate a factor of 20 improvement (see, e.g., [13]), so the ASIC should support 20 such PFBs. We also assume that the ASIC can operate at 400 MHz clock rate, so that a pair of PFBs operating in parallel is needed to handle the 800 MHz sample rate. This allows us to handle 10 input channels per chip, but since this estimate is very preliminary and also to allow for some additional logic, the design implements only 8 channels per chip.

An efficient PFB implementation consists of an FIR filter followed by an FFT (details are beyond the scope of this paper, but may be found in various textbooks including [4]). To handle the 800 MHz sampling rate at 400 MHz clock rate, the structure of Figure 14 is used. The input stream is demultiplexed so that two consecutive samples (“even” and “odd”) are available simultaneously, and these are processed by separate 2048-subchannel PFBs. The two FFT outputs are then combined in pairs by an additional butterfly stage so as to produce an equivelent filter bank of twice the length. Two subchannels are produced simultaneously, one from the upper half of the band and one from the lower.

D. Packaging

Packaging of the F units has not yet been worked out, but it is clearly less difficult than for the X units. Power dissipation should be less, and the circuit topology allows splitting into two or more boards in any of several ways. A single, large board per unit is preferred.

V. F To X Interconnections

As indicated in Fig. 1, there are $40 \times 32 = 1280$ signal paths from the F units to the X units, each of which carries 12.8 Gb/s. The length of these paths is in the 1 to 10 meter range because they are between racks in the same room. It cannot be less because, in current technology, not all of one 800-MHz segment of the system can fit in one rack. It need not be more because each segment should easily fit in a few racks. Since the structure of the system changes drastically between the F and X units (from antenna-oriented to baseline-oriented), and since the signals need to be sorted, this is the place in the signal path where separation via cables is appropriate.

Our strawman design serializes all data for one signal path into a single bit stream and transmits it via optical fiber on one optical carrier. This minimizes the parts count and cable count, and it is feasible in current technology. However, there are alternatives that might be less expensive at present. Each path might be implemented as 4 or 8 bitstreams at lower speed so as to use components (especially the high-speed multiplexers, demultiplexers, optical transmitters and optical receivers) that are cheaper by factors larger than 4 or 8, respectively. Due to the short distance, the cost of fiber is not significant so a separate fiber can be used for each bitstream. On the other hand, the higher parts count may make packaging difficult, resulting in a higher overall cost in spite of large savings at the component level.

Finding the optimum arrangement requires a more detailed design than we have accomplished so far. As shown in Section VII below, this is important because the cost of the digital and photonic components to support these interconnections is nearly as much as the cost of the components that accomplish the processing.

VI. Other Parts Of The Signal Processing

Here we briefly discuss other components of the signal processing shown in Fig. 1.

A. Digitization

We assume that the signals are converted from IF to baseband with a quadrature downconverter, obtaining what may be regarded as the real and imaginary parts of a complex signal. These are both digitized, and the result is treated as complex in all processing. Sampling of each channel is at 800 MHz and quantization is to (6,6)bits (real, imaginary). This is cost effective with currently available ADCs. The Nyquist bandwidth is then 800 MHz, and we support 4 such channels at 2 per polarization. Most of the quantization is intended to provide headroom against interference; the system noise level will normally be set for 2b quantization.

B. Tracking

For each channel, delay tracking is implemented with a FIFO to a resolution of 1 sample and by an FIR interpolating filter for finer resolution. For the FIFO memory, 120 MB per channel allows tracking over the whole sky at 3000 km from the array reference point. The other processing is implemented via FPGAs.

C. Phased Array Summation

To keep the signal transmission cost within acceptable limits, it is expected that some antennas of the SKA will be organized into stations, and for each station only the phase- and delay-tracked sum of the antenna signals will be brought to the center. The signal summation represents a very small amount of processing whose cost can be neglected. After summation, 8b+8b per sample are retained to avoid loss of precision. An optional enhancement would be to compute several phased array sums per channel per station, with the phased array beams pointing to different parts of the primary beam. This requires duplicating the trackers and adders for each additional beam. The station-to-center bandwidth would also be multiplied by the number of beams. The design presented here supports a single beam per station.

D. Signal Transmission

We need to transmit 38.4 Gb/s from each antenna. This fits nicely into four OC192-rate optical channels, one for each signal channel, with about 3% available for formatting overhead. Each optical channel is on a separate fiber to avoid the extra cost of optical multiplexing components. For antennas that are grouped into stations, transmission links are needed from each to the summing point as well as from the summing point to the center. For the latter, 6 optical channels are needed because of the finer quantization. Since these stations are likely to be the most distant, we use a single fiber for all six channels via wavelength division multiplexing, increasing the cost for components but reducing the fiber cost substantially.

VII. Cost Estimate

Table 1 presents a preliminary cost estimate for this design. It is broken down to the level of major components, including the ASICs and other large ICs. Many details are omitted, but these are not expected to affect the cost by more than a few percent. All components are available now or can be fabricated in current technology, and the cost of each is our best estimate for purchases made in CY2002 in the quantities needed for this instrument.

Part B of Table 1 lists the costs at the subsystem level, where it can be seen that the dominant item is the signal transmission from the stations to the center. The central processing hardware, which is the main subject of this paper, accounts for about 12% of the total. All of the items listed (except development work) have a cost proportional to the total bandwidth (3.2 GHz), so it is clear that the bandwidth is limited by the cost of signal transmission and not by the cost of computation. It would be straightforward and relatively inexpensive to increase the bandwidth of the central processing by adding 1.6 GHz sections, since each operates independently.

The cost shown for signal transmission includes only the electronic and photonic components required, not the optical fibers nor their installation. We estimate that each 10 Gb/s link will cost \$4000, including digital multiplexer, laser, modulator, photodetector, clock/data recovery, and demultiplexer. The receiving-side components are actually part of the F units, but their cost is counted as part of signal transmission. In keeping with the configuration of the U.S. stawman design [16], 160 outer stations are actually groups of 13 antennas phased together. The links from the summing points to the center are estimated at \$7000 each in order to include the optical multiplexing, demultiplexing, and amplification needed to put all optical channels onto one fiber. There are 2320 inner stations (single antennas), making 2480 stations altogether; the capacity of the central processing is slightly larger at 2560 stations.

Digitization is another high cost, and our estimate is based on using commercial ADC devices. The sampling rate is chosen for cost effectiveness with currently available parts.

The tracking cost of \$670 per channel is based on an FPGA implementation extrapolated from designs for other telescopes, and it assumes that several channels can be implemented in one module.

Part C of Table 1 provides a breakdown of the central processing costs. Again we have counted the cost of the internal connections (F to X) separately, even though they are physically included in the F and X units. This enables us to see that the interconnections account for a large fraction of the cost of the central processing. Each link is cheaper than a similar-speed antenna link because the short distance allows the use of directly-modulated, short wavelength lasers.

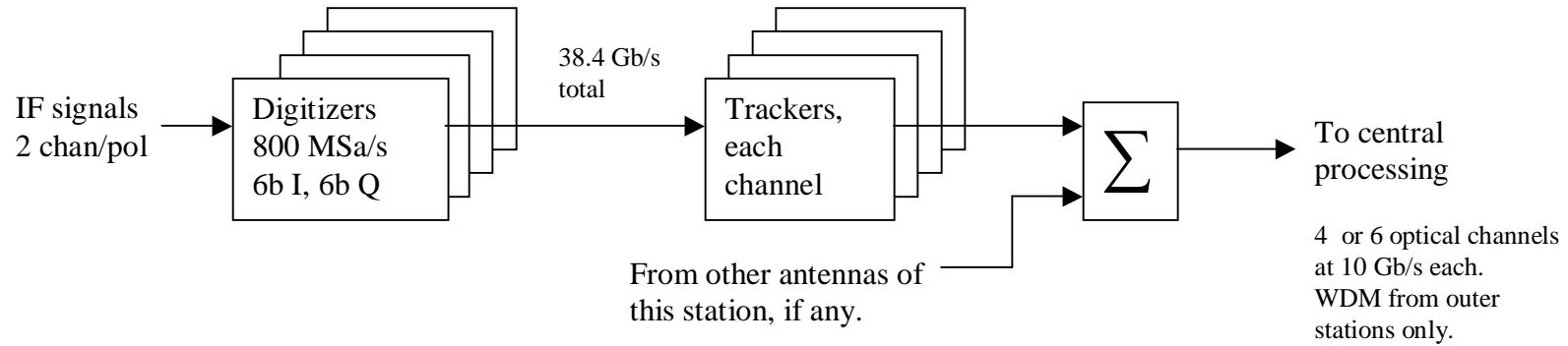
Within the processing units, the ASICs are the most critical and complex components, but they account for only 16% of the cost. We have estimated \$250 each as the marginal production cost of one 20x20 mm² packaged and tested chip. We have also allowed \$1M for development of each ASIC, including design and the NRE portion of fabrication.

Overall, we estimate that the central processing (correlator) for this huge array can be built today for about US\$17M, including development. This is less than 1.2% of the current estimate of the total cost of the telescope. Signal transmission to the center (not including the transmission medium) is estimated at US\$77M, or 5%; and the remaining signal processing items (digitizing and tracking) cost \$20M, or 1.3%. By the time of actual construction it is safe to assume that these costs will be considerably less, both in monetary units and relative to the cost of antennas, buildings, and land.

REFERENCES

- [1] J. Greenberg, private communication, 2002 Jul 01.
- [2] Y. Chikada et al., "Digital FFT spectro-correlator for radio astronomy." In *Indirect Imaging*, ed. J. A. Roberts, Cambridge Univ. Press, 1984.
- [3] J. Bunton, "A cross multiply accumulate unit for FX correlators." ALMA memo 392, 2001 Aug 25.
- [4] R. E. Crochiere and L. R. Rader, *Multirate Digital Signal Processing*. Prentice-Hall: New Jersey, 1983.
- [5] J. Bunton, "An improved FX correlator." ALMA memo 342, 2000 Dec 19.
- [6] H. Groginsky and G. Works, "A pipeline fast fourier transform." *IEEE Trans. Computers*, vol **C-19**, pp 1015–1019, 1970.
- [7] L. Urry, "The ATA imager." In *SKA: Defining the Future* conference, Berkeley, July 2001:
<http://www.skatelescope.org/skaberkeley> .
- [8] L. D'Addario, "Correlators: general design considerations." ATA Memo 24, 2001 Mar 13.
<http://intranet.seti.org/docs/ata/Memo/memo24.pdf> .
- [9] B. Cooper, "Correlators with Two-Bit Quantization," *Aust. J. Phys.*, vol 23, pp 521–527, 1970.
- [10] M. Green *et al.*, "OC192 transmitter in standard 0.18 μ m CMOS." IEEE International Solid State Circuits Conf., 2002.
- [11] J. Cao *et al.*, "OC192 receiver in standard 0.18 μ m CMOS." IEEE International Solid State Circuits Conf., 2002.
- [12] D. Ferris, J. Bunton, and B. Stuber, "A 2 GHz digital filterbank correlator." In *SKA: Defining the Future* conference, Berkeley, July 2001:
<http://www.skatelescope.org/skaberkeley> .
- [13] A. Cataldo, "AMI tweaks gate array for FPGA conversions." *EE Times*, 2002 Jan 22:
<http://www.eetimes.com/story/OEG20020122S0031> .
- [14] E. E. Davidson *et al.*, "The Design of the ES/9000 Module." *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, Vol. 14, No. 4, Dec. 1991.
- [15] D. P. Seraphim, "A New Set of Printed-Circuit Technology for the IBM 3081 Processor Unit," *IBM J. Res. Develop.*, Vol. 26, No. 1, January 1982.
- [16] USSKA Consortium, "The square kilometer array preliminary strawman design: Large-N, Small-D." SKA Memo No. 18, July 2002:
http://www.skatelescope.org/ska_memos.shtml .

Antenna-based processing



Central processing, each 800 MHz dual-polarization

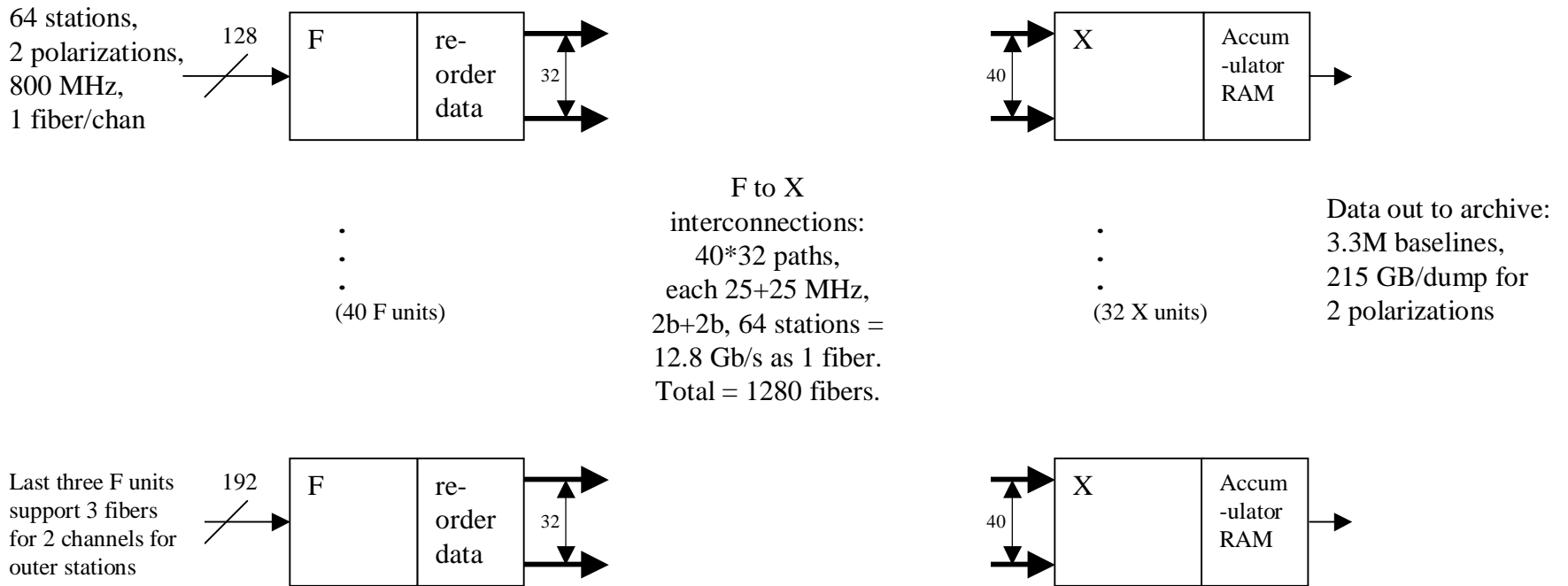


Figure 1: Digital signal flow diagram.

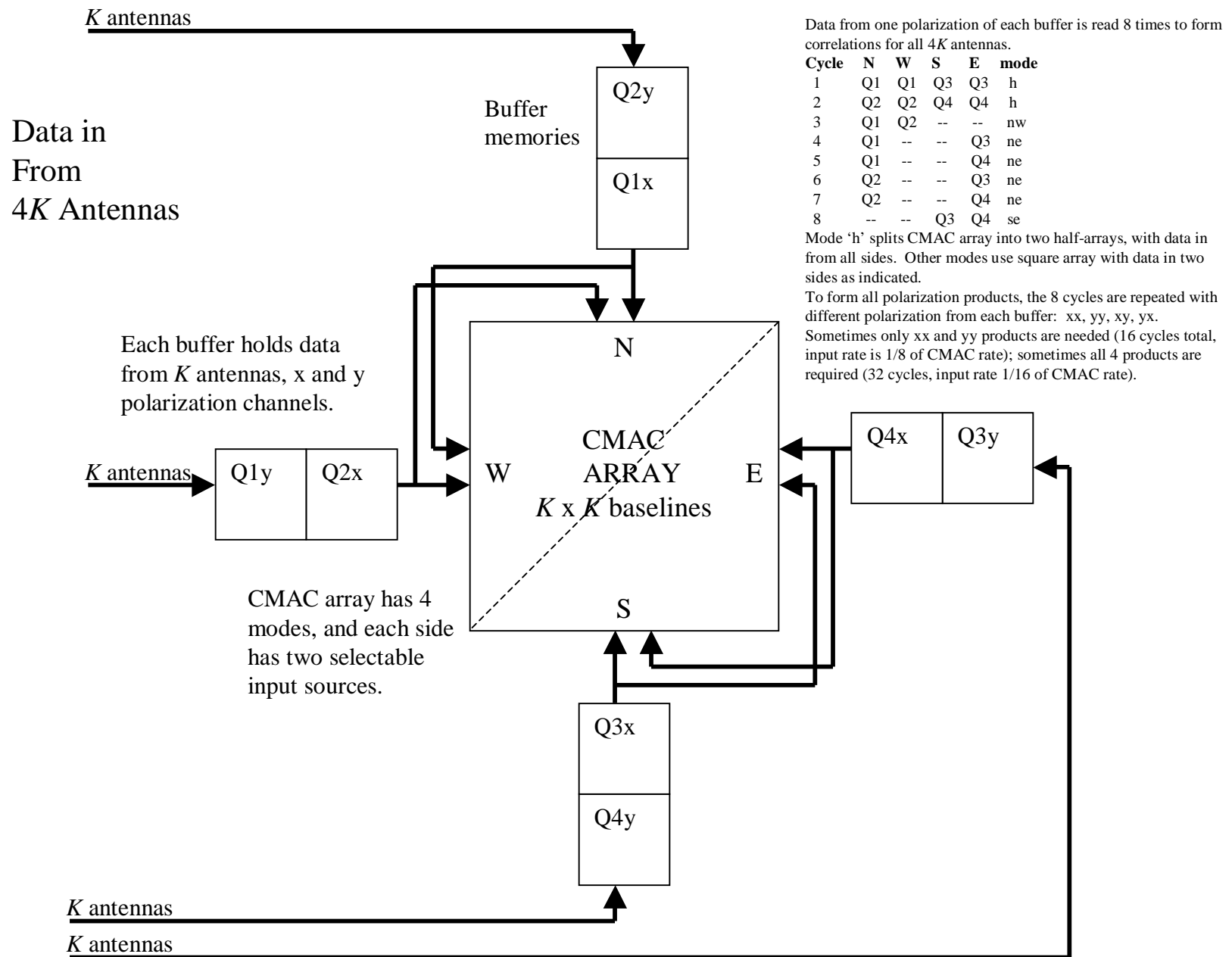


Figure 2a: Architecture of one X unit. In this paper, $K=640$. Signals from $4K=2560$ antennas are processed. Data from two channels of each antenna, x and y, are buffered and processed sequentially.

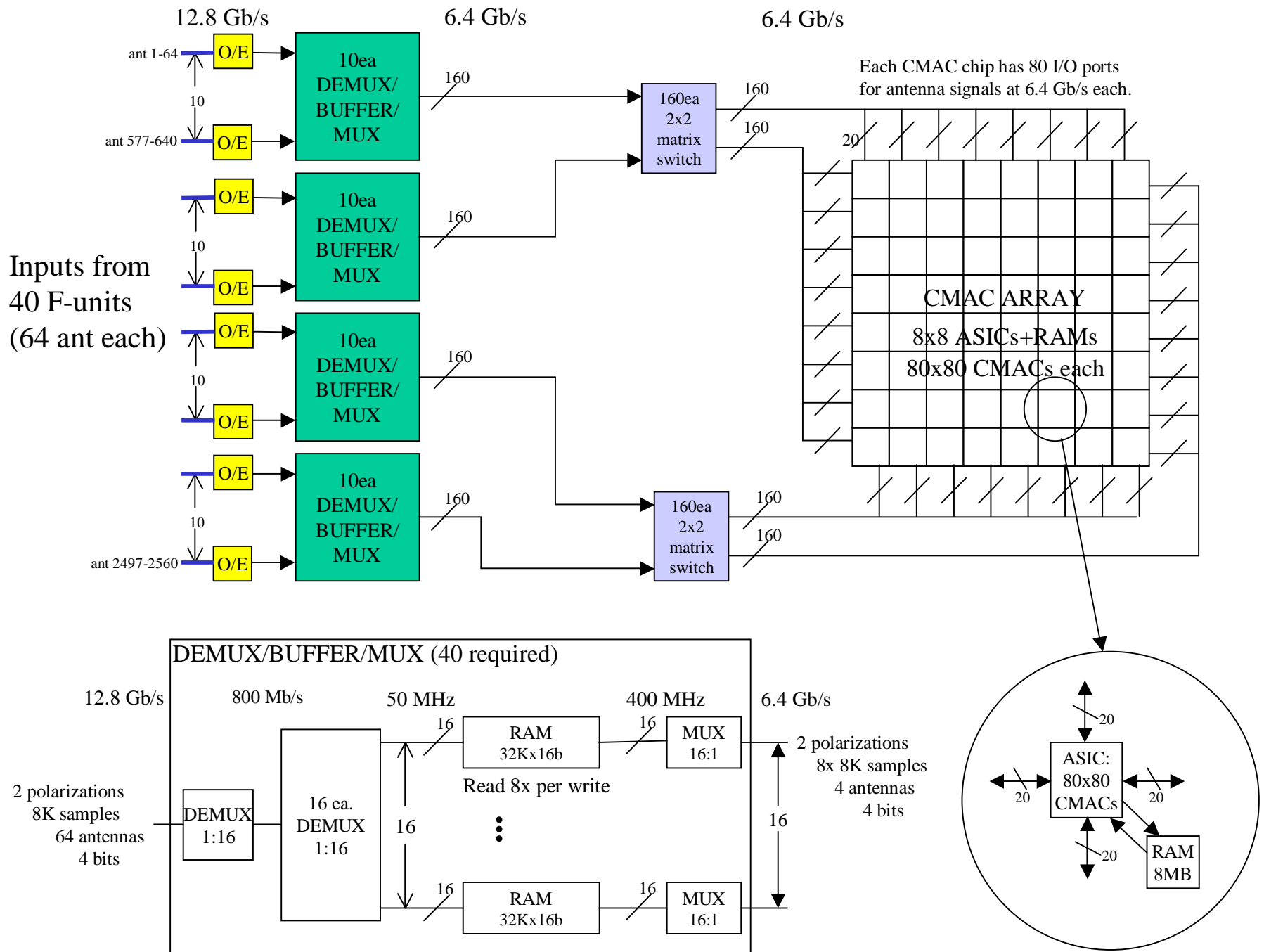


Figure 2b: Implementation of one X unit.

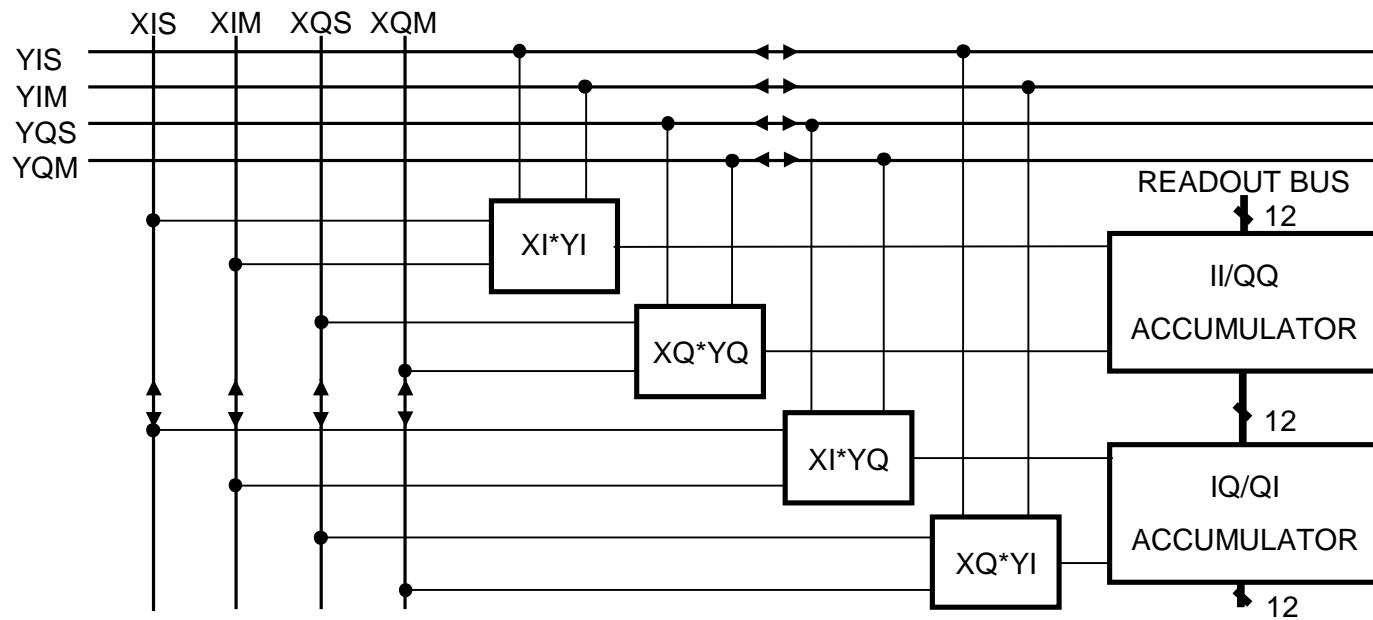


Figure 3: Block diagram of a complex multiplier-accumulator (CMAC). The 2x2 bit multipliers ($XI*YI$, $XQ*YQ$, $XI*YQ$, and $XQ*YI$) are operating on data encoded on 2 bits (S,M). Both the X and the Y buses are bi-directional.

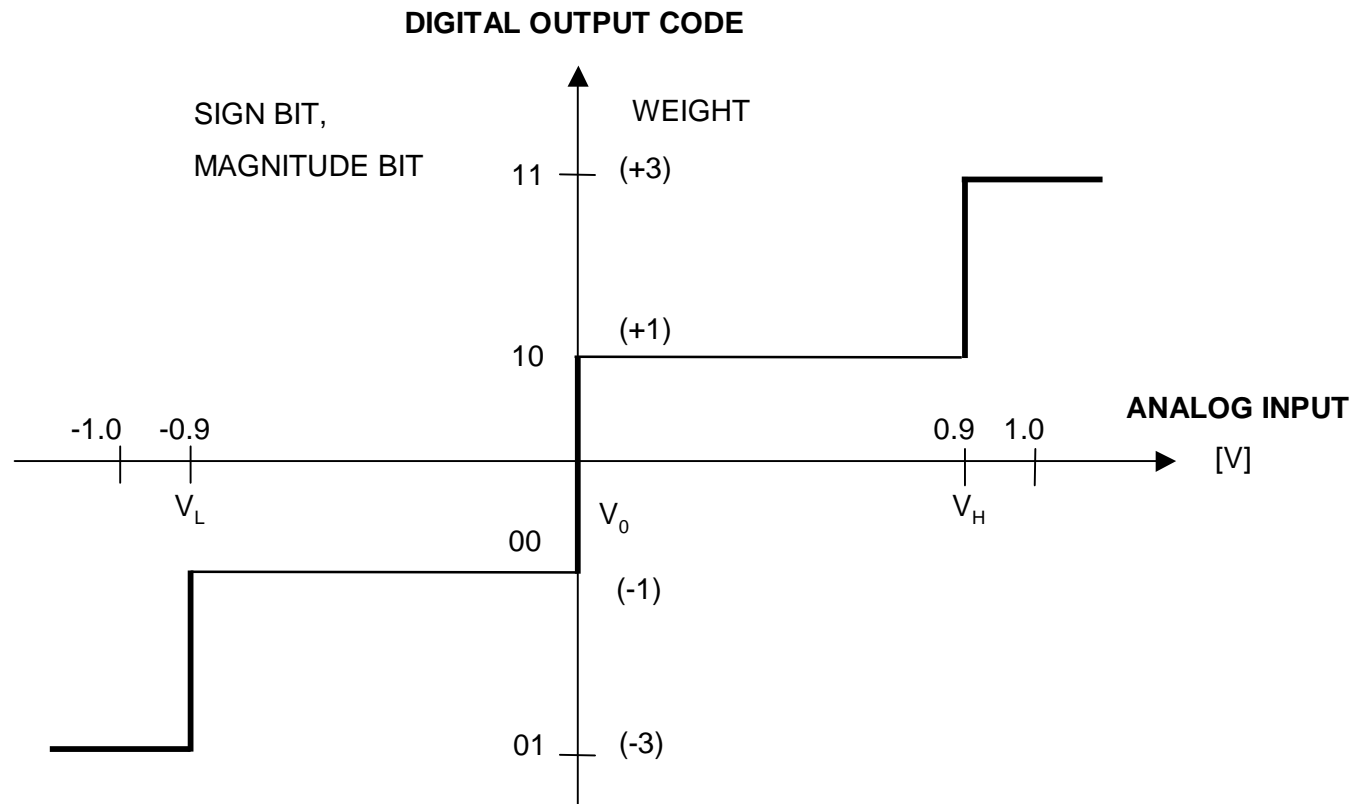


Figure 4: The transfer characteristic of a digitizer with four quantization levels encoded on two bits (sign and magnitude). Weight factors of (-3,-1,+1,+3) are assigned to each quantization level.

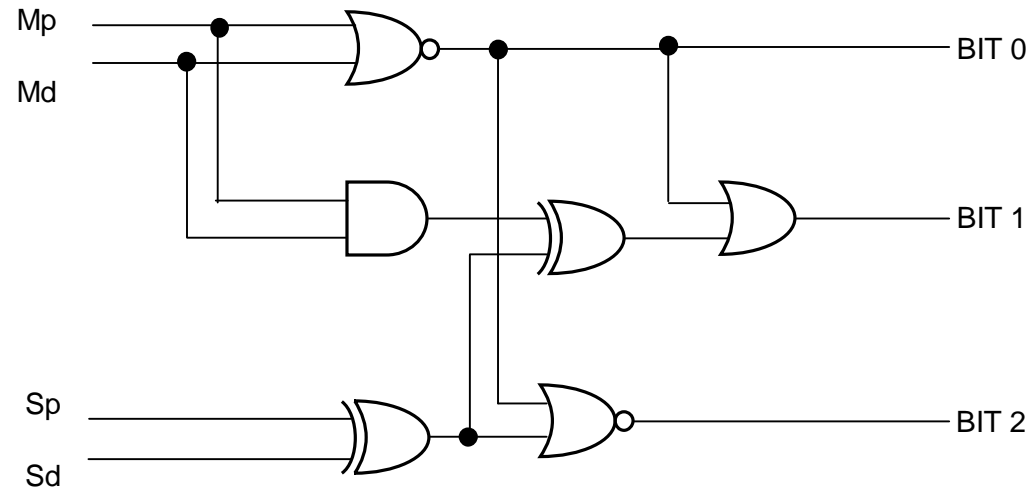


Figure 5: A logic diagram of the 2x2 bit Cooper multiplier

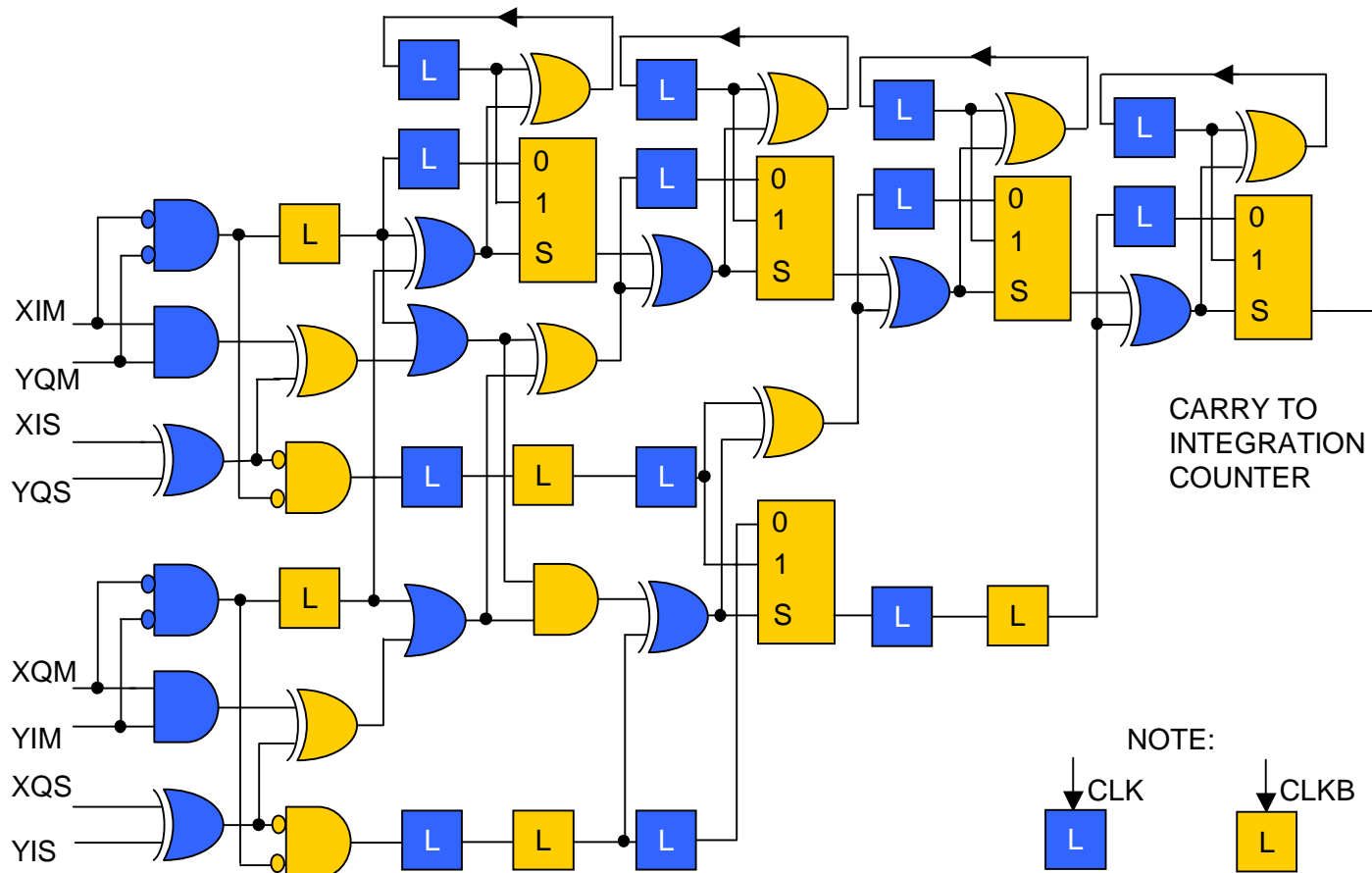


Figure 6: Logic diagram of a pipelined, 2x2 bit, dual multiplier-accumulator. Two dual multiply-accumulate circuits form a complex multiplier-accumulator (CMAC). Each logic gate is latched so that the multiplier-accumulator operates in pipeline mode with complementary clocks.

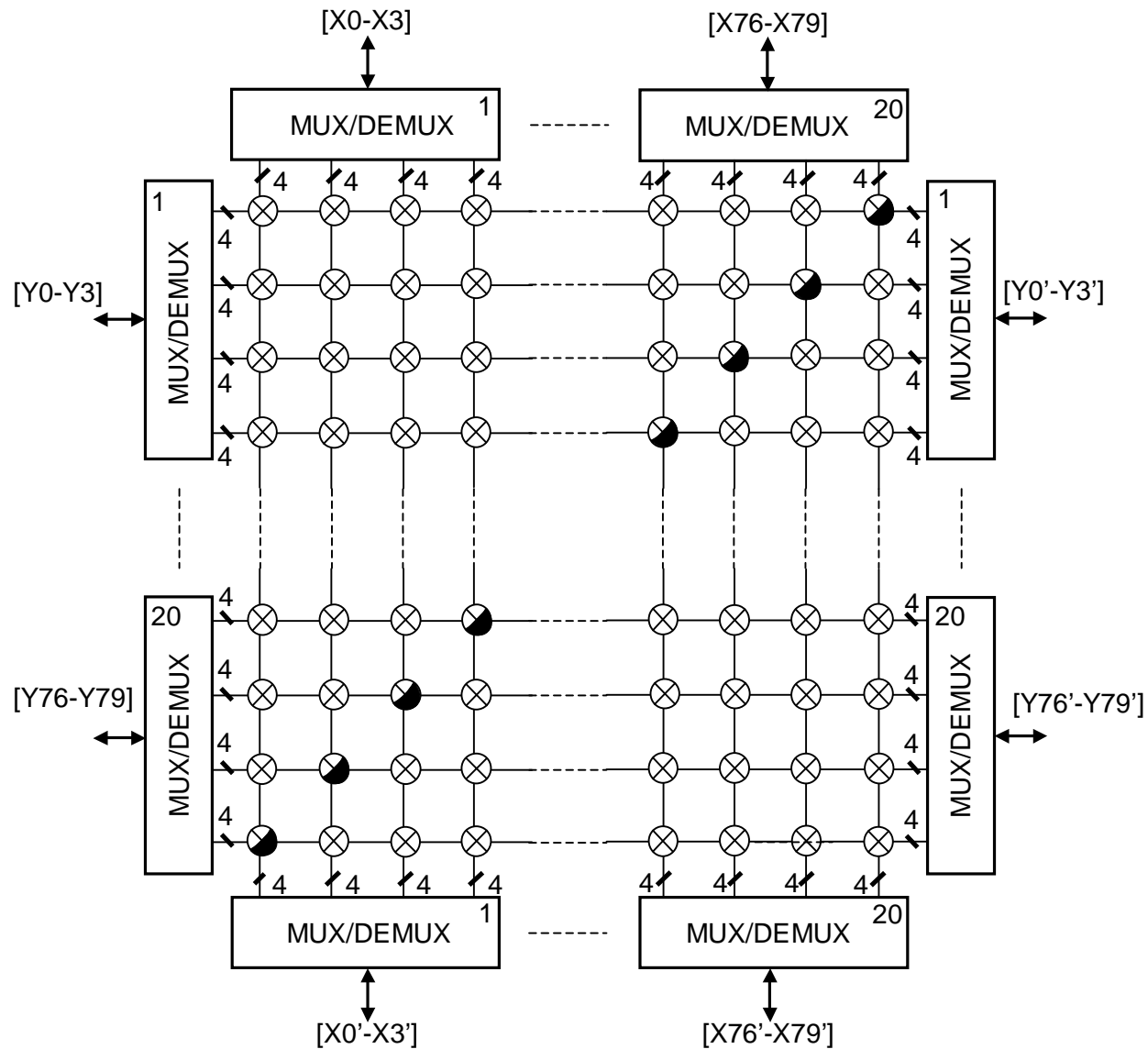


Figure 7: Structure of the 80x80-baseline cross-correlator chip with bi-directional ports (X_i, Y_j) and (X'_i, Y'_j). A complex (I,Q) multiplier-accumulator (CMAC) is located at each cross-point. The array can be split to operate as two independent halves along the top-right to bottom-left diagonal.

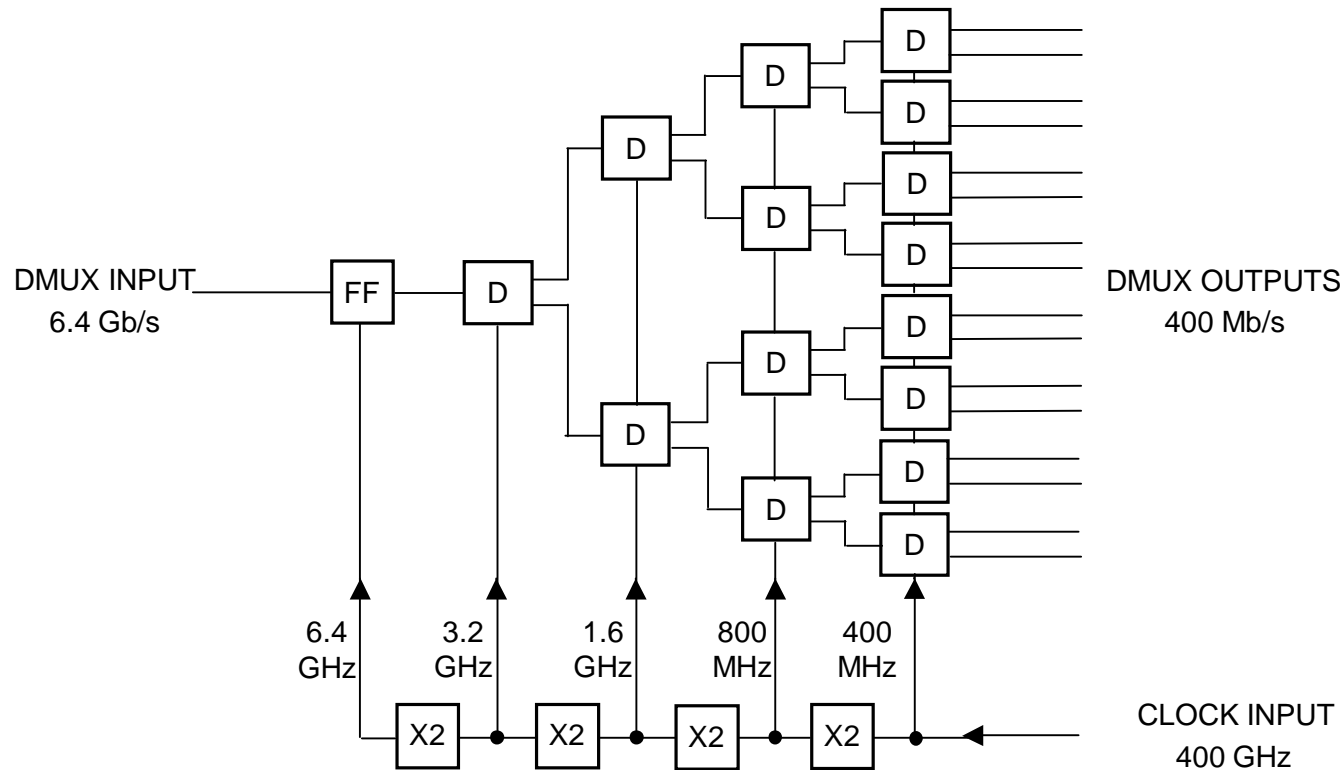


Figure 8: Block diagram of the 1:16 demultiplexer operating at an input data rate of 6.4 Gb/s and an output data rate of 400 Mb/s. The flip-flop (FF) retimes the input data to increase the input phase margin. All the appropriate clocks for the 1:2 demultiplexers (D) are generated on the chip by clock multipliers from a 400 MHz clock.

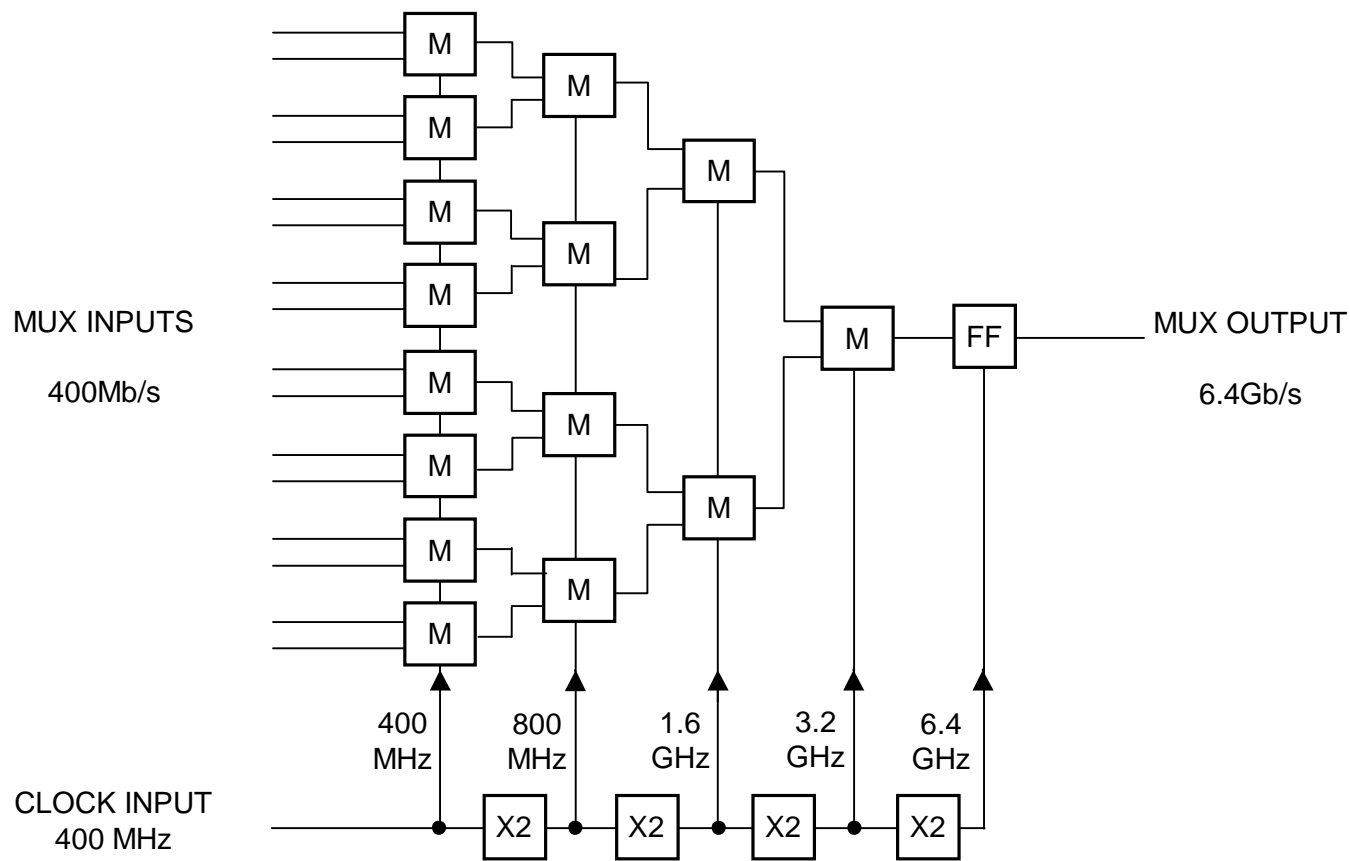


Figure 9: Block diagram of the 16:1 multiplexer operating at an input data rate of 400 Mb/s and an output data rate of 6.4 Gb/s. The FF retimes the output data to reduce the jitter at the multiplexer output. All the appropriate clocks for the 2:1 multiplexers (M) are generated on the chip by clock multipliers from a 400 MHz clock.

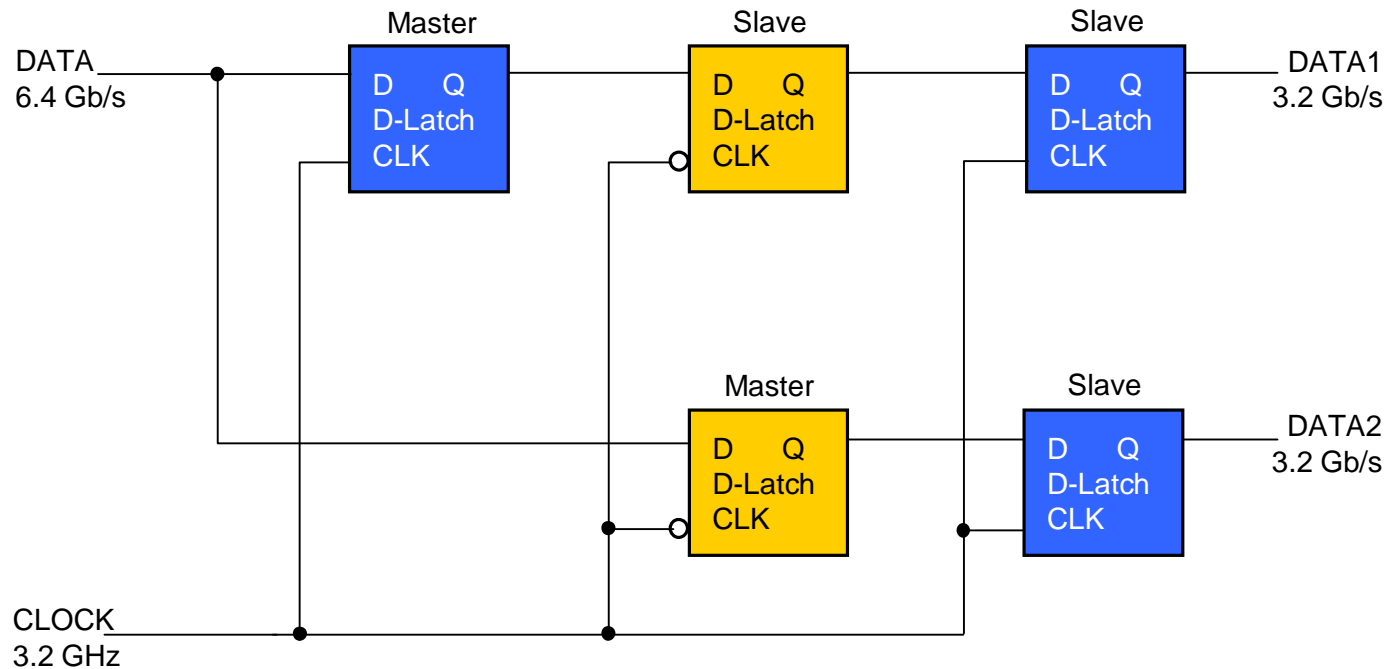


Figure 10: Logic diagram of 1:2 demultiplexer (D) block. The input data arriving at a rate equal to twice the clock frequency is demultiplexed into two data streams each with a data rate equal to the clock frequency.

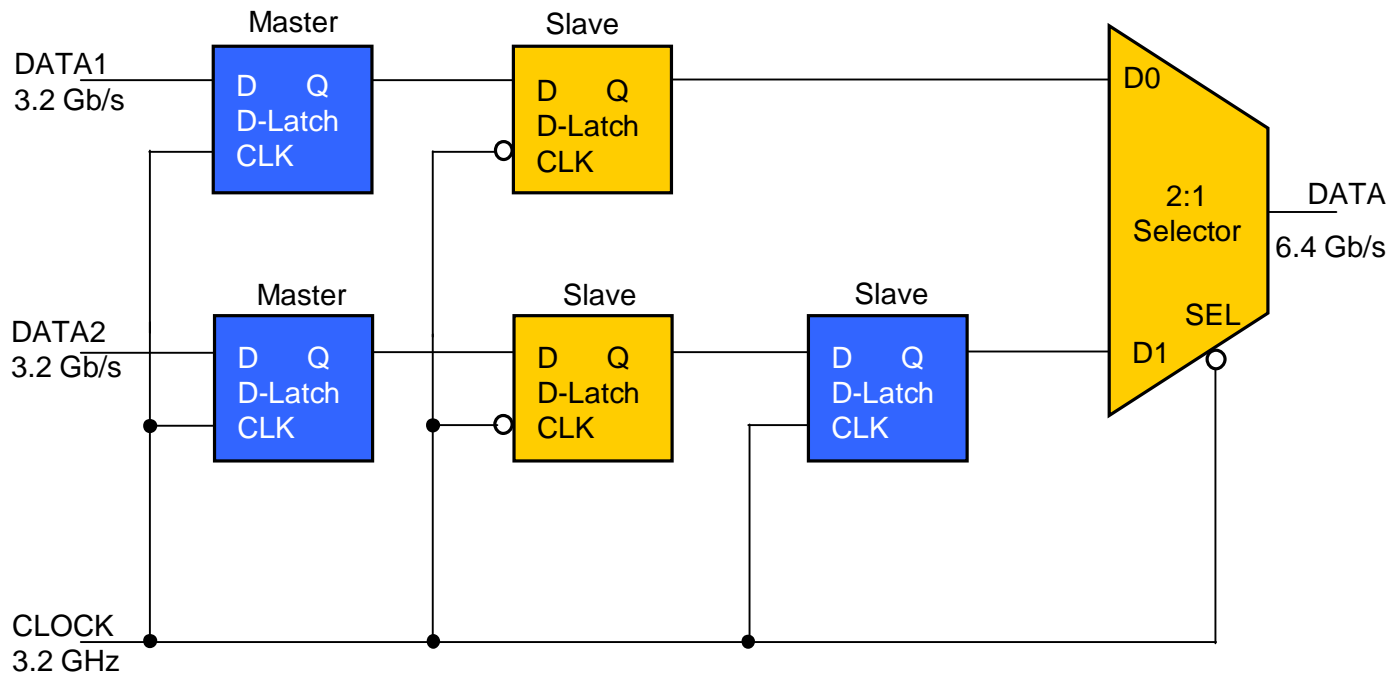


Figure 11: Logic diagram of 2:1 multiplexer (M). The two input data streams arrive at a rate equal to the clock frequency and are interleaved by the 2:1 selector in an output data stream having a data rate equal to two time the clock frequency.

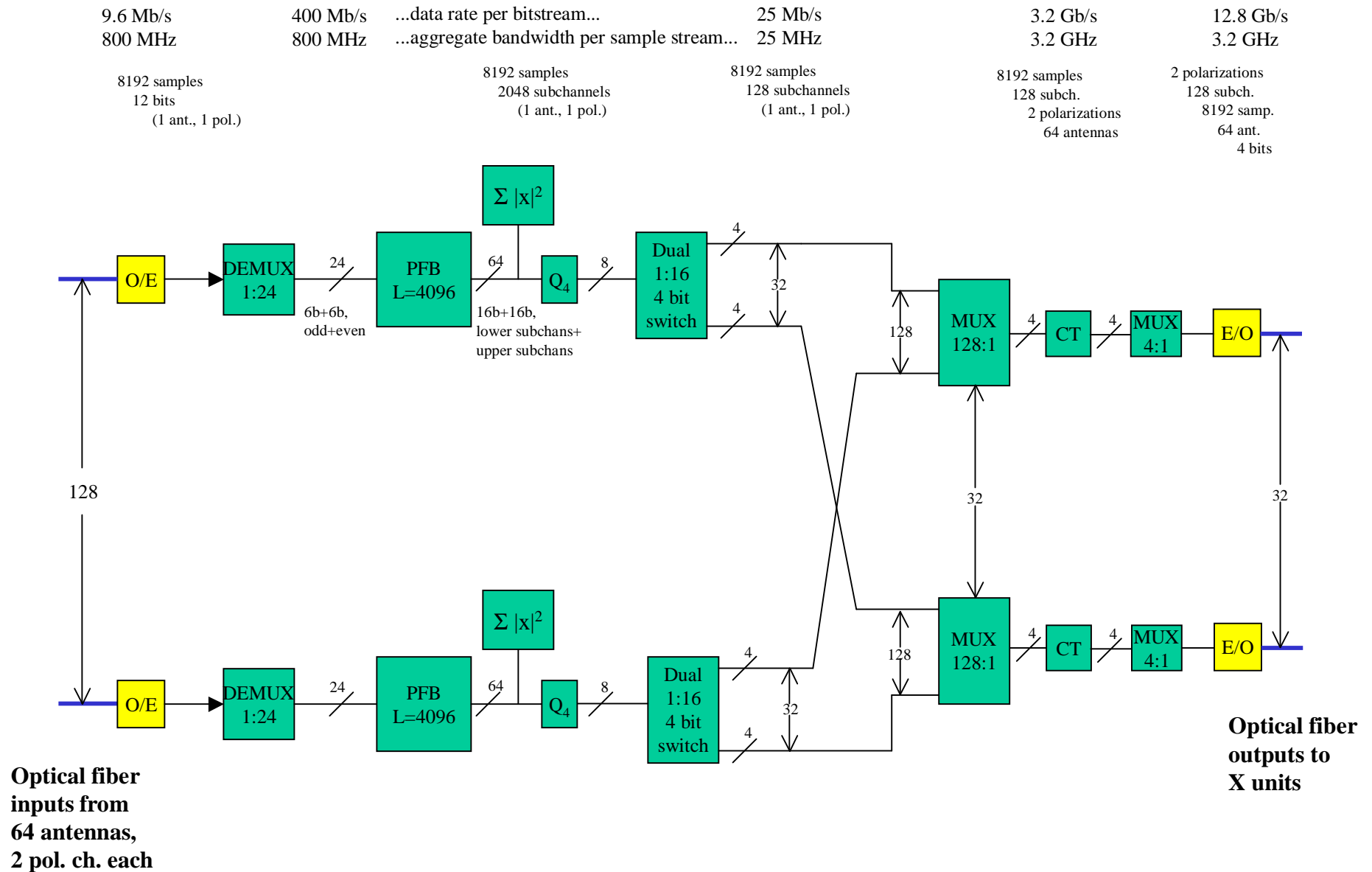


Figure 12: Architectural block diagram of one of the 40 F units. O/E: optical to electrical; PFB: polyphase filter bank; Q_4 : re-quantize to 4 levels; CT: corner turner; E/O: electrical to optical.

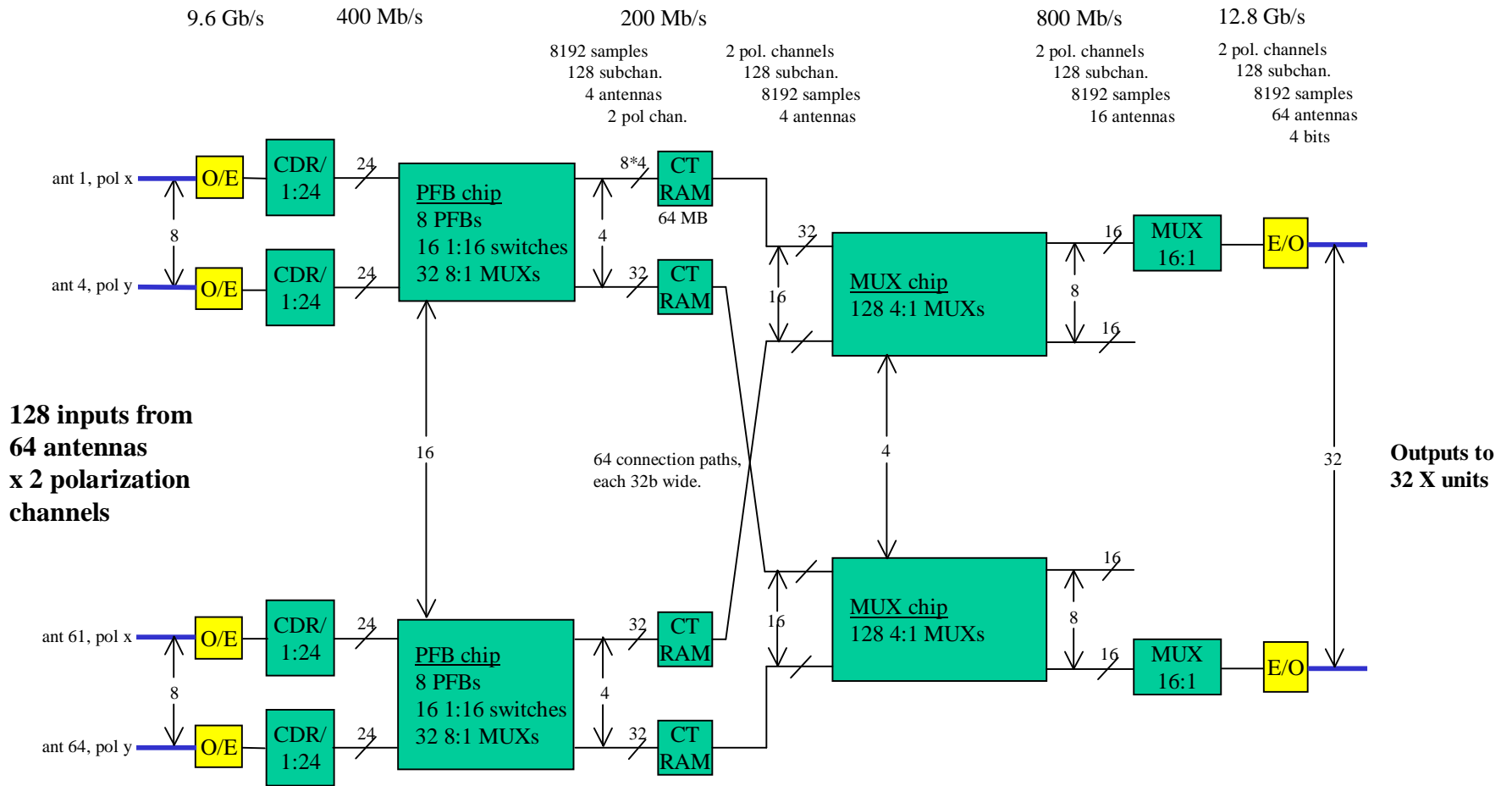


Figure 13: Implementation of one F unit.

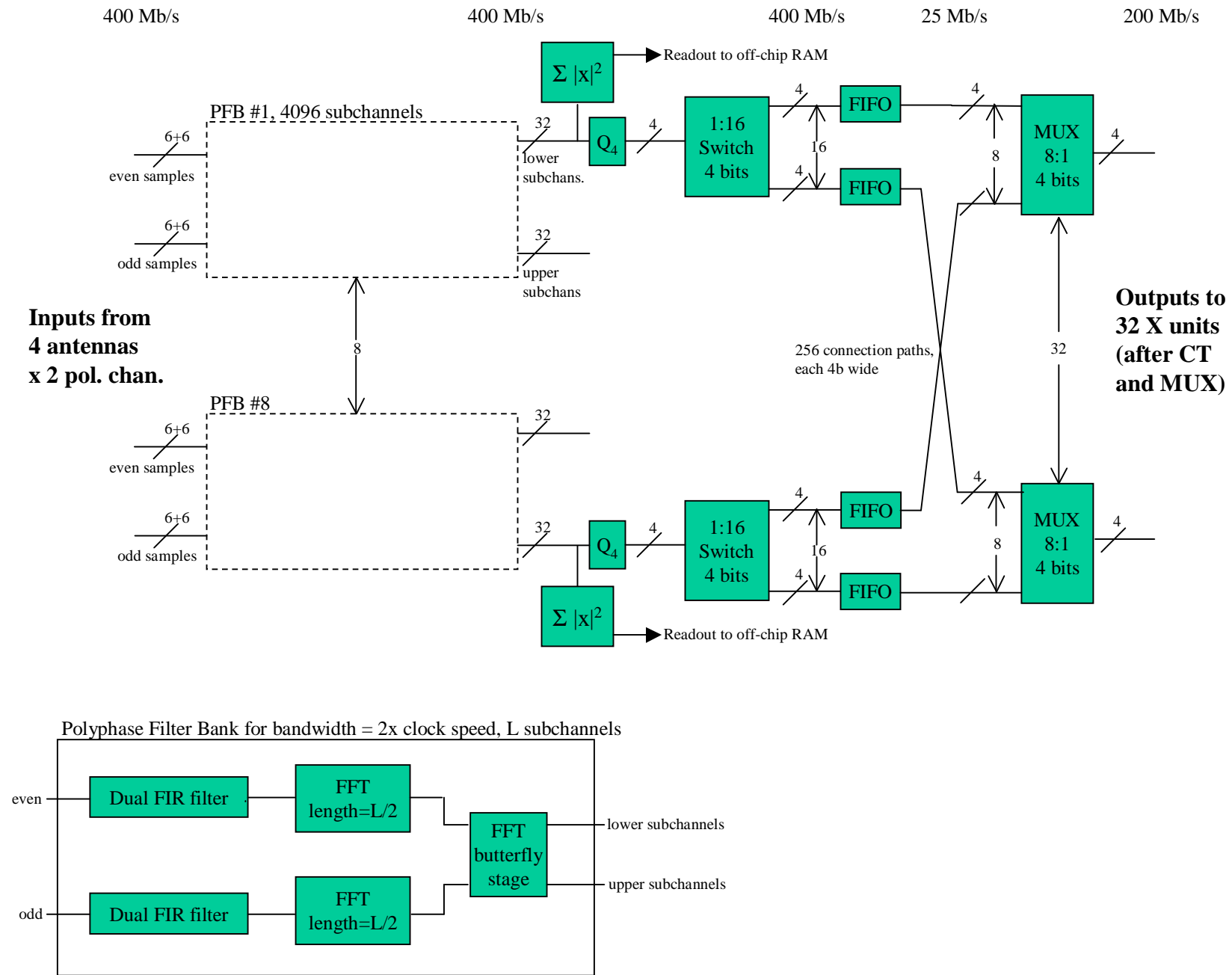


Figure 14: PFB chip block diagram. This ASIC analyzes 8 signal channels of 800 MHz bandwidth into 4096 subchannels, then organizes the data as 32 streams, each carrying all channels and 128 subchannels.

Table 1: Signal Processing Costs

LRD

8/4/2002

A. Basic Parameters			
Inner stations	2320	1	2320
Outer stations	160	13	2080
Total stations, antennas	2480		4400
Beams/station (outer)		1	
Digitization channels, BW/MHz		4	800

all costs in k\$ (2002)

B. Cost by major subsystem	quantity	each	total	Notes
Digitizers	17600	0.45	7,920	extrapolation of ATA design; per channel
Trackers				
inner antennas	9280	0.67	6,218	extrapolation of ATA design; per channel
outer antennas	8320	0.67	5,574	
Signal transmission, stations to center				electronics and photonics only
antennas (all)	17600	4	70,400	
outer stations	960	7	6,720	extra for WDM
Central processing sections	2	7,003	14,006	for 3.2 GHz total bandwidth
ASIC development	2	1,000	2,000	
Other development	1	1,000	1,000	
Total			113,838	

C. Breakdown into components	quantity	each	total	Notes
<u>Central Processing Section</u>				
Spectroscopic filter (F) units	40	17.28	691	800 MHz dual polarization, 2560 stations
F to X Connections	1280	2.01	2,575	64 stations per unit
Cross correlator (X) units	32	116.76	3,736	25+25 MHz per unit
Total			7,003	
<u>Spectroscopic Filter (F unit)</u>				
ASIC clock rate, MHz	400			128 chan. @ 800 MHz, 4096 subchan.
PFBs per ASIC	16			
Quad opt rcvr: 9.6 Gb/s	32	3.000	0.00	included in signal transmission
Quad CDR/demux16: 9.6 Gb/s	32	0.500	0.00	included in signal transmission
PFB ASIC	16	0.250	4.00	(bandwidth*antennas/units) / (filters*speed)
CT RAM: 64MB	64	0.120	7.68	
MUX chip	4	0.280	1.12	FPGA
Board	3	1.000	3.00	2 boards for input receivers at 64 chan each
Power, control	148	0.010	1.48	\$10 per large IC
Total			17.28	
<u>Cross Correlator (X Unit)</u>				
CMACs/ASIC	6400			2560 antennas, 25+25 MHz bandwidth
ASIC clock rate, MHz	400			
16x Demux16+RAMctrl: 400MHz	40	0.280	11.20	FPGA (XC2V1000)
Input buffer RAM: 16bx32k	640	0.010	6.40	
Quad mux16, 6.4Gb/s	160	0.400	64.00	
CMAC ASIC	64	0.250	16.00	
LTA RAM: 36bx512K	192	0.050	9.60	CY7C1370B-200, 18Mb
Board	5	1.000	5.00	4 boards for buffers
Power, control	456	0.010	4.56	\$10 per large IC
Liquid cooling mechanical hardware	1	2.500	2.50	
Total			116.76	
<u>Interconnections</u>				
Quad MUX16: 12.8 Gb/s	8	1.000	8.00	per F unit
Quad opt xmtr: 12.8 Gb/s, short	8	4.000	32.00	per F unit
Optical rcvr: 12.8 Gb/s, short	40	0.500	20.00	per X unit
CDR/demux16: 12.8G	40	0.250	10.00	per X unit
Optical connectors and cable	1	0.012	0.01	per connection path