NATIONAL RADIO ASTRONOMY OBSERVATORY

Charlottesville, Virginia

March 13, 1972

Scientific

VLA MEMORANDUM #13

TO: VLA File

FROM: B. G. Clark

LINE SYSTEM FOR THE VLA - DATA HANDLING

Clearly there are three major problem areas for doing line work with the VLA. They are 1) the cost of the correlator itself, 2) the data handling problem, and 3) the data presentation, storage, and accesability problems. This memo is a first quick look at the data handling problem.

SCOPE This look assumes a set of digital correlators, 351 of them to cover all possible baselines, with 128 channels per correlator (128 lag channels eventually results in only 64 frequency channels, unless you have double sideband data). We will talk about a ratio of field-of-view to beam size of about 128 to one, and therefore a fourier transform of 256 x 256 on the sky plane. We will therefore wish to sample the correlator data at intervals of ≈ 1 radian/512 ≈ 30 seconds. To summarize then, the output will be 44928 words (89856 for double sideband data) at 30 second intervals. Each word needs, as input, at least a precision of $1/2 \log_2$ (sample rate x integration time) + 3 = 18 bits.

This is a maximal system, probably beyond what we shall be willing to build, but it is informative to examine this case to see what the computing load will be.

LOBE ROTATION The analogue continuum correlators are most easily designed with sinewave outputs, but digital correlators are intrinsically perfect integrators. Therefore instead of using the continuous lobe rotation with offset frequencies as proposed in VLA Volume I, we would either stop the fringes entirely, or use the stepwise rotation scheme suggested

by John Granlund in the ITT report, and used in the line system on the Green Bank Interferometer.

If no lobe rotation is used, the clippers must be extremely good to avoid a DC offset in the correlators. Over a reasonably long interval, the number of "1"s output from the clipper must equal the number of zero's to within a few parts in 10^6. This is probably achievable with a partially digital servo on the clipping level, or by recording the relative proportions of 1's and 0's for generating a correction.

The two-state scheme proposed by Granlund makes a complete cycle in 32 switches, and has both state pairs with equal times on all baselines. The two state phase switching can be used for either of two purposes: a) to make each observation a difference of in-phase and anti-phase observations, eliminating the need for such tight control on the clipper DC offset, or, b) separating the sidebands of an observation made with DSB front ends.

To do both of these things, it is possible to do a four state switching scheme, which gives all four phase differences equal time. However, the cycle lasts for 1024 states and I believe it is impractical.

It is also possible, at a cost of an additional 11% in signal-to-noise ratio, to analyze sine-wave data, dumping the correlators a few times a second and using a table look-up to select whether the data is to be added in phase of 0°, 90°, 180°, or 270°. There will be an additional source of noise also, because the cancellation of offsets and separation of sidebands is not exact. This appraoch impresses me as being less profitable than the two state cycle.

CORRELATOR CONFIGURATION The basic correlator will probably consist of a fast memory-counter, to record the data as it comes in, and a minimum cost store to hold the results until they can be processed. In the Model 3 correlator, the memory-counter is the 8281 counter-register, and the low cost store is core.

The smallest module of mass memory economically feasible is about 8K. If we can do a READ-ADD-STORE in 1 μs, then the time required to

fill the module is about 8 ms. To hold the loss of observing time down to 5% during the read/store phase, the correlator can be read about 6 times/second only. The counter-memory must therefore have a length of 22 bits, of which 14 must be brought out to the slow store, to avoid truncation noise. (The length of the fast store can be reduced by modularizing it in smaller chunks than 8K, and reading out at shorter intervals, keeping all but one of the submodules going during the readout.) The slow store must have a length of 22 bits, of which 18 must be further processed.

The number of words of counters required is just the number of channels - 44928. The number of words of slow store required is (44928) x (2 to allow for the possibility of reducing double sideband data) x (2 Buffers so that data can be reduced in place) = 179712 words = 22 8K modules.

CORRECTIONS Bad channels can be eliminated and the van Vleck corrections may be applied in a straight forward manner without consuming very much computer time. The sine need be taken only if the correlation coefficient is greater than 0.001, so except for masing lines the correction will take only about 7 core cycles/word-10000/second. In the worst case it is only 100000/second.

FOURIER TRANSFORM FOR SPECTRUM The trickiest part about this is getting the scaling right, so that none of the intermediate sums overflow and that truncation noise remains negligible. This appears possible in 18 bits only if great care is excercised - it is not clear that commerical units handle rounding in the proper fashion. The problem is straightforward in 24 bits, but, even here, one probably needs to do some scaling. If no rounding is employed, the truncation noise will be about 6×10^{-6} of the largest number. With a pure tone into the correlator, this number can be as much as 128, so if no scaling is done, the truncation noise is unacceptably high, 7×10^{-4} when receiver noise is 4×10^{-5}. It appears necessary to specify a scale factor. Automatic scaling would result in a scale factor varying over the sky and would require about 40,000 core cycles per second to unscramble. The alternative is to have the observer specify a scale factor.

The longest FFT hardware unit whose timing I have on hand is an 18 bit unit, which does a 128 point transform in 4 ms. Therefore it can do 351 transforms in 1.5 seconds. A 24 bit unit should not be much slower.

The use of a 24 bit hardware FFT device does not necessarily imply the use of a 24 bit CPU to control it, but it would certainly simplify things.

After the Fourier transform, it seems likely that the spectra could be truncated to 16 bits without loss of SNR. The total output data rate is thus ~50,000 16 bit words every 30 seconds, or 7×10^7 words for a 12 hour observation. A 1600 BPI computer tape will hold about 1.8×10^7 16 bit words. We have three choices at this point: We may add the cost of buying and storing 8 tapes a day to the operation cost of the VLA, or we can develop an automatic edit, correct, and average scheme, and save only the data gridded onto a u,v map, or we may purchase a more compact form of data storage, such as the Ampex rotating head digital tape recorder. (This device puts 4.4×10^{10} bits on a reel of two inch tape.) Similar densities are available in punched plastic cards.

EDITING AND CORRECTION This must be held to a minimum. The times for this reduction can be estimated from the times for the continuum correction and editing: the line case is essentially identical except for the addition of one complex multiplication per channel. This complex multiplication consumes about 60 core cycles, not counting bookkeeping. Doubling this number to account for the bookkeeping results in 2×10^5 operations/second, not at all excessive, compared to what is already done at this stage. However, it is enough that one would like to do it all in one pass. With suitable organization, this can be done.

SORT-MERGE This is the bad part. All alternatives are expensive. The ones I can think of are listed below:

1) Random access memory, 70 million words. A single read, store operation sorts the data completely. To be economically feasible the cost must be below about .05¢/bit.

2) Rotating memory. The 2314 disk in Charlottesville stores 1.4×10^7 16 bit words per spindle. Utilizing all eight spindles gives the possibility of a sort with seven output units, without unnecessary access arm motion. Under this condition, the sort could take place at about 35,000 words/second, 35 minutes per pass. A seven output sort requires about $\log_7 (7 \times 10^7) = 10$ passes at most to complete.

3) A completely dedicated computer with four magnetic tape units (two input, two output) doing a multiple pass binary sort. This would take about 26 passes. In order that the entire data set reside on two reels, the tape units must have a density higher than 3000 8 bit bytes per inch, or must be built to accomodate non-standard reel sizes. The data transfer rate must exceed 90 kilobytes/second. This is probably only possible with two selector channels, running two drives simultaneously.

The most attractive of these possibilities appears to the the second.

MAP MAKING The hardware FFT device which is used for calculating spectra may be used also for map making. One would, first average the data onto the u,v grid (taking about the same time as 1 sort pass). Then, each u,v plane would be read in a column at a time, Fourier transformed, and output. Assuming a FFT time of 10 ms per 256 point column, a bookkeeping overhead time of 25 ms, and an I-O time of 15 ms this takes 13 seconds. The rows and columns are then interchanged (if the output from the first stage was properly organized this will take only 5 seconds or so) and the process repeated. Thus each map will take about 30 seconds, and the entire 128 maps are produced in slightly over an hour. If you make a 512 x 512 point map, or larger, this time becomes a major consideration; otherwise you are limited by the sort/merge stage.

SUMMARY The hardware requirements in addition to those required for continuum and to the correlator itself are:

1 CPU, with hardware FFT, 32 K memory perferably 24 bits, 18 possibly acceptable, ~$100,000.

2 ultra high density archival memory drives, ~$300,000.

100 M words rotating memory, ~$200,000.

Hardware phase switching/differencing in the correlator

Output device(s), $?

The computers time will be spent as follows (thousands of operations/sec):

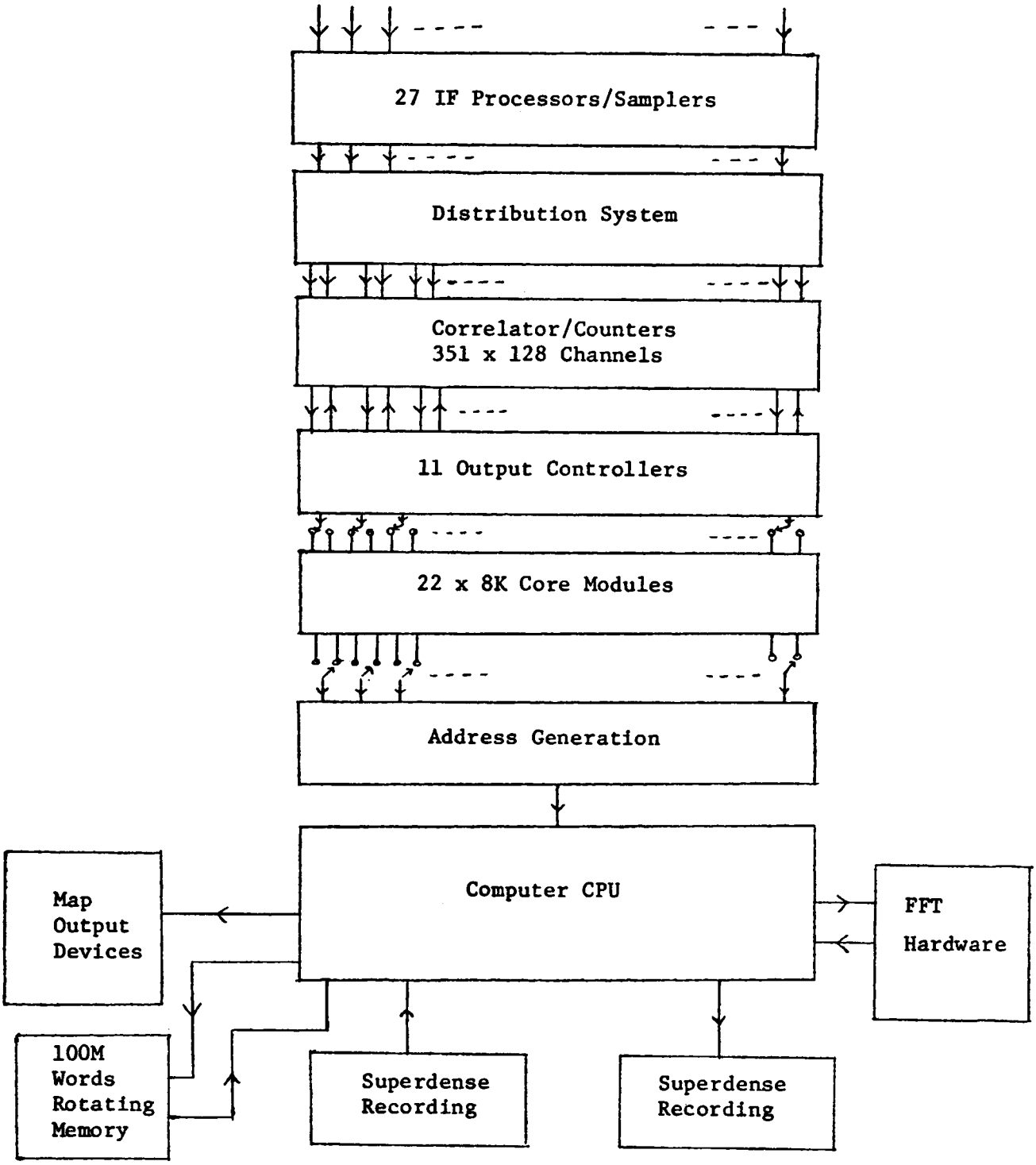
Real time:

Input from correlator	3	
Van Vleck Corrections	100	(max)
Delay realm editing	5	
FFT bookkeeping	100	
Hardware FFT (equivalent)	50	
Output to archive	5	
	263	

Assynchronous:

Input from archive	10	
Ten pass sort	500	
FFT Bookkeeping	70	
FFT Mapping (Equivalent)	30	
Output display	?	
	610 +	
TOTAL	1000 +	

CONCLUSIONS 1) The computer configuration listed in the summary above would cost about $700 thousand. The 22 x 8K core modules (which I have regarded as part of the correlator, though they might equally well be regarded as part of the computer) would, with interfacing, cost about $300 thousand. Therefore, the cost of the computer to handle a correlator channel will, in sufficient quantity, cost 20-40% of the cost of the channel itself. The computer and IF processor equipment are comparable in cost. 2) 16 Bit precision is marginal or submarginal for some purposes. The problem is more severe with a hardware FFT device because you cannot re-scale during the transform. The continuum study must also be re-examined to see what operations need double precision.



VLA Multichannel Correlator-schematic connection.