NATIONAL RADIO ASTRONOMY OBSERVATORY
Charlottesville, Virginia

August 1, 1973

VLA SCIENTIFIC MEMORANDUM NO. 112

B. G. Clark

A general-purpose scientific computer has been specified for the data reduction of continuum data. If this computer is also used for a cross-correlation line receiver reduction, it will clearly have rather limited capabilities--to perform the reduction done for the continuum system on N channels of data clearly requires N times as much computing. It is the purpose of this memo to estimate what the limitations are. The estimates below are rather sloppy, but I believe conservative.

The computing load will be assumed to consist of the following principal components: (1) Fourier transform in the lag-frequency domain to find spectra; (2) post-observation correction and editing; (3) sort from (correlator, time) order to (u,v) order; (4) prepare (u,v) plane and Fourier transform to (x,y) plane; (5) display maps; and (6) clean. The time required for each of these processes will be estimated below. In these estimates, it will be presumed that a twelve hour observation has been made, primarily of one source, with, perhaps, interspersed calibrators. There are N channels of cross-correlation (N/2 frequency channels with no grading, N/4 with Hanning weighting), sampled every S seconds, from K correlators. The data will be mapped onto an MxM grid in the (u,v) plane, presumably at about two points per beam--that is, the most extended baseline used goes about halfway to the edge of the (u,v) plane.

There is clearly a relationship between M and S. It is roughly appropriate that the longest baseline should move through about half a cell during S seconds. Therefore

$$S \approx 13750/M.$$

The time estimates are as follows:

(1) Fourier transform for spectrum. Since the transform will be done in a special-purpose device, most of the time will be I/O time and multiplication by a grading function. A total time of 10 μs per point is reasonable, giving a total time, in seconds

$$T_F = 0.43\ NK/S$$

(2) Post-observing correction should amount to a sine/cosine pair per correlator and time point plus a complex multiplication per frequency point. Allowing a factor of three (for overhead) above the minimum code gives

$$T_E = (13 + 0.33\ N)K/S$$

(3) Sorting, like the FFT, is a nlogn process. It also depends, to some extent on the amount of memory allocated as a buffer, especially for the first stage. The time is also dependent on the length of the records. In this case, I presume (u,v) 1-1/2 bytes each, data (Re,Im) 4 bytes,frequency, correlator number, 1-1/2 bytes each, total 10 bytes. If we utilize an 80 K byte buffer area, the sort will require a total time of about

$$T_S = 0.8\ NK(14 + \log_2 NK/S)/S$$

On the other hand, it is sometimes convenient to average data as you go, into the appropriate (u,v) cells. This process precludes using three-dimensional transforms to correct for the curvature of the sky and, of course, any post-observing editing. We may estimate the time for sorting in this fashion by noting that it is similar to one pass on length 43200NK/S, and $\log_2 \frac{43200\ NK}{S}$ passes on length NM^2

$$T_S{}' = 0.8\ NK/S + 0.000018\ NM^2\ (13 + \log_2 \frac{NK}{S}).$$

This method is of advantage when

$$M \lesssim 3K$$

(4) The transform itself will again be done in a peripheral unit. The manipulations required to place the data in the u,v plane and to grade the plane amount to about one complex add per input point and two complex multiplies per output point (presuming that no convolution is done in the u,v plane). The total time required is

$$T_T = 0.08\ NK/S + 0.000035N\ M^2$$

(5) It remains far from clear to me just what needs to be done in the way of display. Any estimate is a wild guess. I shall presume that this time is the sum of times necessary to make a ten-point-per-beam gray-scale type display in (α,δ), one map per f, plus the same in (α,f), one map per δ plus the same in (δ,f), one map per α.

$$T_D = M^2N(0.0022 + 0.000018\ \log_2 NM^2).$$

(6) Clean is assumed done iteratively between the (x,y) and (u,v) planes. The most time-consuming operations occur in the u,v plane, and total one complex multiply per cleaned component per u,v point. It is practicable to clean only about $M^2/10$ components. In fact, a perfectly reasonable map often results from cleaning only M^2 /100 components, and this will be used for the time estimate.

$$T_C = 1.2 \times 10^{-6}\ NM^4 + 2\ T_T$$

Three cases will be considered:

Case A: Bare bones system. All baseline parameters and system temperatures known ahead of time so that all corrections may be applied at observe time, no post observing corrections are applied. Sorting is done while observing, and observations are combined into (u,v) cells while observing. A single display will be provided, and no clean. Total time given by

$$T = T_F + T_E + T'_S + T_T + T_D.$$

Case B: A very comfortable system. Provision made for post-calibrating amplitudes and phases, two transforms done with different gradings, with two displays at each, as well as a real time transform and display.

$$T = T_F + 2T_E + T_S + T_S + 3T_T + 5T_D$$

Case C: System with clean. Same as Case B, but with T_C added, and two more displays.

A major part of the system time in every case is spent on map and u,v manipulations and displays. These are independent of the number of correlators and antennas. Therefore, the system requirements are much less dependent on the number of antennas than on the numerical field of view, M. This also implies that these calculations, based on twelve hour observations, may not be conservative enough if shorter observations are used.

A tabular presentation of the formulae above is given in Table I. For each of the three cases, and for various values of M, two numbers are tabulated. These numbers are the number of lag channels (the number of frequency channels is about half this) that can be handled without a growing backlog, by the synchronous system computer. The first number, N_O, is the number of channels if only a single pair is connected as an interferometer. The second number, N_A, is the number of channels which can be handled if all 351 interferometer pairs are present.

Finally it should be noted that the time limiting operations do occur in relatively simple loops, for which small-scale computers are very much more cost-effective than a general-purpose machine.

BGC/pj

Numerical View of Field	Integration Time	Case A Minimum Processing		Case B Comfortable System		Case C Clean	
M	S	N_O	N_A	N_O	N_A	N_O	N_A
32	420	18000	14000	3800	1400	3200	1300
64	210	4600	3600	960	480	620	380
128	110	1100	890	240	150	80	70
256	53	290	220	60	43	7	7
512	26	72	56	15	12	1	1
1024	13	18	14	4	3	-	-