THE VLA SPECTRAL LINE SYSTEM

A Progress Report

January 1976

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The VLA was originally conceived as a high resolution array for continuum observations. However, as spectral line radio astronomy developed during the sixties, it became clear that the VLA also could become a very powerful array for spectroscopy. A preliminary (and sketchy) design of a spectral line radiometer was therefore included in the VLA system.

A useful spectral line system requires mass storage and data handling capabilities of staggering proportions. Fortunately, the rapid development of electronic and computer technology has reduced the problems of obtaining these capabilities to manageable levels, and the present report describes a VLA spectral line system that will give acceptable performance at a reasonable cost. The report is not a complete and final answer to the spectral line design problems, but it is a status report of the continuing development as of January 1976. Much more work needs to be done on the spectral line system before the design is complete.

The VLA is expected to have limited spectral line capability sometime during 1978.

Chapter II. Scientific Needs and Capabilities

In this chapter we will attempt to define the minimum requirements for a scientifically useful VLA spectral line system. In addition, we will try to describe those further capabilities which, although not absolutely necessary, are highly desirable. The problem divides into three moderately distinct areas: data acquisition, data processing and display, and data storage. We will treat each of these areas separately.

A. Data Acquisition

1. VLA System Parameters

We will assume that the line system will be restricted to the basic VLA configurations and frequencies. The stations comprising the four VLA configurations are listed in Table II-1 and the four VLA frequency bands are listed in Table II-2. In the latter table we also list the anticipated system temperatures, synthesized beamwidths, and map noise values. The noise is determined using the formula (Greisen 1975)

$$\sigma(T_B) = c \frac{\kappa \; T_S}{\sqrt{\Delta\nu \; t}} \left(\frac{\lambda}{\beta_o}\right)^2 G(\gamma/\beta)$$

where

c = 1.77 for 10 telescopes, 0.63 for 27 telescopes

κ = array efficiency factor determined by the taper

TABLE II-1

NATIONAL RADIO ASTRONOMY OBSERVATORY

VLA PROJECT

ANTENNA LOCATIONS

WEST ARM			EAST ARM			NORTH ARM		
STATION	DIST. FROM ORIG. M	DIST. FROM ORIG. FT.	STATION	DIST. FROM ORIG. M	DIST. FROM ORIG. FT.	STATION	DIST. FROM ORIG. M	DIST. FROM ORIG. FT.
DW1*	-80.00*	-262.47*	DE1*	-40.00*	-131.23*	DN1	0	0.00
DW2 - CW1	44.85	147.15	DE2 - CE1	44.85	147.15	DN2 - CN1	54.86*	180.00
DW3	89.93	295.05	DE3	89.93	295.05	DN3	94.86*	311.22
DW4 - CW2 - BW1	147.33	483.37	DE4 - CE2 - BE1	147.33	483.37	DN4 - CN2 - BN1	134.86*	442.45
DW5	216.07	708.89	DE5	216.07	708.89	DN5	194.82	639.17
DW6 - CW3	295.43	969.26	DE6 - CE3	295.43	969.26	DN6 - CN3	266.38	873.95
DW7	384.89	1262.76	DE7	384.89	1262.76	DN7	347.04	1138.58
DW8 - CW4 - BW2 - AW1	484.00	1587.93	DE8 - CE4 - BE2 - AE1	484.00	1587.93	DN8 - CN4 - BN2 - AN1	436.40	1431.76
DW9	592.40	1943.57	DE9	592.40	1943.57	DN9	534.15	1752.46
CW5	709.79	2328.71	CE5	709.79	2328.71	CN5	639.99	2099.70
CW6 - BW3	970.50	3184.06	CE6 - BE3	970.50	3184.06	CN6 - BN3	875.07	2870.96
CW7	1264.35	4148.13	CE7	1264.35	4148.13	CN7	1140.03	3740.26
CW8 - BW4 - AW2	1589.92	5216.27	CE8 - BE4 - AE2	1589.92	5216.27	CN8 - BN4 - AN2	1433.58	4703.35
CW9	1946.03	6384.61	CE9	1946.03	6384.61	CN9	1754.67	5756.79
BW5	2331.65	7649.77	BE5	2331.65	7649.77	BN5	2102.37	6897.54
BW6 - AW3	3188.09	10459.61	BE6 - AE3	3188.09	10459.61	BN6 - AN3	2874.59	9431.07
BW7	4153.40	13626.64	BE7	4153.40	13626.64	BN7	3744.98	12286.68
BW8 - AW4	5222.90	17135.50	BE8 - AE4	5222.90	17135.50	BN8 - AN4	4709.31	15450.49
BW9	6392.69	20973.39	BE9	6392.69	20973.39	BN9	5764.08	18911.02
AW5	7659.48	25129.53	AE5	7659.48	25129.53	AN5	6906.29	22658.43
AW6	10472.87	34359.81	AE6	10472.87	34359.81	AN6	9443.03	30981.07
AW7	13643.92	44763.52	AE7	13643.92	44763.52	AN7	12302.27	40361.78
AW8	17157.23	56290.12	AE8	17157.23	56290.12	AN8	15470.10	50754.92
AW9	21000.00	68897.64	AE9	21000.00	68897.64	AN9	18935.00	62122.70

West Arm: *On South Extension of North Arm

East Arm: *On South Extension of North Arm

North Arm: *Adjusted to allow tracks to pass through.

Table II-2: Expected system parameters
(27 antennas, 1 hour, 1 kHz, 0 taper)

λ(cm)	21	18	6	2	1.3
ν(GHz)	1.35-1.72	1.35-1.72	4.5-5.0	14.4-15.14	22-24
T_s (°K)	50	50	50	300	500
β_o (D) "	52	45	15.5	5.0	3.2
β_o (C)	16	14	4.7	1.52	0.97
β_o (B)	4.8	4.1	1.43	0.46	0.30
β_o (A)	1.46	1.26	0.44	0.14	0.09
σ(D) °K	4.3	4.3	4.3	26	43
σ(C)	47	47	47	280	467
σ(B)	504	504	504	3026	5043
σ(A)	5442	5442	5442	32652	54421

Table II-3: The effects of taper

Taper db	κ	(β/β_o)
0	1.00	1.00
5	0.57	1.28
10	0.40	1.52
15	0.33	1.72
20	0.28	1.90
25	0.25	2.60
30	0.23	2.21
35	0.21	2.36

T_s = system temperature

$\Delta\nu$ = bandwidth in kHz

t = observing time in hours

λ = wavelength in cm

β_o = synthesized beamwidth at zero taper in "

γ = array spacing in "

β = synthesized beamwidth in "

G = slowly-varying function which puts σ in °K

and where "zero taper" is defined as the Fourier transform weighting in which each cell is "sampled" either once or not at all. The effect of adding gaussian taper to this weighting may be estimated from Table II-3. The noise values in Table II-2 are for 27 telescopes, zero taper, 1 kHz, 1 hour, and γ/β = 0.25.

2. Potential VLA Line Sources

Taking these system parameters as given, let us review the potential areas of spectral line research with the VLA.

(a) Neutral hydrogen (1420.405 MHz):

(i) Line interferometers have been very profitably used to synthesize the HI distributions of external galaxies with spatial resolutions of 2' to 4' (e.g., Rogstad and Shostak 1971, Rogstad, et al. 1973, 1974b, Emerson 1974) and of 25" (e.g., Allen, et al. 1973, Rots, 1974). Peak brightness temperatures of 20-50°K are found, but interesting outer regions of the observed galaxies have peak brightness temperatures less than or on the order of 5°K. Even more extended regions need to be observed in order to determine the full rotation curves and total masses of galaxies. These outermost regions will have brightness less than 1°K.

In addition to high sensitivity, observations of external galaxies will often require considerable dynamic range. This requirement arises because among the more interesting galaxies are those containing strong continuum sources.

The total widths of extragalactic emission features are generally in the range 200-500 km s^{-1} (1.0 - 2.5 MHz). The useful single channel bandwidth depends on the synthesized beamwidth (measured in kpc at the source). If this beamwidth is large, then detailed information on velocity features will not be of interest since the many narrow spectral features within the beam will be blended together. Single channel bandwidths of 10-20 km s^{-1} (50-100 kHz) have been used in such observations. For nearby galaxies, rather more detailed velocity information will be desirable. Five km s^{-1} will be quite adequate to observe systematic "peculiar" motions such as those arising from spiral density waves while 2 km s^{-1} will provide a wealth of information on the relative motions of the large gas clouds embedded in the basic galactic structures.

Although interstellar HI has structures as small as 1 pc, signal-to-noise considerations show that such resolutions will not be useful for external galaxies. Resolutions of 20"-90" (C and D configurations) will be useful and will provide resolution as small as 70 pc on the nearest galaxies (M31 and M32). With 20" resolution, galaxies closer than 200 Mpc will be at least partially resolved. Galaxies come in all angular sizes up to about 5° (M31) with many interesting ones in the range of 4'-20'.

(ii) Line interferometers have also been used to synthesize the distribution of the neutral hydrogen within the Galaxy which is seen in absorption in the spectra of galactic and extra-galactic continuum radio sources (Greisen 1973). The brightness temperatures are determined by the background sources and may be extremely high. However, because some absorption lines are very deep ($\tau \gtrsim 5$) and because the distribution of optical depth ("τ") across fields containing a number of weak sources is of interest, high sensitivity observations will also be performed. Since the signal-to-noise ratios determined by the receivers and sampling will often be very high and since the low τ portions of absorption spectra are important in describing the intercloud medium, very accurate calibration (<1%) will be required for this type of observation.

HI absorption profiles typically contain some narrow components which are partially blended at 4 kHz (0.8 km s^{-1}) single-channel bandwidths. A single-channel bandwidth of 1 kHz (able to measure a 6°K thermal broadening) appears adequate to resolve all spectral features. Absorption features probably occur only within the velocity range of Galactic HI emission. This velocity range is typically 150-250 km s^{-1} (0.7-1.2 MHz) in the Galactic plane and <50 km s^{-1} (0.3 MHz) at high latitudes.

Scale lengths of 1 pc are common in the absorbing HI medium. However, concentrations with diameters <0.3 pc are apparently rare (Greisen 1976). Thus resolutions of 20" will be adequate for most absorption observations. Resolutions on the order of 60" will be used for observations of nearby (<2 kpc) HI at the higher latitudes. These angular resolutions are similar to those obtained on molecular lines at millimeter wavelengths. Thus direct comparisons of maps of the different species will be possible. Of course, some projects will use angular resolutions of a few seconds of arc to continue to look for very short scale lengths and to observe distant hydrogen features. Such projects will, undoubtedly, employ intense, moderately small angular diameter background sources. The field of view needed in absorption observations depends principally on the background sources and will range from 2' to 30' or so. With weak background sources it will be necessary to map the full single-dish beam area in order to determine the contribution of the HI emission to the observed brightness temperatures.

(iii) Galactic HI emission has not been mapped with line interferometers. The peak brightness temperatures (with 10' beamwidths) seldom exceed 110°K and are normally 10-50 °K. The emission is currently believed to arise partially from a smoothly-distributed hot gas and partially from discrete cold clouds. Thus, the interesting spatial structure of small angular size in emission may have amplitudes of only a few °K. Preliminary data of Greisen, Cram, and Lockman (in preparation) indicate that intense small-scale structures do occur, however.

Current, low spatial resolution data indicate that velocity resolutions as poor as 2 km s^{-1} (10 kHz) are normally adequate to resolve all spectral structure in HI emission. However, narrow lines are found at times and are likely to be more common at VLA spatial resolutions. The absorption data indicate that frequency resolutions <1 kHz will not be useful. Emission features have total bandwidths of 150-250 km s^{-1} (0.7-1.2 MHz) in the Galactic plane and 50 km s^{-1} (0.3 MHz) at high latitudes.

Galactic HI emission will always fill the full field of view of the single-dish telescopes and, hence, we will always have to map the full field. Because of the need for high sensitivity, these observations will mostly be made with low resolutions (~60"). The HI absorption data indicate that such resolutions will be adequate at least for relatively local hydrogen.

(b) Water molecule: transitions between

7-6	22235.045	MHz
6-5	22235.078	
5-4	22235.121	

The strongest known radio sources, over limited frequency ranges, are the water line maser sources. Water masers are found to be associated with HII regions particularly those containing small-diameter infrared sources and exhibiting emission from other molecular species. A 265-meter baseline, two-element interferometer has been used to show that the H_2O emission arises in many sources from essentially (but not exactly) the same regions as OH maser emission (Hills, et al. 1972). Individual H_2O features are found to have peak fluxes ranging from a few to several thousand Jy and to be unresolved with beamwidths <<1". Thus, since sensitivity will not be a problem with these objects and since these objects exhibit considerable variability in time scales of weeks or less, (Knowles, et al. 1969a), programs to monitor the changes and structure of the maser sources will be of interest. Some H_2O features are also found to be linearly polarized (Bologna et al. 1975).

H_2O emission normally arises over velocity ranges less than 20 km s^{-1} (1.5 MHz). However, several sources have wider total bandwidths with the maximum being 300 km s^{-1} (22.5 MHz) for W49 (Knowles, et al. 1969b). Typical single-channel bandwidths used for H_2O observations have been 30 to 50 kHz (0.4-0.7 km s^{-1}). However, since some features are very narrow ($\lesssim$1 km s^{-1}) and since there is some evidence for the presence of all 3 lines listed above (ibid.), single-channel bandwidths around 7 kHz (0.1 km s^{-1}) or less will occasionally be desirable.

Individual maser regions have been found by VLB observations to consist of several components each $\lesssim 0\overset{''}{.}0005$ in diameter distributed over regions ~1" in diameter (Johnston, et al. 1971). Because of the limitations of VLB data, each spectral component is assumed to arise from a single location. The VLA will allow this moderately arbitrary assumption to be tested with considerable sensitivity. Some HII regions contain more than one maser region with the total source extending over $0\overset{'}{.}5$ to 3' (Johnston, et al. 1973).

(c) Hydroxyl radical (OH)

(i) Ground state $O^{16}H$ transitions:

1612.231, 1665.401, 1667.358, 1720.533 MHz

(i.a) A considerable number of HII regions exhibit maser emission in the basic 18-cm lines of OH. The OH masers typically have peak fluxes in the range 4-20 Jy although much higher fluxes (~500 Jy) are occasionally found (e.g., Robinson et al. 1971). Over half of the emission features are found to be highly polarized with circular polarization predominating but with high degrees of linear polarization also occuring (Ball and Meeks 1968). The spectra in the different polarizations and in the different hyperfine-splitting lines (listed above) show little relationship in most sources. Some emission features are found to vary over time scales of weeks to months (Weaver, et al. 1968).

Maser lines are often quite narrow with full widths $\lesssim$5 kHz (1 km s^{-1}). Single channel bandwidths of 5 kHz have been profitably used to observe OH masers, but bandwidths $\lesssim$0.5 kHz will be needed to fully resolve the spectral features. Total bandwidths of emission features tend to be quite small (<50 kHz), but some sources have features extending over a range of 30 km s^{-1} (150 kHz) or more.

VLB observations have shown that individual maser features have angular diameters much smaller than the minimum beamwidth available with the VLA (Moran, et al. 1968). However, the features tend to cluster in regions a few arc seconds in diameter. Some HII regions have two or more such clusters separated by several minutes of arc (Hardebeck, 1971).

(i.b) Weak OH emission also arises in dark dust clouds (Turner 1973). Peak brightness temperatures at 1667 MHz are found to be ~1 °K but 0.5 °K peaks are more common. The ratios of the hyperfine-splitting lines are nearly, but not exactly, those of thermodynamic equilibrium. The emission is believed to be unpolarized. The emission spectra are narrow (<6 km s^{-1} = 35 kHz) with individual components having linewidths as narrow as 0.3 km s^{-1} (1.6 kHz). Emission is found to extend over most of the dust clouds (i.e. several degrees) but to show significant spatial variation on scales <15' (Turner and Heiles 1974).

(i.c) The basic 18-cm lines of OH are also commonly found in absorption (Turner 1970). The absorption is associated at times with

HII regions and OH emission, but is also found in the spectra of non-thermal galactic and extra-galactic radio sources. Optical depths, averaged over the background sources, are normally small ($\tau \sim 0.01$-0.2). However, aperture synthesis observations of the Galactic center region have found high optical depths ($\tau > 3$) in clouds of ~4' diameter (Bieging 1974). The background-source brightness temperatures can be very high in the case of non-thermal sources, but interesting Galactic thermal sources, with few exceptions, do not exceed a few tens of °K.

Individual absorption features can be quite narrow (<5 kHz), but are typically 10-30 kHz or more. Single-channel bandwidths of 4 kHz will suffice for many observations and 1 kHz should be adequate to resolve all lines. Total bandwidths of 25-30 km s^{-1} (130-170 kHz) are sufficient for many sources, but bandwidths of 0.6 MHz are required for some. The Galactic center region has absorption extending over 300 km s^{-1} (1.7 MHz).

Bieging (1974) presents evidence that OH absorbing clouds are ~3-5' in diameter. His results are for the Galactic center region and probably do not apply to the Galaxy as a whole. If the 1 pc scale found in HI absorption is applicable, then beamwidths of 20"-60" will, in general, be adequate. The fields of view to be observed depend on the background source sizes and will range from ~4' for non-thermal background sources to ~2° for Galactic HII region complexes.

(ii) Excited level $O^{16}H$ transitions: $^2\Pi_{1/2}$, J = 1/2 at

4660.242, 4750.656, and 4765.562 MHz

Emission from OH in the $^2\Pi_{1/2}$, J = 1/2 excited state has been detected from a few sources (Palmer and Zuckerman 1970, Zuckerman and

Palmer 1970b). The strongest line found has a peak intensity ~5 Jy, but the other detected lines have peak fluxes $\lesssim$1 Jy. Little is known about the spatial distribution of the emitting material, but Gardner et al. (1971b) present evidence for an angular scale of 1!2 ± 0!4 at 4660 MHz in Sgr B2. The spectra at 4660 reveal narrow components (~10-20 kHz) and at least one broad component (300 kHz, ibid.). At 4765 MHz the spectral components are probably not fully resolved with bandwidths of 1 kHz and the total bandwidths are small ($\lesssim$100 kHz). The 4750 MHz line has not been detected. Data on the intensity and spatial distribution of the excited OH are important since they bear on the excitation mechanisms and physical conditions of OH masers. The 4765 line appears to be correlated with the 1720 MHz maser emission of OH (Thacker, et al. 1970).

(d) Formaldehyde (H_2CO)

(i) $1_{11} - 1_{10}$ transition: 4829.660 MHz
(hyperfine splitting: 6 components from -19 to +11 kHz)

Formaldehyde is principally observed in absorption against a wide variety of background continuum sources and against the 3°K background (Zuckerman, et al. 1970a, Palmer et al. 1969). The apparent optical depths, averaged over the background sources, are normally <0.1 although a few sources show higher τ (~1). Aperture synthesis observations show high τ ($\gtrsim$2) in small (20"-60") clouds in the Galactic center region (Whiteoak et al. 1974, Rogstad, et al. 1974a). However, such high optical depths are not typical even for H_2CO clouds associated with giant HII regions (Fomalont and Weliachew, 1973). Formaldehyde has been seen in emission (Downes and Wilson 1974), but the peak fluxes are very low ($\lesssim$0.2 Jy).

The widths of individual spectral components range, typically, from 0.7-10 km s^{-1} (11-160 kHz). At smaller bandwidths, the blurring of the absorption features by the hyperfine structure may be visible (Zuckerman et al. 1970). Many H_2CO absorption spectra contain only one component, but typical total bandwidths for multi-component spectra are <60 km s^{-1} (1 MHz). The Galactic center region has a total bandwidth ~3.5 MHz.

Spatial sub-structure in H_2CO has been found to be at least as small as the narrowest beamwidths (20" x 40") so far obtained (Whiteoak, et al. 1974). However, the ensemble of H_2CO clouds is very widely distributed. Thus, full-field (10') and even multiple-field (in HII complexes) mapping will commonly be performed. Because of the low optical depths, signal-to-noise considerations will probably require the use of beamwidths $\gtrsim$5".

(ii) $2_{12} - 2_{11}$ transition: 14488.479 MHz

(hyperfine splitting: 7 components from -20 to +11 kHz)

Because of the difficulty of short wavelength observations, less is known about the 2-cm line of formaldehyde. Like the 6-cm line, it is found in absorption against both continuum sources and the 3°K background. Optical depths are low with τ ~0.03-0.1 and line widths average ~3 km s^{-1} (150 kHz) (Evans et al. 1975). Total feature bandwidths appear to be similar or somewhat narrower than at 6 cm (e.g., 10 to 60 km s^{-1} = 0.5 - 3.0 MHz). The line shows spatial structure at times on a scale of Evans' (ibid.) beamwidth (1!2) and is found in the spectra of the same sources as the 6 cm line.

(iii) Isotopic species: $H_2\,{}^{13}CO$ 4593.089

The C^{13} isotopic species of formaldehyde has also been observed in absorption (Zuckerman, et al. 1969, Gardner, et al. 1971a). The lines are very similar to those from the C^{12} isotopic species, but with the peak optical depths reduced by factors of >10 due to the lower abundance of C^{13}.

(e) Ammonia (NH_3):	(1,1)	23694.48 MHz
	(2,2)	23722.71
	(3,3)	23870.11
	(4,4)	24139.41
	(2,1)	23098.78
	(3,2)	22834.17

Ammonia emission has been observed from a number of HII regions both in the Galactic center region and in other parts of the Galaxy (Morris, et al. 1973). Peak fluxes are typically 24 Jy although the (3,3) line has a peak of 200 Jy near Sgr B2. Some features have bandwidths as small as 2-3 km s^{-1} (160-240 kHz) while others are as wide as 40 km s^{-1} (3.2 MHz). Total spectrum bandwidths around 50 km s^{-1} (4 MHz) are common while the total bandwidth at the Galactic center is ~125 km s^{-1} (10 MHz). Very little is known about the spatial distribution of the NH_3 emission. There is evidence for a scale length of 2'-3' for the Sgr B2 cloud (ibid. and Knowles and Cheung 1971). The metastable lines are expected to have a moderately wide distribution.

However, the non-metastable lines (e.g. (2,1) above) require very high densities and, hence, should have short scale lengths. Turner (private communication) has found scales ~30" for this line with peak flux ~2 Jy.

(f) Formic Acid (H_2CO_2): $1_{11} - 1_{10}$ transition 1638.805 MHz

Formic acid emission has been detected in the Galactic center region (Zuckerman, et al. 1971). The line is weak (~0.14 Jy) and blended with absorption lines of ^{18}OH. The line width is 23 km s^{-1} (125 kHz). Very little is known of the spatial distribution although there is some evidence for a scale length ~6'. Observations of weak lines in the Galactic center region will require very accurate bandpass calibration because of the strong continuum sources present. The lines are interesting, however, because there is evidence for inversion of the level populations ("masering") in formic acid and other complex molecules (such as those listed below).

(g) Formamide (NH_2 - CHO)

(i) $1_{11} - 1_{10}$ transition: F = 1-1 1538.135 MHz

F =	1-1	1538.135 MHz
	1-2	1538.693
	2-1	1539.295
	1-0	1539.570
	2-2	1539.851
	0-1	1541.018

The $1_{11} - 1_{10}$ transitions of formamide have been observed in emission in the Galactic center region (Gottleib, et al. 1973). Peak fluxes ~1 Jy (2-2) and ~0.3 Jy (others) are found. Line widths are ~25-40 km s^{-1} (125-200 kHz) and the spatial extent is $\lesssim 10'$.

(ii) $2_{11} - 2_{12}$ transition:

F =	2-2	4617.12 MHz
	3-3	4618.97
	1-1	4619.99

Emission in the $2_{11} - 2_{12}$ transition of formamide has also been observed in the Galactic center region (Rubin, et al. 1971). Peak fluxes are of order 0.5-1.5 Jy and line widths are 20-40 km s^{-1} (300-600 kHz). The lines are blended with the H112α recombination line. There is no information on the spatial extent of the emission.

(h) Isocyanic acid (HNCO): 1_{01} - 0_{00} transition

21980.52, 21981.46, 21982.06 MHz

The emission of the 1_{01} - 0_{00} transition of isocyanic acid from a cloud near Sgr B2 has been mapped (Buhl _et al._ 1973). The cloud has a peak antenna temperature of 0.4 °K (85' Maryland Point antenna) and a diameter ~5'. Lower resolution observations (Buhl _et al._ 1972) suggest a total flux ~12 Jy and a bandwidth ~25 km s^{-1} (1.8 MHz) which may result, in part, from the blending of the hyperfine components.

(i) Vinyl cyanide (CH_2CHCN): 2_{11} - 2_{12} transition

1371.709, 1371.794, 1371.947 MHz

Vinyl cyanide has been observed in emission in the Sgr B2 molecular region (Gardner and Winnewisser 1975). The peak fluxes are low (~0.04-0.06 Jy) and the feature widths are ~23 km s^{-1} (100 kHz). Information on spatial structure is not available.

(j) Recombination line emission:

A very large number of recombination lines are available to the VLA. These include (i) at 1.3 cm 65α, 66α, and 81β - 83β, (ii) at 2 cm 76α, 95β, and 96β, (iii) at 6 cm 109α - 113α and 137β - 142β, and (iv) at 18-21 cm 156 - 169α and 196β - 212β. The rest frequencies for these and other recombination lines have been given by Lilley and Palmer (1968).

Recombination lines have been extensively observed since about 1964. The only interferometric detections of the lines have been by Sullivan and Downes (1973) who barely detected the H166α line at 21 cm in W3 and by Matthews, Goss, Sullivan, and Wellington (see Habing 1975) who clearly detected and mapped the H109α line at 6 cm in W3 and DR21. The difficulty of interferometric detection results from the weakness of the lines (typically 0.07 - 1.0 °K peak antenna temperature at H109α, Reifenstein et al. 1970). The line to continuum ratio is a function of density, temperature, and frequency. At constant density it is usually found to increase with frequency, but significant enhancements at centimeter wavelengths can occur for high density regions having low electron densities (Dupree and Goldberg 1970).

Recombination lines are found to be quite wide and comparatively simple in spectral shape. Line widths are typically 24-36 km s^{-1} (120-180 kHz at 1420 MHz, 1870-2800 kHz at 23.4 GHz) with the extremes being ~20-100 km s^{-1} (100-500 kHz at 1420 MHz, 1560-7800 kHz at 23.4 GHz) (Reifenstein et al. 1970). Line broadening mechanisms other than normal Doppler shifts are found to be unimportant (Churchwell and Edrich 1970) and, hence, linewidths are proportional to frequency. The interferometer maps at 6 cm have revealed structure in the center velocity of the line with a total variation of center velocity of 27 km s^{-1} over DR21 (30 seconds of arc) and with substructures on scales of a few km s^{-1}. These maps show a variation across the source in the width of the line.

Low frequency recombination lines are thought to arise principally in fairly low density gas and hence to have angular scales similar to the full HII regions. These scales range from 0ʹ.5 to 20'. The high frequency lines, however, are expected to come from areas of high density. Many HII regions contain such areas with angular scales of 3" - 20" (Balick 1972, Webster and Altenhoff 1970). The relative success of the two interferometer experiments cited above bears out these expectations.

This rather long summary is intended to give an overview of the parameters of the line sources currently known to be available to the VLA. Considering the rapid rate at which radio lines are now being discovered, there will undoubtedly be several more lines known by the time the VLA spectral line system comes into operation. The VLA, with its narrow beamwidths and large collecting area, will probably contribute to the discovery of further molecular lines from regions of very high density and small angular size. The VLA should also be very useful in observing molecular lines, particularly OH, H_2O, and H_2CO, in external galaxies.

3. Correlator Parameters

In Table II-4 we summarize the potential areas of research in terms of the desired bandwidths. The first 3 columns identify the line source, columns 4 and 5 give typical and extreme total bandwidths, columns 6 and 7 give our estimates of typical ("useful") and extreme ("occasionally useful") single channel bandwidths, and columns 8 - 10 list the number of frequency channels implied by the numbers in the previous columns. The single channel bandwidths listed represent a compromise between known or suspected spectral structure bandwidths and the need for wide bandwidths to improve the signal-to-noise ratios on weak lines. Let us assume, as is currently intended, that the VLA will employ a recirculating correlator system, e.g.

$$N_{ch}\, B = K \qquad B \geq B_o$$
$$N_{ch} = N_o \qquad B \geq B_o$$

where B is the total bandwidth, N_{ch} is the number of frequency channels, and K, N_o, and B_o are constants of the system. Earlier discussions of the VLA have centered on K = 320 MHz, B_o = 5.0 MHz, and N_o = 64 channels. Table II-4 indicates that, except for the very weak lines, this specification will require most line programs to use between 2 and 7 separate local oscillator settings (and hence several days of telescope time). An alternative specification which uses the same clock rate and correlators would be K = 320 MHz, B_o = 1.25 MHz and N_o = 256. Such a specification would be adequate for a wide range of observing programs. However, it would create serious constraints for observations at 2 and 1.3 cm and for observations of W49, the Galactic center, and other heavily observed, wide-total bandwidth objects. A factor of 2 change in the number of correlators, changing B_o to 2.5 MHz, N_o to 256, and K to 640 MHz would remove a good part of these constraints. This last specification also has the attribute that at 40 MHz total bandwidth, the single-channel bandwidth would be 2.5 MHz which will allow full-field or nearly full-field mapping of continuum sources without the effects of the so-called "delay beam" (D'Addario 1974). In the A configuration at

			total ΔB MHz		single channel Δν kHz		N chan.		
			typical	extreme	typical	extreme	minimum	typical	extreme
HI	1420	ext gal	1.0	3.0	25-50	10	20	120	300
HI	1420	gal abs.	0.5	1.5	4	1	125	500	1500
HI	1420	gal emis.	0.5	1.5	10	2	50	250	750
H_2O	22235	(emis.)	1.5	23	40	7	40	215	3300
OH	~1665	maser emis.	0.1	1	4	0.5	25	200	2000
OH	"	emis.	0.05	?	2	0.5	25	100	—
OH	"	abs.	0.5	1.8	4	1	125	500	1800
OH	4660	(emis.)	0.1	0.5	4	0.5	25	200	1000
H_2CO	4829	(abs.)	1.0	3.5	8	2	125	500	1750
H_2CO	14488	(emis.)	2.0	6.0	50	10	40	200	600
NH_3	23700	(emis.)	4	10	80	25	50	160	400
H_2CO_2	1638	(emis.)	0.2	?	30	10	10	30	—
NH_2CHO	1539	(emis.)	0.3	?	30	10	10	30	--
NH_2CHO	4618	(emis.)	0.6	?	75	20	8	30	--
HNCO	21981	(emis.)	2.0	?	200	50	10	40	--
CH_2CHCN	1371.7	(emis.)	0.1	?	30	10	5	10	—
HII	21 cm	(emis.)	0.2	0.6	25	4	10	50	150
HII	6 cm	(emis.)	0.5	1.5	60	10	10	50	150
HII	2 cm	(emis.)	1.5	4.5	180	30	10	50	150
HII	1.3 cm	(emis.)	2.5	7.5	270	45	10	60	170

21 cm, this line system would be 7.4 times more sensitive than a 40 MHz continuum system in the detection of point sources at the edges of the primary beam (D'Addario 1975).

The current VLA continuum system has eight 50-MHz channels. Because there are two independent IF systems, the effective total bandwidth is $\lesssim$100 MHz (i.e. K = 800 MHz). Using the continuum clock rate and correlators, a line system could be created with N_o = 256 and B_o = 3.125 MHz. Such a system would be better than the 640-MHz system described above not only because of the higher value of K but also because of the greater flexibility inherent in the independent IF systems. In an appendix to this report, it is argued that an even larger system (K = 1600 MHz) is required for full-field continuum observations and for line observations of clusters of galaxies. In any case, some line programs will require the use of two or more separate local oscillator settings ("velocity offsets"). Provision must be made throughout the VLA line system for the interleaving of observations with different velocity offsets.

Because of the narrow bandwidths involved in line observations, receiving system noise is almost always a very serious problem. Everything reasonable should be done to reduce the noise for line observations. One-bit quantization is significantly noisier than three-level or other multiple-bit procedures. The noise computations of Table II-5 are based on the three-level scheme used by the current VLA continuum system. As shown by Table II-4, some line programs do not require large numbers of frequency channels. For such programs, users would happily sacrifice the number of channels for improved signal-to-noise ratio. Such schemes as oversampling, sampling both RR and LL polarizations on unpolarized sources, etc. should be considered. Any schemes which become implemented will have to be optional and, preferably, under computer control in order to keep the wider bandwidth, high frequency resolution, and polarization capabilities available.

4. Fields of View

Table II-5 represents the range of angular resolutions which are likely to be used with the spectral line VLA. In the table columns 1-3 describe the line source, columns 4-6 give a VLA configuration, taper, and resulting beamwidth, columns 7-8 list typical and extreme source sizes (where >SD means sources fill or exceed the single-dish field of view), columns 9-10 give the required numeric fields of view (to a power of 2 with $\gtrsim$ 4 points/beam), column 11 lists possible single-channel bandwidths, columns 12-15 list the 1 σ noise levels which should result in 12 hours for 10 and 27 antennas (in units of milli-Janskys/beam and °K), and column 16 lists comments on how good the signal-to-noise ratios are expected to be. The anticipated extreme numeric fields of view are, with three exceptions, $\lesssim$ 512. Of the exceptions, the OH and H_2O masers have small source complexes within fairly large fields. Such sources should be mapped at low resolution (e.g. since the sources are strong and time variable, by using very high taper) to determine which limited areas should be mapped with full resolution. Note that the direct Fourier transform must be used in mapping such limited areas in order to avoid the aliasing of other emission regions into the small map area. The other exception (HI absorption) is likely to be the result of an overestimate of the background source size over which high resolution will be desired. The noise figures encompass the range likely to be

Table II-5. Possible beamwidths, map areas, map noise

Species	ν	Object	Config.	Taper db	Beam "	Source Typical '	Size Extreme '	Map size Typical	Map size Extreme	Δν kHz	Then σ in 12 hours (@ T_A = 0°K) 10 Ant m. Jy	27 Ant m. Jy	10 Ant °K	27 Ant °K	"S/N" (27 Ant.)
HI	1420	ext. gal	D	5	66	10	>SD	64	128	50	2.12	0.76	0.28	0.10	good
			C	5	20			128	512	10	4.75	1.70	6.81	2.44	ok - inner regions
HI	1420	Gal. abs	C	15	28	3	20	32	256	4	7.88	2.82	6.23	2.23	good - excellent
			B	15	8			128	1024	1	15.72	5.63	134.6	48.61	good - excellent
HI	1420	Gal emis.	D	10	79	>SD	--	128	--	10	4.72	1.69	0.45	0.16	good
			C	5	20			512	--	2	10.64	3.81	15.2	5.46	poor
H_2O	22235		C	10	1.5	0.05	>SD	16	512	40	23.54	8.43	23.9	8.56	excellent
			A	5	0.12			128	4096	7	56.97	20.4	948.	339.6	excellent
OH	1665	maser emis.	B	5	5.2	0.50	10	32	512	4	7.51	2.69	116.3	41.63	excellent
			A	5	1.61			128	2048	0.5	21.28	7.62	3550.	1271.	very good
OH	1665	emis.	D	15	77	>SD	--	128	--	4	7.88	2.82	0.58	0.21	poor
			A	5	1.61			128	2048	0.5	22.31	7.98	17.6	6.30	bad
OH	1665	abs.	D	10	68	5	>SD	32	128	4	7.46	2.67	0.70	0.25	good (non-thermal)
			C	5	18			128	512	1	15.05	5.39	21.6	7.72	poor (thermal)
OH	4660		D	15	27	2	?	32	?	4	7.88	2.82	0.59	0.21	good
			C	5	6			128	?	0.5	21.28	7.62	30.5	10.91	ok
H_2CO	4829		D	15	27	5	>SD	64	128	8	5.56	1.99	0.42	0.15	as OH absorption
			C	5	6			256	512	2	10.64	3.81	15.2	5.46	good (non-thermal only)
H_2CO	14488		D	15	9	$\gtrsim$SD	--	128	--	50	13.35	4.78	0.98	0.35	poor - ok
			C	15	2.6			512	--	10	29.86	10.69	53.0	18.97	poor
NH_3	23700		D	15	6	0.5	>SD	32	128	80	17.59	6.30	1.28	0.46	ok
			C	5	1.2			128	512	25	30.08	10.77	43.1	15.43	poor
H_2CO_2	1638		D	15	77	6	?	32	?	30	2.88	1.03	0.21	0.076	poor
			C	15	24			64	?	10	4.97	1.78	3.94	1.41	bad
NH_2CHO	1539		D	15	83	10	?	32	?	30	2.88	1.03	0.21	0.076	poor
			C	10	23			128	?	10	4.72	1.69	4.78	1.71	bad
NH_2CHO	4618		D	15	27	$\gtrsim$SD?	?	128	?	75	1.82	0.65	0.13	0.048	poor
			C	15	8			512	?	20	3.52	1.26	2.79	1.00	bad
HNCO	21981		D	15	5.5	>SD	--	128	--	200	11.12	3.98	0.81	0.29	bad
			C	15	1.7			512	--	50	22.26	7.97	17.7	6.32	bad
CH_2CHCN	1372		D	15	90	10?	?	32	?	30	2.88	1.03	0.21	0.076	bad
			C	5	20			128	?	10	4.75	1.70	6.81	2.44	bad
HII	~1420		D	10	79	4	>SD	16	128	25	2.99	1.07	0.28	0.10	poor
			C	5	20			64	512	4	7.51	2.69	10.8	3.86	bad
HII	5008		D	10	24	4	>SD	64	128	60	1.93	0.69	0.18	0.065	poor - ok
			C	5	6			256	512	10	4.75	1.70	6.81	2.44	bad
HII	14690		D	10	4.9	>SD	--	128	--	180	6.67	2.39	0.61	0.22	poor ?
			C	5	2			512	--	30	16.48	5.90	23.6	8.45	bad
HII	23404		D	10	4.9	>SD	--	128	--	270	9.08	3.25	0.87	0.31	poor ?
			C	5	1.24			512	--	45	22.43	8.03	32.1	11.50	bad

found since we've put the wider beamwidth with the wider bandwidth and the narrower beamwidth with the narrower bandwidth. The noise values look very good for HI, H_2O, most OH, most H_2CO, and some NH_3 observations. The recombination lines and other molecules look less promising. However, in the latter case, the angular extents of the line sources (which make them seem unpromising) have been deduced from wide beamwidth data and are very uncertain.

From the table and preceding discussion it is clear that many proposed sources will fill or exceed the primary beam (single dish field of view). This fact imposes additional requirements on the VLA line system. Extremely accurate antenna pointing will be required. (See Berge and Greisen 1969 for a discussion of the effects of poor pointing.) Maps will need to be corrected for the antenna pattern of the single-dish telescopes. To avoid degrading the accuracy of the maps, this antenna pattern must be known to the accuracy of the maps (i.e. $\lesssim 1$ %). The VLA does not provide adequate coverage of the u,v plane at baselines less than about 40 meters. The absence of zero-spacing information causes a complete loss of information concerning the "zero level" of the maps while the absence of short-spacing data causes an extended and significant negative bias in the central portions of the synthesized beam pattern. Both effects are detrimental to the mapping of extended objects, particularly spectral line sources. Thus single-dish observations must be included in the synthesis process. Although smaller dishes would provide useful data, the full short-spacing data can be recovered only from maps made with antennas greater than 80 meters in diameter. To recover these data, the antenna pattern of the large single dish must be accurately known. The single-dish observations must be made in the total power mode (except for those cases in which no use is to be made of continuum maps) and must be calibrated using assumptions similar to those made for the regular VLA data. We may visualize the 140-foot and 300-foot antenna being routinely used to provide the single-dish data in advance of scheduled VLA observations.

The effects of the sampling time (i.e. the on-line integration time) have not been properly analyzed. Clark (1973) has suggested the criterion that the longest baseline should move through no more than half a cell during the integration time. He derives the formula

$$S \simeq 13750/M$$

where S is the sampling time in seconds and M is the one-dimensional numerical field of view. Under this criterion the fundamental sampling time of the VLA line and continuum systems must be less than the 3.3 seconds appropriate to full-field mapping in the A configuration ($M = 4096$). For the smaller numeric fields of view appropriate to <u>most</u> line problems, the data may be averaged over longer intervals before being stored by the real-time computing systems. D'Addario (1974) derives a criterion for the number of data samples which suggests that the fundamental sampling interval should be as low as 1.5 seconds.

5. Accuracy

We have already discussed the anticipated noise on synthesized maps due to the system temperatures and the methods by which such noise may be reduced. We will now consider the accuracy of those portions of the VLA mapping system which follow the antennas and RF amplifiers. It is believed that the atmosphere will limit the ultimate accuracy to ~1°/GHz in phase and that sidelobes in the synthesized beam pattern will limit map accuracies to ~1%. The specifications on various portions of the VLA common to both line and continuum problems (e.g. the LO chains) have been set accordingly. The VLA line system should retain these standards.

However, since atmospheric effects, LO instabilities, and sidelobes are broadband problems, the VLA line system may be able to do rather better than the above specifications. This is true only for spectral line sources (of which there are many) in which the line to continuum ratio is small. In current line interferometers it has been found that the relative complex gains of the narrow channels are considerably more stable (and hence able to be calibrated more acurately) than are the absolute complex gains. Considerable effort should be expended to insure that this is the case for the VLA as well. A reasonable specification for the accuracy of the relative calibration is <0.1%. All data processing steps which affect the relative accuracies of the narrowband maps must also adopt this specification or a more stringent one.

6. Polarization

Among the most promising of the line sources are the OH and H_2O maser emission regions. The various spectral components have been found to tend to be highly polarized with up to 100% linear or circular polarization. For these objects to be properly studied, the VLA line system must be capable of mapping all of the polarization parameters. Because of the high data rate of the VLA line system and because most of the other line observations will be of essentially unpolarized objects, it does not seem reasonable to require all four polarizations (RR, LL, RL, LR) to be recorded simultaneously. Because the calibration and analysis of polarization data taken at different times is less accurate and more difficult, it is highly desirable to allow the user the option of swapping spectral channels for additional polarizations. In any case, since the OH and H_2O maser regions contain features of different polarizations and since they often vary on time scales as short as a week, the VLA line system should be capable of switching between the polarizations on a fairly short time scale and, preferably, under computer control. If some of the other molecular lines are found to be masering, then polarization observations will undoubtedly be attempted on them as well.

Polarization observations of HI absorption features may also be attempted. Verschuur (1969) has reported Zeeman splitting in the HI absorption lines of Cas A and other sources. His results have been called into question by a reappraisal of the effects of HI spatial structures in the polarized side lobes of his antenna and by the discovery of spatial structures in the absorbing medium which are much smaller than his beamwidth. The high directivity of the VLA will make the Zeeman experiment worth trying again.

B. DATA MANIPULATION

The subject of data manipulation separates into two distinct areas: (1) the control of observations as they are being made and (2) the manipulation of data after they have been recorded. The latter area includes the subjects which we will call quality control, synthesis, analysis, interpretation, and publication. This section will be subdivided along these lines.

(1) Real time data manipulation

The most fundamental aspect of real time data manipulation is the users' control over the parameters of the observation. The user must be able to control, in as simple a manner as possible,

(a) the source position (α, δ and their epoch),

(b) the duration of the observation,

(c) the observing frequency (e.g., the line rest frequency, the velocity offset, and the rest frame in which velocities are measured),

(d) the total bandwidth,

and (e) the polarization (RR, LL, RL, and/or LR).

These functions should be under full computer control so that the user may interleave observations at different polarizations and/or velocity offsets. Such interleaving will be required in some programs by instrumental limitations on the number of simultaneous spectral channels and polarizations. The choice of calibration procedures should be up to the user only after consultation with VLA advisory personnel. Considering the cost and complexity of the VLA, user interaction with the real time control program probably should not be direct and instantaneous (e.g. via terminals). Adequate user control is maintained with punched cards submitted to a human operator. This array operator need not be a PhD. astronomer, but he must be able to interact with the user, to advise the user on such subjects as the hardware feasibility of requested changes and the effects of the changes on data calibration.

Real-time user interaction with the data depends on the nature of the data already recorded and currently being recorded. It is assumed that the monitoring of VLA hardware functions will not be of direct concern to the user. However, minor malfunctions may manifest themselves only through peculiarities (which may not be recognized as such by the operator) in the data. Since VLA time will be quite valuable, it is senseless to waste 12-hours integration time on sources which turn out to be significantly weaker or stronger than had been anticipated. Users should have "back-up" sources to substitute in such circumstances. To view the data, the user and array operator should have a variety of real-time, or nearly real-time, displays available.

Perhaps the most basic such display is a hard-copy (printer, chart recorder) display of phase and amplitude as a function of time. The user should be able to select, in as simple a fashion as possible (thumbwheels?), at least a single spectral channel or a single sum of consecutive spectral channels. Another basic and very useful display is a plot (using a CRT or other temporary medium) of the current visibility function spectrum from a selected interferometer pair. Very desirable attributes for both of these display procedures are:

(a) to have a number of such displays for separate antenna pairs,

(b) to allow rapid user selection of which interferometer pairs are displayed (via selector switches for example),

and (c) to allow the data, particularly the spectra, to accumulate (average) for some period of time.

Attribute (a) seems particularly important. To properly monitor an interferometer pair the user will have to watch it for several minutes. To monitor the full VLA in twelve hours with a single display, the user could devote only 2 minutes to each interferometer pair. Three is probably the minimum useful number of display devices.

One of the best methods to judge the performance of the VLA is to look at its ultimate output - maps. A gray-scale display of limited spatial (e.g. 128 x 128 or 256 x 256) and intensity (e.g. 5-6 bits) accuracy would provide all the information the user needs in real time. The user should be able to select a single spectral channel, a single sum of consecutive channels, or a single channel minus the average of a set of consecutive channels for the display. It is acceptable to require the user to make this selection once at the beginning of the observations of each source. This display need not be updated continuously but could be updated at discrete intervals (e.g. at 0.25, 0.5, 1, 2, 6, and 12 hours after the start of the observations).

(2) Post-observation data manipulation

The VLA real-time system will produce an enormous mass of data which, because of its size and its location in the (u,v) plane, will be virtually unintelligible. Powerful computing systems are required to calibrate, edit, sort, and Fourier transform the data. However, particularly in the spectral-line case, the maps produced at this stage of the data manipulation are still essentially useless. Before the maps may be interpreted in astrophysical terms and published, the user must be able to comprehend them, to extract global and detailed information from them, and to describe and display them in formats suitable to the interpretations. Powerful computing systems are required at these stages of the data manipulation as well. It is important to recognize that, in order to realize the full potential of the VLA, computing facilities must be available for large quantities of processing on the direct output maps.

In the following paragraphs we will divide the subject of post-observation data manipulation into five areas which we call quality control, synthesis, analysis, interpretation, and publication. These subsections are only partially distinct. Information obtained during later stages of processing may cause users to return to earlier stages for additional processing. These subsections will contain discussions of data manipulation procedures currently known to be useful. It is our feeling that, within reason, the VLA data processing system eventually should provide all of these procedures. Any new or improved procedures will, of course, be welcome. The VLA should initially provide a very basic set of routines and then expand the set over the years. All of these processing areas have numerous display modes in common. We simply list them here:

i. print numbers (line printer, CRT)

ii. print numbers as symbols (line printer)

iii. display profiles as function of any 1 coordinate (line printer, plotter, CRT)

iv. contour map of low quality (plotter, CRT, film, TV)

v. gray-scale display (film, TV)

vi. ruled surface display (plotter, CRT, film, TV)

Display modes iv, v, and vi should also have high quality versions for publication purposes. A final subsection will contain discussions of some operating considerations.

(a) Quality control

By quality control we mean the data manipulation actions of calibration and editing. This term is meant to emphasize our fundamental assumption that, because of the enormous quantity of data produced by the VLA, almost all calibration and editing will eventually be done by automatic procedures operating in a combination of the synchronous and asynchronous systems. These procedures will apply all the usual phase corrections (for atmosphere, nutation, source position, baseline parameters, unequal LO electrical lengths, receiver drifts) and gain corrections (for clipping, atmosphere, receiver drifts) properly accounting for the fact that each channel is at a different IF and RF. The calibration of the complex gains of the channels will be done (hopefully) to the relative accuracy specified previously ($\lesssim 0.1\%$). Provisions for the calibration of polarization data will also be made. Those data affected by detected hardware malfunctions will be deleted.

The user's interaction with his data at this stage will almost always be restricted to checking the results of the automatic procedures. The user will need to view both the calibrated data and the data taken on calibration sources. Routines to display these data en masse (e.g. gray scale plots on baseline-time and u,v axes) are needed to locate questionable data and detailed display routines (e.g. printed numbers, spectra) are needed to isolate the defective data points. Users are not likely to want to delete individual spectral channels. Narrowband terrestrial interference will produce ringing throughout the spectrum while correlator problems occur in the lag domain and, hence, affect all channels. Thus users will require editing routines which make it convenient to delete individual baselines and times, but not individual frequencies.

Occasionally the data will need to be recalibrated as a result of bad data points, unanticipated discontinuities, and/or abnormal system time constants. If the automatic calibration routines are sufficiently flexible, then the re-edited data may simply be run through suitable portions of the automatic routines using the revised information on system time constants and discontinuities. The user may also wish to recalibrate his data if there is a single strong point source of either line or continuum emission in the field. More accurate calibrations of all parameters (except the relative complex gains of the narrow channels) may be obtained by using such a source as a continuous calibration. The capacity to perform this type of calibration is quite desirable, but, since it is unclear how often the situation will arise, it should probably be given low priority.

Since permanent displays are normally not required at this stage of the data manipulation, most of the editing and some calibration processes are suitable to interactive processing. However, decisions on calibration parameters such as smoothing time constants may require the examination of bodies of data too large for thoughtful comprehension on the time scales of interactive processing. Thus, much of the determination of calibration parameters and all of the application of those parameters to the data may be done in batch processing modes.

(b) Synthesis

By synthesis we mean the processes of data manipulation which convert calibrated visibility function data to properly registered and corrected maps. The principles of aperture synthesis have been extensively discussed in the VLA Scientific Memorandum series. However, since confusion on a number of fundamental points appears to remain, we will briefly review the subject here. An interferometer does not measure the true visibility function V but rather an average (over finite bandwidths, integration times, and antennas) of V at some discrete locations in the u,v plane. For simplicity let us neglect the effects of bandwidth and integration time and note that the effects of finite antenna size are equivalent to a multiplication of the sky brightness temperature by the power pattern of the antennas. Then, assuming a classical Fourier transform approach, our sampled visibility function is given by

$$V' = V \cdot S$$

where the sampling function may be expressed as the weighted sum of delta functions

$$S = \sum_{i=1}^{N} W_i \, {}^2\delta(u-u_i,\ v-v_i).$$

To use the FFT algorithm, we must estimate V' at the intersections of a rectangular grid. This creates a sampled, smoothed, and re-sampled visibility function V" given by

$$V'' = III \cdot (C * (V \cdot S))$$

where * represents a convolution, C is a convolving function, and III is the re-sampling function given by

$$III = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} {}^2\delta(u-j\Delta u,\ v-k\Delta v).$$

It is V" to which we apply the FFT to obtain

$$T'' = \overline{III} * (\overline{C} \cdot (T * B))$$

where T", $\overline{III}$, $\overline{C}$, T, and B are the Fourier transforms of V", III, C, V, and S, respectively. In mapping the full field or partial, uncrowded fields we may usually ignore the effects of the convolution with $\overline{III}$ ("aliasing") and correct T" for $\overline{C}$ to obtain

$$T' = \frac{T''}{\overline{C}} \simeq T * B.$$

The direct Fourier transform yields T' = T * B exactly. Except for possible interactions with the aliasing, the choice of the convolving function is irrelevant. For computational speed, the VLA will probably use the "pill-box" function

$$C(u,v) = 1/(\Delta u \ \Delta v) \qquad |u| < \Delta u/2 \text{ and } |v| < \Delta v/2$$

$$C(u,v) = 0 \qquad \text{otherwise}$$

although other functions are more effective in suppressing aliasing. The pill-box function makes the smoothing and resampling operation very similar to an averaging (i.e. estimating the true visibility function) operation, but the similarity is coincidental. Thus, using the classical techniques, an interferometer array measures not the true brightness distribution, but rather the true distribution (modulated by the single-dish power pattern) convolved with the Fourier transform of the weighted sampling function ("synthesized beam pattern").

We assume that the VLA hardware and software will automatically handle the basic operations of synthesis:

(i) Determination of u,v from t, B, V,

(ii) Setting the scale of the (u,v) plane,

(iii) Gridding (FFT),

(iv) Fourier transform,

(v) Correction for the Fourier transform of the convolving function (FFT),

(vi) Production of the synthesized beam pattern,

and (vii) Re-registration of maps so that maps at all frequencies are evaluated at the same (x,y) points.

User interaction at this stage of the data manipulation will involve only the selection of the data to be synthesized, the weights to be used, and, if necessary, the phase center. The choice of weights (W_i) is important in a general or over-all way, but the details of the exact choice are of little importance. To obtain the narrowest synthesized beamwidth with reasonable sensitivity and beam pattern, the W_i are set inversely proportional to the local density of data samples (roughly $W_i = 1/N$ where N samples occur in the same cell as the i'th sample). To obtain the highest sensitivity (in terms of flux per synthesized beamwidth), the W_i are set inversely proportional to the noise in the data sample (roughly $W_i = 1$ for similar receivers on all baselines). To improve brightness sensitivity and lower the sidelobe levels, the W_i may also contain a tapering function. Considering the distribution of data samples taken by the VLA, it is reasonable to restrict users to weights of 1/N times a tapering function. The VLA system should automatically determine a "map" of the number of samples in each cell (N) for cell sizes suitable for full-field maps. This map may be determined in real-time and then used off-line either directly for full field maps or after smoothing to larger cell sizes for less than full-field maps. Since the signal-to-noise ratios on maps are determined by the degree of tapering, the VLA system must allow the user either to experiment with various tapers or to smooth his maps in the (x,y) plane. In the former case the user would try various tapers on the data from one to, at most, four spectral channels or sums of channels before transforming all the data with the selected taper.

The user must be allowed to select the data to be Fourier transformed. If the synthesis devices have essentially unlimited dynamic range (e.g., software) then this selection will involve only which channels to transform, whether single-dish "visibility function" data should be included, and whether data from the outermost (poorly-sampled) cells should be excluded. However, if the synthesis devices have limited dynamic range (e.g. optical devices), then the user will often have to exercise further choices. In particular, the visibility functions of strong continuum sources will have to be subtracted from the data before the Fourier transform. The user will have to select which spectral channels define this continuum visibility function.

It appears likely that the numeric fields of view for VLA line maps will be limited. Such limitations appear quite reasonable on the basis of the discussions given in part A of this chapter. However, certain OH and H_2O maser sources will be suitable to A-configuration, large-field mapping. Users should be allowed to select, with suitable restrictions, the size of the region within the primary beam which is actually mapped. When limitations on the numeric field of view are important, the user must be allowed to select the phase center or centers to be used in the synthesis process. In this way users could obtain full resolution maps of several small regions within the primary beam. When numeric field of view limits permit, the user could be restricted to full field maps if such maps are desired for archiving purposes.

(c) Analysis

By analysis we mean those processes by which the user examines and becomes familiar with his data, converts the data into physically meaningful formats, and extracts certain fundamental parameters. The first of these areas consists principally of the various display media listed previously. Since the user will be interested in speed with limited accuracy, the examination area is suited to interactive processing. However, he will need to view limited portions of the data with full accuracy. The user will normally desire various displays along one and two-dimensional curves and planes through the three dimensions (x,y,V) of his data. He will carry out various arithmetic operations (e.g. adding to or multiplying maps by constants or other maps). He will need statistical information such as integrals, averages, and r.m.s. fluctuations in maps or parts of maps.

The processes used to convert maps into physically meaningful formats are quite varied. Zero- and very short-spacing data can be added to the maps at this stage if they have not been added during synthesis. Such data should be obtained and prepared (as visibility function data of correct format) prior to the observing run. If the CLEAN algorithm is deemed necessary, it should now be applied to the data. Done by the usual digital methods, this algorithm is both slow and of dubious validity. We may hope that the sidelobes will be sufficiently low that it is not required. A limited version of CLEAN - the "permanent" subtraction of a small ($\lesssim 6$) number of strong point sources - will be needed to increase the dynamic range for weak signals. Following the above operations, large-field maps will require correction for the power pattern of the single-dish telescopes. The coordinate systems used for different phase centers will not be aligned. Therefore, when an object is so large that it has been mapped using several phase centers, an interpolation onto a common grid is necessary in order to put all the observations together. If more than one velocity offset has been observed, the various three-dimensional maps should be combined taking into account the overlaps, if any. Maps may require frequency and/or spatial smoothing in order to improve signal-to-noise ratios. If extragalactic observations contain channels with velocities near zero, then the maps in these channels are likely to be affected by gas in the Galaxy. Estimates of these effects can be obtained from the amplitudes of strong continuum point sources within the external galaxy and the maps corrected accordingly. Of course, spatial structure in the Galactic gas will make such corrections uncertain.

The conversion procedures listed, so far, act to correct the maps for various effects. To convert the maps to physically meaningful units requires additional steps. Polarization maps need to be converted to maps of total intensity, polarized intensities, etc. That part of the maps which is independent of frequency ("the continuum") must be determined using channels known to be free of line radiation or using various window and cut-off methods. The continuum map may then be subtracted from line emission maps or used to convert absorption maps to maps of optical depth. Note that optical depth, velocity, and some polarization maps are not reliably defined over the full map area. The analysis and display of such maps requires a "blanking array" map which divides the array values into the catagories: (1) reliable, (2) too uncertain to use except for interpolations near the edges, and (3) totally useless.

The processes used to extract "fundamental" parameters from the maps are probably as varied as the sources themselves. Some of these processes, known to be useful in a wide variety of cases, are described below. Maps of the predominant velocity in the spectrum (or a portion of the spectrum) are very useful summaries of data. Several methods to define the predominant velocity (e.g. the first moment, at peak line strength, at median line signal) certainly should be available. Maps of the zero (line sum) and second (line width) moments are also useful summaries. Users should be allowed to apply window and cut-off methods to improve the reliability of such maps. Interactive routines to find the flux, position, half-widths, and position angles of source components are often useful. Another analysis technique which is likely to be of general interest is the decomposition of the spectra into sums of Gaussians. These Gaussians give estimates of interesting physical parameters and represent a form of frequency smoothing which reduces the total number of maps. This procedure is likely to be useful in terms of speed and reliability only for comparatively simple spectral distributions.

(d) Interpretation

By interpretation we mean those data manipulation and examination processes by which the user combines his present data with current observational and theoretical astrophysical information in an attempt to understand and describe the implications of his data. The methods of detailed interpretation of VLA line maps will be limited only by the imaginations of observers and theorists. It is clear that the VLA cannot support, with any priority, complicated, special purpose interpretive procedures desired by only a few observers. However, it is also clear that (1) some significant information can be gleaned from the data only through such procedures and (2) most observers will not have, on their own, the funds to create and operate such procedures. We may anticipate the need for a general purpose computing facility supported in part by VLA programmers for both general and special purpose analysis and interpretation of VLA data. This is a very serious problem which deserves further discussion, but it is beyond the scope of this chapter.

(e) Publication

By publication we mean those display procedures which make high-quality representations of VLA data suitable for publication. These procedures could well be implemented as a separate package on a mini-computer since they would be used solely for display of input data with little or no analysis. The minimum set of displays includes (1) a multi-pen plotter for two and three dimensional contour and ruled-surface maps and for one-dimensional profiles and (2) an optical image recording system for properly labeled black and white and color photographs with or without intensity coding and overlayed contours. For publishable displays a good photo lab is required including a very accurate (in both optical qualities and in the setting of reduction factors) horizontal bench camera capable of holding at least 24" x 24" negatives and a large, vacuum operated contact printing machine. For preparing optical overlays, an accurate and easy to use measuring machine and procedure to determine the position of field stars is also highly desirable.

(f) Data-manipulation operating procedures

As explained above, the typical user will spend minimal time interacting with his data in the quality control and synthesis stages. However, all users will require quite a lot of time to look at, and to try to understand, their maps. They probably will need a 1-2 week period after the observations and another week at some later time just for the analysis and publication stages. It is reasonable to limit and to schedule (on a 24-hour schedule) the use of the computing facilities. Two interactive terminals would support 16-12 users in 3-4 hour shifts. Such shifts should be sufficient to exhaust the user and to provide adequate hard copy output to keep him busy for at least the rest of his working day. The time limitations of the user's attention span and his shift on the terminal require the processing programs available to the terminal to be quite efficient. Lengthy computations will have to be run in a batch mode. Because of the time limitations users will have to prepare carefully for their scheduled processor time and should be assisted by menu-prompting procedures (see IDAPS Manual) and by VLA personnel (e.g. "friends" of the VLA available on a 24-hour basis).

C. DATA STORAGE

The data storage problem in the VLA line system is extreme. With 10-second sampling rate, a 256-channel, 27-antenna VLA produces 7.76×10^8 data words (plus associated u,v values, calibration and monitor data, etc.) in 12 hours. If the resulting maps are 1024 x 1024, the map data occupy 2.68×10^8 words in the T_V (x,y) maps alone. The net data rate is then $\sim 2.6 \times 10^9$ words/day. A standard data tape (1600 bpi, 2400 feet) with no record gaps will hold 0.023×10^9 words. Thus, the VLA line system will fill at 75% efficiency 152 tapes per day. This is an extreme data storage problem which we may divide into three parts: (1) active storage, (2) VLA archival storage, and (3) user storage. Although the data storage problem of the line system is more extreme, the following remarks also apply to the VLA continuum system. We should insure that similar standards are applied to both systems.

1. Active storage

By active storage we mean those storage media readily accessible by the VLA data processing computer system. At any one time during data processing that part of the data set currently being used (say maps) resides on some medium such as high-speed disk which can be very quickly accessed by the computer. The remainder of the data set (visibility functions) can be on a slower medium such as digital film, tapes, etc. Because of the limited processing time available to users, the transfer of data from slow to fast media must either be very efficient or be handled off-line by a mini computer. Because of the need to check and perhaps perform asynchronous calibration and editing, the visibility data must be accessible in baseline-time order. The map data base must be designed so that access to various one- and two-dimensional cuts is efficient both in CPU and, more importantly, real time.

2. VLA archival storage

By archival storage we mean those large-volume, principally digital media used by the VLA to store inactive data for indefinite, but long, periods of time. These media must, of course, be computer readable, but the rate at which that reading can take place is not of direct concern to the user. VLA personnel must maintain archived data in such a fashion that they remain readable by currently supported programs. Archived visibility-function data are useful for two operations: (1) recalibration and re-editing after the initial maps have been thoroughly examined and (2) mapping regions within the primary beam not mapped initially. The first operation requires data accessible in baseline-time order, but appears to be useful only for a limited period of time. After the user has had a reasonable time (probably about one year) in which to analyze and publish the data, the baseline-time ordered data can be discarded. The second operation has usefulness over an arbitrarily long period of time and is best handled by archiving either full-field, full-resolution maps or gridded (u,v)-planes.

Archived map data are useful over indefinite periods of time principally for (1) comparison with maps taken at different epochs or frequencies and (2) for re-analysis and interpretation in the light of new ideas and techniques. Ideally the map data should be archived essentially forever. A minimum archival period of five years from the date of observation (i.e. about three years from the date of publication) is needed to allow astronomers adequate time to find out about and use the map data. Archived data should become public property after some suitable proprietary period.

3. User storage

By user storage we mean the various media used by the observer to take his data home. These media will include the hard copy outputs of the data processing system (gray-scale photos, line plots, printer output, CRT-screen copier output) plus large volume digital storage media (tape, film). Many users will not be able to afford sophisticated special purpose VLA data devices (e.g. automatic digital film readers, video tape readers) unless they are very inexpensive, nor will they have access, in general, to the most modern computing devices (e.g. super high density magnetic tape readers). Such users will almost certainly do very little large-volume computing on VLA data at their home institutions. Thus, some users may not require the large-volume digital output. Most users will want magnetic tapes containing at least some of their map data. Because of the wide variety of computing facilities, the VLA must provide these tapes in a wide variety of formats (binary, BCD, EBCDIC, ASCII) and densities. A few observers may have sufficient research funds to make use of their full collection of data. For such observers the VLA should provide more efficient storage media (such as will be used for VLA archival storage) and advice on how to use them.

For observers to utilize VLA data at their home institutions, the data formats must be extremely well documented. Good documentation on such matters as coordinate systems and useful data extraction and display algorithms must also be provided. Such documentation must be clearly written and should not be hidden away piecemeal among a wide variety of reports, memoranda, and the like.

We must also consider the position of NRAO staff scientists as VLA users. The anticipated computing facilities at the VLA site will not support general purpose data analysis programming. Thus NRAO scientists will want to carry some digital data to a general purpose facility such as the one currently in Charlottesville. We must decide whether or not NRAO will have sophisticated data devices (e.g. digital film readers, 6400 bpi tape units) and professional programmers to support the staff scientists. If the NRAO does provide such support we may expect that outside users will also wish to take advantage of the support.

D. REFERENCES

Allen, R. J., Goss, W. M. and van Woerden, H. 1973, Astr. and Ap., 29, 447.

Balick, B. 1972, Ap. J., 176, 353.

Ball, J. A. and Meeks, M. L. 1968, Ap. J., 153, 577.

Berge, G. L. and Greisen, E. W. 1969, Ap. J., 156, 1125.

Bieging, J. H. 1974, unpublished thesis, California Institute of Technology.

Bologna, J. M., Johnston, K. J., Knowles, K. J., Mango, S. A., and Slvanacker, R. M. 1975, Ap. J., 199, 86.

Buhl, D., Snyder, L. E. and Edrich, J. 1972, Ap. J., 177, 625.

Buhl, D., Snyder, L. E., Schwartz, P. R. and Edrich, J. 1973, Nature, 243, 513.

Churchwell, E. and Edrich, J. 1970, Astr. and Ap., 6, 261.

Clark, B. G. 1973, VLA Scientific Memorandum No. 112.

D'Addario, L. 1974, VLA Scientific Memorandum No. 116.

D'Addario, L. 1975, VLA Scientific Memorandum, No. 120.

Downes, D. and Wilson, T. L. 1974, Ap. J., 191, L77.

Dupree, A. K. and Goldberg, L. 1970, Ann. Rev. Astr. and Ap., 8, 231.

Emerson, D. T. 1974, M.N.R.A.S., 169, 607.

Evans, N. J., Zuckerman, B., Morris, G. and Sato, T. 1975, Ap. J., 196, 433.

Fomalont, E. B. and Weliachew, L. 1973, Ap. J., 181, 781.

Gardner, F. F., Ribes, J. C. and Cooper, B. F. C. 1971a, Ap. Letters, 9, 181.

Gardner, F. F., Ribes, J. C. and Sinclair, M. W. 1971b, Ap. J., **169**, L109.

Gardner, F. F. and Winnewisser, G. 1975, Ap. J., **195**, L127.

Gottlieb, C. A., Palmer, P., Rickard, L. J. and Zuckerman, B. 1973, Ap. J., **182**, 699.

Greisen, E. W. 1973, Ap. J., **184**, 363 and 379.

Greisen, E. W. 1975, VLA Scientific Memorandum No. 117.

Greisen, E. W. 1976, to be published in Ap. J..

Habing, H. J. 1975, Space Sci. Rev., **17**, 629.

Hardebeck, E. G. 1971, Ap. J., **170**, 281.

Hills, R., Janssen, M. A., Thornton, D. and Welch, W. J. 1972, Ap. J., **175**, L59.

Image Data Processing System: Users Manual, NASA Marshall Space Flight Center, Huntsville, Alabama, 1975.

Johnston, K. J. et al. 1971, Ap. J., **166**, L21.

Johnston, K. J., Sloanaker, R. M. and Bologna, J. M. 1973, Ap. J., **182**, 67.

Knowles, S. H., Mayer, C. H., Cheung, A. C., Rank, D. M. and Townes, C. H. 1969a, Science, **163**, 1055.

Knowles, S. H., Mayer, C. H., Sullivan, W. T. and Cheung, A. C. 1969b, Science, **166**, 221.

Knowles, S. H. and Cheung, A. C. 1971, Ap. J., **164**, L19.

Lilley, A. E. and Palmer, P. 1968, Ap. J. Suppl., **16**, 143.

Moran, J. M. et al. 1968, Ap. J., **152**, L97.

Morris, M., Zuckerman, B., Palmer, P. and Turner, B. E. 1973, Ap. J., **186**, 501.

Palmer, P., Zuckerman, B., Buhl, D. and Snyder, L. E. 1969, Ap. J., 156, L147.

Palmer, P. and Zuckerman, B. 1970, Ap. J., 161, L199.

Reifenstein, E. C., Wilson, T. L., Burke, B. F., Mezger, P. G. and Altenhoff, W. J. 1970, Astr. and Ap., 4, 357.

Robinson, B. J., Caswell, J. L. and Goss, W. M. 1971, Ap. Letters, 9, 5.

Rogstad, D. H. and Shostak, G. S. 1971, Astr. and Ap., 13, 99.

Rogstad, D. H., Shostak, G. S. and Rots, A. H. 1973, Astr. and Ap., 22, 111.

Rogstad, D. H., Lockhart, I. A. and Whiteoak, J. J. 1974a, Astr. and Ap., 36, 252.

Rogstad, D. H., Lockhart, I. A. and Wright, M. H. C. 1974b, Ap. J., 193, 309.

Rots, A. H. 1974, unpublished thesis, Rijksuniversiteit te Groningen.

Rubin, R. H., Swenson, G. W., Benson, R. C., Tigelaar, H. L., Flygare, W. H. 1971, Ap. J., 169, L39.

Sullivan, W. T. and Downes, D. 1973, Astr. and Ap., 29, 369.

Thacker, D. L., Wilson, W. J., and Barrett, A. H. 1970, Ap. J., 161, L191.

Turner, B. E. 1970, Ap. Letters, 6, 99.

Turner, B. E. 1973, Ap. J., 186, 357.

Turner, B. E. and Heiles, C. E. 1974, Ap. J., 194, 525.

Verschuur, G. L. 1969, Ap. J., 156, 861.

Weaver, H., Dieter, N. and Williams, D. R. W. 1968, Ap. J. Suppl., **16**, 219.

Webster, W. J. and Altenhoff, W. J. 1970, A. J., **75**, 896.

Whiteoak, J. B., Rogstad, D. H. and Lockhart, I. A. 1974, Astr. and Ap., **36**, 245.

Zuckerman, B., Palmer, P., Snyder, L. E. and Buhl, D. 1969, Ap. J., **157**, L167.

Zuckerman, B., Buhl, D., Palmer, P. and Snyder, L. E. 1970a, Ap. J., **160**, 485.

Zuckerman, B. and Palmer, P. 1970b, Ap. J., **159**, L197.

Zuckerman, B., Ball, J. A., and Gottlieb, C. A. 1971, Ap. J., **163**, L41.

CHAPTER III. IMPLEMENTATION

A. INTRODUCTION

In this chapter we consider some possible configurations of hardware which will result in a system meeting most, if not all, of the requirements described in Chapter II. Where possible, we consider details of the implementation in terms of specific devices. We begin by taking the following minimum specifications as given:

Visibility function - 351 baseline channels (one polarization each);
256 spectral channels at 2.5 MHz and smaller bandwidths; proportionally fewer channels at larger bandwidths to at least 20 MHz;
minimum total bandwidth ≤ 250 kHz;

Maps - digital maps or maps convertable to digital form with up to 1024 x 1024 elements for each spectral channel; noise introduced in mapping ≤0.5% of the maximum brightness within each spectral channel, ≤.1% channel-to-channel error.

Further specifications could be given, but for now we take them to be flexible and dependent on equipment availability and cost.*

We assume no changes from the continuum system in the antennas, front ends, local oscillator systems, and IF transmission system. Thus, IF signals at selectable bandwidths of up to 49 MHz will be delivered to the spectral line processor. The processing of these signals divides naturally into a sequence of independent operations,

* It will turn out that the visibility function specifications, given here on the basis of Chapter II results, can be exceeded in several important ways at modest cost; see Section III.C.

whose implementation is discussed in sections C through E. They are:

1. Correlation: Computation of the time-averaged complex product of the IF signals for each antenna pair in each spectral channel. It is assumed that the IF signals will be immediately sampled and digitized, and that the correlation will be done entirely in digital hardware.

2. Calibration and Editing: Each number provided by the correlator is divided by the complex gain of the corresponding narrow-band interferometer to provide a calibrated complex visibility measurement. It may also be necessary to identify and delete data which is unusable due to equipment malfunctions.

3. Mapping: Computation of the maps from the calibrated visibility data. This includes whatever sorting and re-formatting is needed for the mapping procedure employed (e.g., gridding for the FFT).

4. Display: Providing access to the data, primarily the maps, for the user.

5. Analysis: Computational operation on the maps.

B. GENERAL CONSIDERATIONS

1. Storage

As will become apparent in the ensuing discussion, the signal processing after correlation is dominated by data mainpulation and storage, rather than by arithmetic operations. (Most of the arithmetic is in the mapping task, and if we assume that this is done by a linear operation like the Fourier transform, then the computations are easily implemented by analog or digital hardware now available.) Storage requirements become more severe as we progress through the processing: first, the correlator must provide storage for one integration period, typically 10 sec; next, the calibrated visibilities for a particular source may involve as much as 12 hours of data (possibly more for special observations) all of which must be rapidly accessible to the synthesis operation; and finally, the maps for a particular source involve nearly as much data as the calibrated visibilities (although some compression can occur in the mapping process), and storage for at least five years of maps is expected to be needed.

It is important to distinguish several types of storage. Storage can be permanent or temporary (where anything stored on an erasable medium is considered temporary), and we must decide at each stage of the processing how long the data must be retained. Next, storage can be direct-access or sequential-access. By direct-access we mean that the average time required to access one element of a data set is about the same for all elements; thus, magnetic disks and photographic film are usually direct-access storage, but magnetic tape requires sequential access.

The size of data sets which must be kept in direct-access storage is of critical importance to the design of the post-correlation processing system. Most intermediate storage ("buffering") can be sequential, or eliminated entirely, if "pipelining" is used: that is, each operation is performed on a small block of data and the result is passed to the next operation, so that not until the end is it necessary to store a large block of data. Carried to the extreme, this means that the input and all intermediate results are lost, precluding re-processing with altered parameters. To the extent that re-processing is anticipated, sequential recordings can be made at strategic points. For the VLA, strict pipelining would mean that the first output containing data from an entire observation would be a set of maps, and these would be constructed as the observation progresses. Sequential records of correlator output or calibrated visibilities could be kept for archival or backup purposes, but very little fast, direct-access storage would be needed. However, in addition to the difficulties involved in maintaining real-time calibration data (discussed in section D) and possibly in performing real-time reformatting for synthesis (section E), certain observing modes which seem highly desirable would not be possible in such a system. These include what might be called "source-hopping", in which observations of several sources are interspersed through an observing session; "frequency hopping", in which the center frequency or bandwidth is changed periodically in order to extend the frequency coverage; subarray observations, in which the 27 antennas are partitioned into two or more arrays observing different sources; and polarization switching, in which polarization information is obtained sequentially by periodically changing the antenna polarizations.

None of the special modes just mentioned alters the data rate, but each greatly increases the complexity of the data organization, compared with the simple situation in which one unknown source is observed continuously at one center frequency (interrupted only for calibration). More maps must be computed per unit observing time, although in some cases each can have a smaller space-bandwidth product. The trouble is that the data must be sorted by source, center frequency, bandwidth, and polarization prior to synthesis. This implies that, unless the number of values of such variables is severely limited, a large fraction of the data from an observing session (preferably all) must be kept in direct access storage prior to mapping. It seems that at least 12 hours worth of such storage should be provided (this is 1.2×10^{10} b, assuming 16-b words and allowing 25% for indexing and overhead; it would fill 20 Memorex 3330 disks or 15 DEC RP-04's).

Another question, separate from considerations of special observing modes, is whether the output maps need to be kept temporarily in direct access storage. This depends on the users' display requirements (discussed in section E), which at present seem hard to define; so a design philosophy of maximum flexibility and interactivity at the display stage in the processing, more than at any other stage, seems justified. This implies the need for a large amount of direct access map storage. Enough storage for maps of all channels at the largest map size (1024 x 1024) is probably a minimum (2.1×10^9 b at 8 b/pixel and 25% overhead; 2-1/2 RP-04's). The need to compare several observations, to compute and display polarization maps, etc., and probably to support more than one user at a time implies a need for direct-access map storage several times larger than this.

The above requirements for direct-access storage must be considered a minimum, based on the use of pipelined processing wherever possible. Requirements for additional flexibility and user interaction with the data, particularly at the early stages of processing, can be met only by vastly increasing the storage capacity of the system.

Finally, consider permanent storage. Most maps, and probably also visibilities, will need to be stored as archives and also supplied to the user. Clearly such data grows linearly (or faster) with time; and since many years of operation are anticipated, ultra-dense storage is called for.

2. Flexibility

It is easy to say that the entire system should be made as flexible as possible. In view of the variety of desirable observing modes which we can currently enumerate, plus those which will no doubt be invented later, a rigidly-designed machine which performs a single, well-defined task seems unacceptable. But we must also take into account the cost of flexibility: not only the increased capital cost, but the decreased performance on routine tasks compared with that of a specially-designed system.

It seems impossible to study exhaustively all the tradeoffs of this sort. Therefore, the approach of this chapter will be to present specific design concepts and to analyze them critically, taking into account the extent to which they are flexible enough to accomodate a variety of options.

3. Interactivity

The extent and nature of the user's interaction with the telescope (which includes all data processing) is part of the problem of flexibility, and may be the subject of some controversy. It is likely that the degree of data interaction ("massaging") which has been common with smaller telescopes will not be feasible with the VLA, particularly for the spectral line system. Assuming that tradeoffs will be required (especially in the allocation of storage as discussed above), we must decide at what points and to what extent user interaction will be possible. Current thinking leans toward emphasizing interaction with maps by means of a flexible display and analysis facility, and de-emphasizing interaction in the calibration and map-computing processes.

4. System Block Diagram

Whereas a major investment has already been made in the hardware and software for the continuum VLA, we have decided to retain for the line system the same division into subsystems, namely the correlator, the synchronous computer, and the asynchronous computer subsystems (see Figure III-1). The asynchronous computer performs all tasks for which pipelining is not necessary or preferable. The boundaries between the subsystems occur at natural interface points in the processing, so that each subsystem is to a large extent autonomous. This simplifies the management of the development effort, allowing separate groups to work in parallel on the three subsystems.

The data rate across the correlator-to-synchronous computer interface is about 225 kb/sec maximum (based on 256 spectral channels, 10 sec of correlator integration, and 16 b words). This is the same data rate which would exist at this interface in the full continuum system as presently designed (the increase in correlator integrating time from 0.3125 sec to 10 sec just cancels the increased number of channels from 8 to 256). Across the synchronous-to-asynchronous computer interface the maximum data rate is also about 225 kb/sec, which is a substantial increase over the data rate at this interface in the continuum system.

The block diagram of Figure III-1 reflects some notable decisions about the allocation of tasks among subsystems. First, the correlator is thought of as delivering visibility measurements for each frequency channel of each baseline, so it incorporates an FFT processor for converting from the lag domain to the frequency domain. Second, and more importantly, the synchronous computer subsystem is expected to apply calibration information to the correlator-supplied measurements as they come in, and to maintain the stored calibration information (i.e., the complex gains of all spectral channels for all baselines), updating it when calibration sources are observed. (It is expected that an independent facility for post-observation calibration will have to be implemented in the asynchronous computer subsystem, but it is hoped that this will be needed only rarely; further discussion on this point is given in Section D of this Chapter.) Next, the synchronous subsystem is expected to make some data-quality decisions based on information from the monitor system, and to mark the measurements accordingly. It is hoped that this will be adequate to identify most, if not all, unusuable measurements. In the asynchronous computer subsystem, the

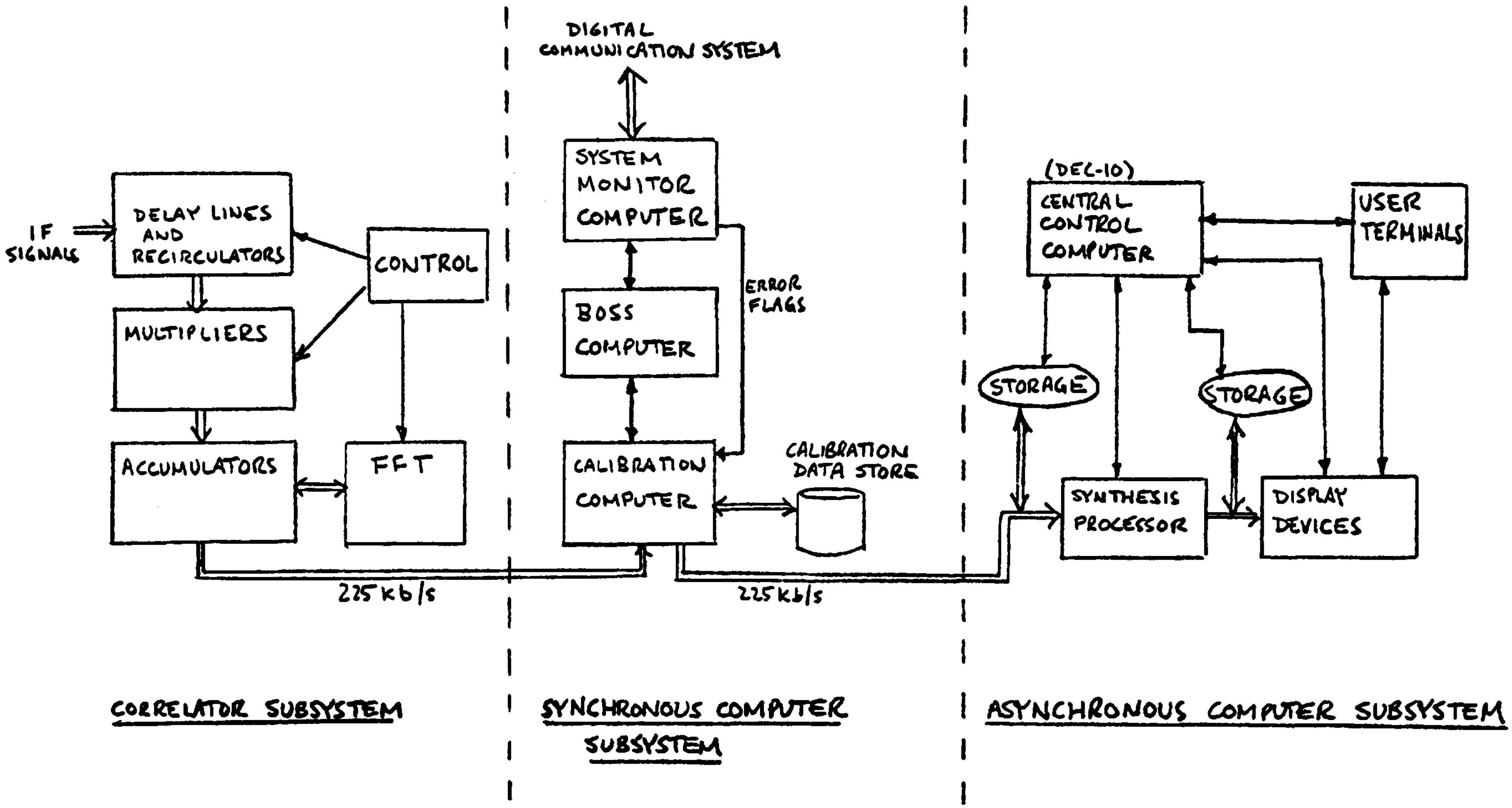


Figure III-1. System Block Diagram

block diagram identifies only the major tasks of synthesis, storage, and display. The DEC-10 computer, already acquired for the continuum system, is expected to play a major role in managing the resources of the asynchronous subsystem, but in its present form it is inadequate for handling the major line system tasks just mentioned. We envision the construction of a special-purpose processor for synthesis -- i.e., the construction of maps from the visibility data -- and some design alternatives are considered in section E.1 of this chapter. We also envision the acquisition of vastly increased mass storage capacity of various kinds. For displays, possible devices are discussed in section E.5; some of the digital devices now being acquired for the continuum system may also be useful for the line system.

C. THE CORRELATOR SUBSYSTEM

1. Specification

From the survey of the literature given in Chapter II, we conclude that the requirements for nearly all experiments currently forseeable can be met if about 250 spectral channels are available at 2.5 MHz total bandwidth, with at least the same number of channels at smaller total bandwidths. The ability to use much larger total bandwidths, possibly with smaller numbers of channels, is highly desirable. This situation is ideally suited to the use of high-speed logic in connection with recirculating buffers, as described in the next subsection.

The basic specifications are thus those of Table III-1. The integrating time of 10 sec is the maximum that can be used with the A array without noticable loss of information, and we specify that the correlator should provide all of this integration in order to minimize the data rate at the correlator-to-synchronous computer interface.

Table III-1: Correlator Specifications

Number of baselines	351
Number of spectral channels	256 at 2.5 MHz bandwidth ≥256 at smaller bandwidths
Maximum bandwidth	≥20 MHz (design goal: 100 MHz)
Integrating time	10 sec to 40 sec, programmable
Polarization tensor elements	1 at a time, rapidly switchable
Desirable flexibility	(a) more polarization elements by sacrificing spectral channels, (b) more spectral channels by sacrificing baselines.

Also indicated in Table III-1 are some flexibility features which would be desirable if they can be implemented at small incremental cost; while keeping the total number of correlation channels constant, we desire to be able to trade spectral channels for more polarization tensor elements, and to trade baselines (antennas) for more spectral channels.

2. The Recirculation Concept

Most of the electronics required for a multi-baseline correlator is taken up in multipliers and accumulators. In a multi-delay-channel cross-correlator such as we need for the spectral line system, the number of products which must be formed within one sampling period (equal to 1/2W, where W is the total bandwidth) is twice the product of the number of baselines and the number of spectral channels (as large as 179,712 in our case). Fortunately, we can avoid actually implementing this number of multipliers by taking advantage of the fact that digital integrated circuits now available can operate at speeds much greater than the sampling rate $f_s = 2W$ for most bandwidths W of interest. The idea is to build a small number of high-speed multipliers operating at clock frequency f_c, and to time share them among the various multiplications which need to be computed. One multiplier can compute $N \equiv f_c/f_s$ products per sampling period, so the number of multipliers required is reduced by a factor of N. For example, f_c can be 100 MHz; then for W = 2.5 MHz, we have N = 20. In addition, it is possible to time share a portion of the accumulator associated with each multiplication.

To see in more detail one way to accomplish the time sharing, consider Figure III-2, which gives a conceptual block diagram of a recirculating cross-correlator. We require a serial-in, serial-out memory

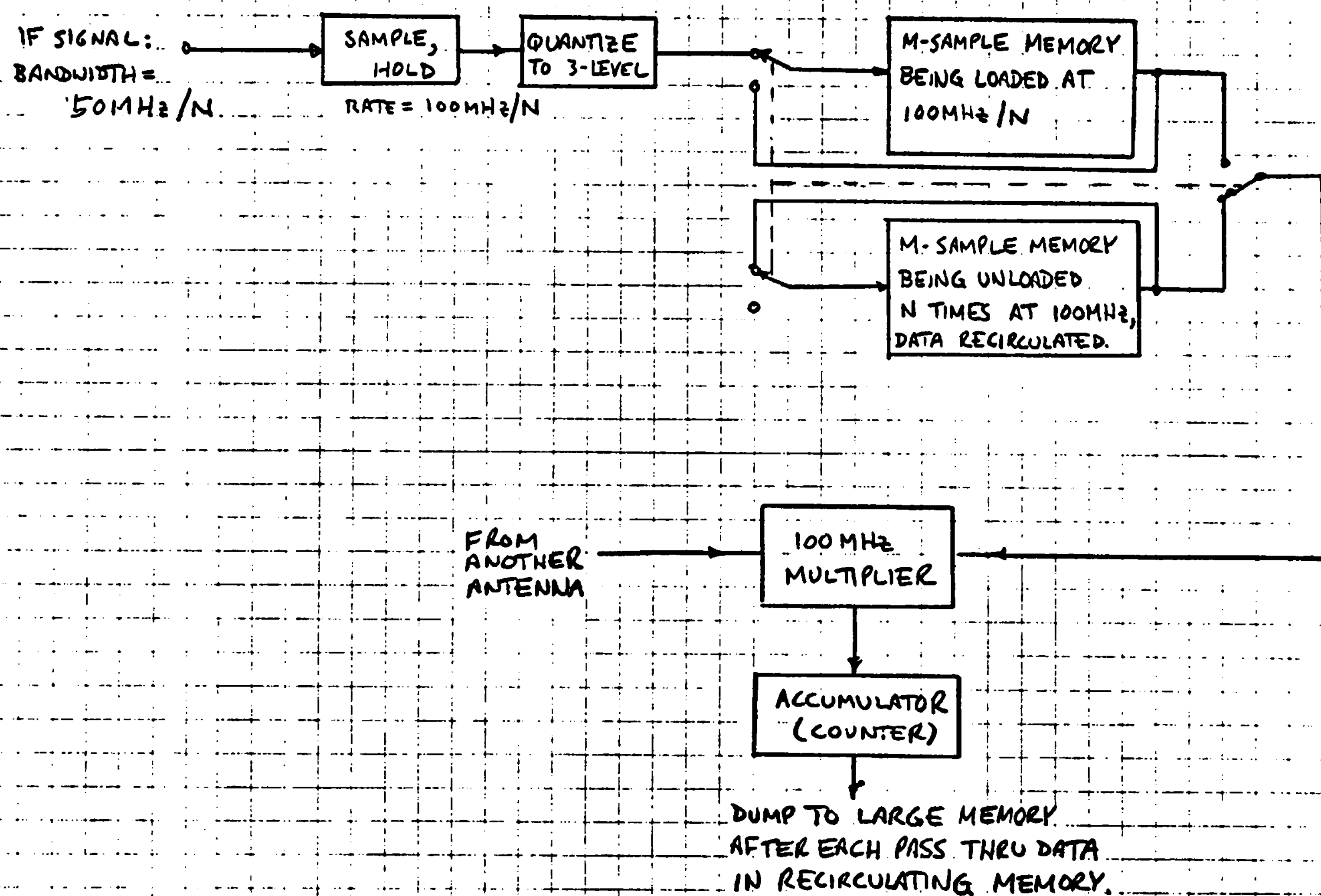


FIGURE III-2: THE RECIRCULATION CONCEPT

(conceptually, a shift register) which holds enough samples for one "basic integrating period" (explained below) and which can be loaded slowly (rate f_s) and unloaded rapidly (rate f_c). Two such memories are shown in Figure III-2; one is being loaded with current data while the other, loaded previously, is being unlaoded N times into a high-speed multiplier with the data being maintained in the memory by recirculation (N is thus known as the recirculation factor). The other input to the multiplier is fed with corresponding data from another antenna. The sample products are accumulated, so that at the end of one complete pass through the memory, one correlation has been computed for one basic integrating period. The accumulated count is then dumped into a large (but slow) memory. Next we clear the accumulator and begin reading the same memory into the multiplier a second time, but this time we begin with the second sample in the memory rather than the first, so that at the end of this pass through the memory the accumulator contains the correlation for a slightly different lag (delay) than the first time. Continuing in this way, we compute correlations for N lags after N memory passes. By that time, the other memory has just finished filling with new data and the roles of the two memories can be interchanged.

Note that on the Nth pass, N-1 samples are unused; thus it is important to have M>>N for efficient use of the incoming data. However, by a more clever memory arrangement than shown in Figure III-2, it is possible to reduce the number of unused samples to much less than N and also to reduce the required memory size from 2M to slightly more than M samples. A large value of M is desirable to reduce the rate of dumps to the large memory, and thus to minimize the speed at which it must operate;

but a small value of M is desirable to minimize the cost of the recirculating memory. This tradeoff may result in the choice of a value of M which gives a rather short basic integrating time; additional integration can then be done in the large memory.

The basic concepts for such a recirculating correlator originated with A. E. E. Rogers, G. Papadopoulos, and J. Carter at Haystack Observatory, and were first used by J. A. Ball to construct an autocorrelator (Ball 1973).

It is important to note that the multipliers of a recirculating correlator may be organized to produce a larger number of channels if the input spectrum is divided into smaller bands before correlation. If an input spectrum of bandwidth W is divided into K bands, each of width W/K, the effective number of channels is increased by a factor of K. The maximum bandwidth which can be analyzed with a given maximum sampling rate is also increased by K. On the other hand, some overlap of band edges is needed to assure accuracy in aligning the baselines of each, and the cost of analog processing equipment, samplers, delay lines, and recirculators is increased by K.

3. Proposed System

A system meeting the requirements stated in Table III-1 has been devised utilizing the following principles:

1) Use as much as possible hardware in common with the continuum system.

2) Use the recirculation concept to reduce required hardware.

3) Use the K-partition principle described in the preceding section to enhance capability.

4) Keep as much flexibility as possible with regard to polarization and multi-channel, wide-field continuum use.

5) Build the system in a modular form to allow a sub-set of the system to be tested before committing construction of the entire system. Modularity also simplifies design and maintenance.

A block diagram showing the relevant aspects of the continuum system presently under construction and the modifications needed for spectral-line operation is given in Figure III-3. The system contains, for each antenna, two 50 MHz bandwidth I.F. bands on each of the two front-ends which are orthogonally polarized and may or may not be at the same wavelength (i.e. 1.2, 2, 6, or 18-21 cm). The four I.F. bands have separately programmable center frequencies, lobe rotation rates, delays and bandwidths. For continuum operation each I.F. signal is divided into 0° and 90° phase-shifted components before sampling with 3-level quantization; eight 2-bit delay lines are then required. The digitized and delayed I.F. signal is then multiplied in-phase (i.e. 0°x0° or 90°x90°, but not both) and in quadrature (i.e. 0°x90° or 90°x0°, but not both) with both the co- and cross-polarized signals from another antenna. The continuum system thus requires four multiplications per I.F. signal per antenna pair, 16 multiplications per antenna pair, and 16x351 = 5616 total multipliers.

The proposed modifications of this system and their costs are summarized below:

1) The four I.F. Receivers per antenna will be modified or augmented to allow variable bandwidths for spectral line work ($280K)

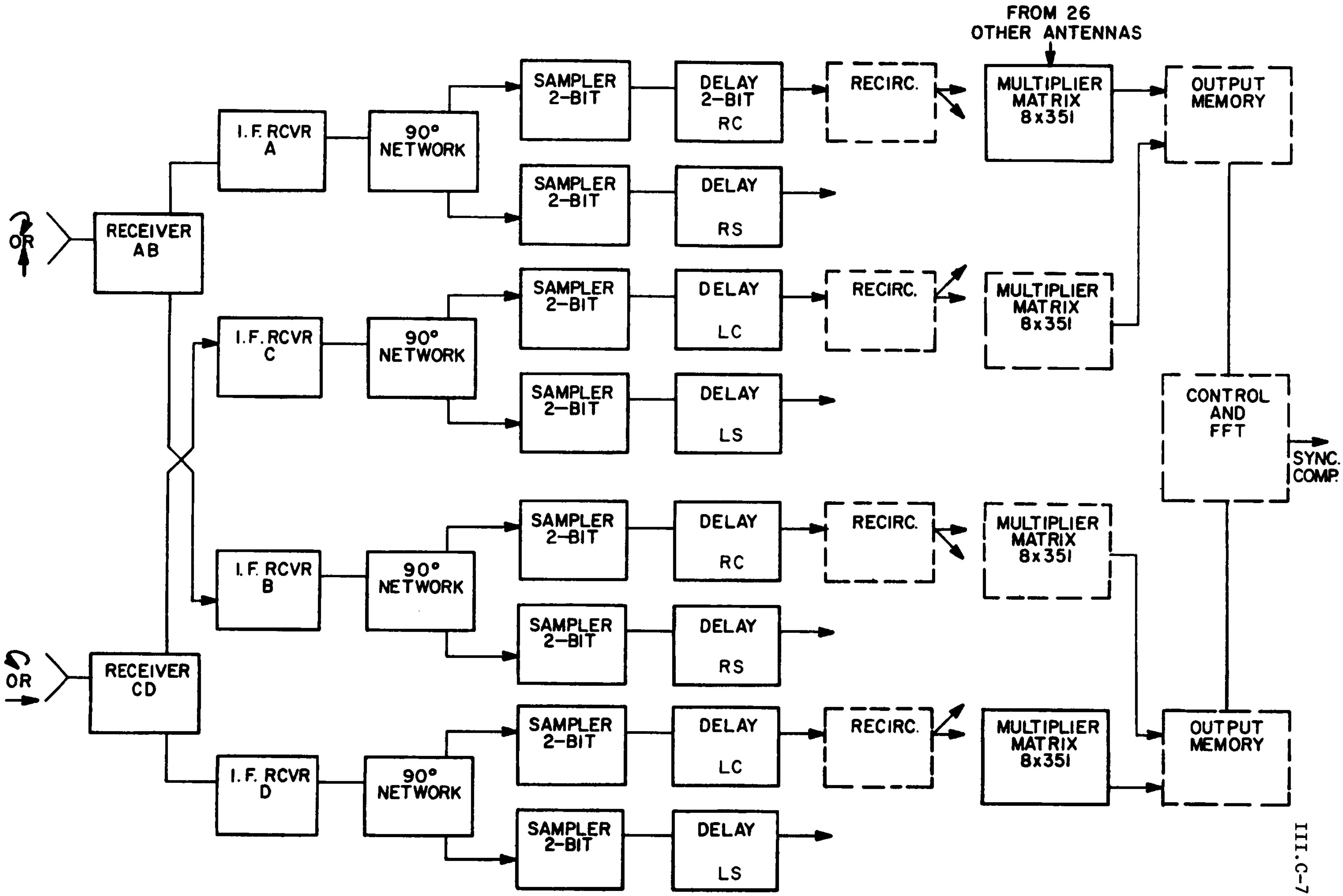


Figure III-3

Block Diagram of VLA Electronic System. Portions shown in dotted lines are for spectroscopy. For simplicity connections to multiplier matrices are not shown.

2) Four 73,728-sample dual recirculating memories will be added per antenna ($176K)

3) The number of 2808-multiplier matrices will be increased from two to four ($232K)

4) Output memory will be increased to allow recirculation of up to 32 times and integration of up to 40 seconds ($200K)

5) A hardware Fast-Fourier Transform calculator and a control computer will be added ($90K)

6) The number of read-out multiplier bits will be increased and multiplexers will be added to allow changes of mode of operation ($50K)

The proposed correlator consists of four processing quadrants grouped into two identical systems, with one controller and FFT device. The quadrants can be connected together in five different modes which are described below. The bandwidth and resolution of the system in the various modes is given in Table III-2.

a. Continuum Mode

This is the normal continuum mode in which two 50 MHz dual-polarized bands are correlated to give all four polarization products. Only two out of four quadrants are required for this task; the redundant quadrants will be connected and used in case of failure. (The continuum system utilizes in-phase and quadrature multipliers to determine the complex correlation coefficient whereas a plus and minus lag method is proposed to measure the complex cross-power spectra. An initial investigation of using either method for both line and continuum has indicated that it is more efficient to use the different methods.)

b. Single-Band Spectral-Line Mode

A single I.F. band will be processed by all four quadrants in series; i.e. each quadrant will correlate at different lag values. The number of frequency channels will be four times and resolution will be one-fourth the values for a single quadrant.

c. Dual-Band Spectral-Line Mode

Two I.F. bands which may be orthogonally or identically polarized and may or may not be at the same frequency are individually

TABLE III-2(a) PARAMETERS OF ONE CORRELATOR QUADRANT

BANDWIDTH MHZ	MAXIMUM FREQUENCY CHANNELS	RESOLUTION(1) KHZ	RECIRCULATOR(2) FACTOR	OVERSAMPLE(3) FACTOR
50.	4	18,000(4)	1	1
25.	8	3,780	2	1
12.5	16	944	4	1
6.2	32	236	8	1
3.12	64	59	16	1
1.56	128	14.7	32	1
.78	128	7.4	32	1
.39	128	3.7	32	2
.195	128	1.8	32	2
.097	128	0.9	32	2

(1) Half-power width of equivalent filter.

(2) Increases, due to recirculation, in the number of correlations computed.

(3) Ratio of sampling frequency to Nyquist rate. Quantization increases noise by 1.23 if oversample factor = 1, by 1.12 if oversample factor = 2, compared with unquantized (analog) processing.

(4) 15,120 kHz in single-band and double-band modes.

TABLE III-2(b) MULTIPLIER FACTORS FOR DIFFERENT MODES OF OPERATION

BANDWIDTH	CHANNELS	RESOLUTION	MODE
X1	X4	÷4	Single-Band
X2	X2 (per band)	÷2	Double-Band
X4	X1 (per band)	÷1	Four-Band
X1	X1	÷1	Polarization

correlated (no cross-products), using two correlator quadrants for each band. The number of frequency channels per band will be twice and the resolution will be one-half of the values for a single quadrant. The two bands will be separately outputted to the synchronous computer system.

d. <u>Four-Band Spectral-Line Mode</u>

Four I.F. bands - two on each orthogonal polarization - are individually correlated. The number of frequency channels <u>per band</u> is equal to the value for one quadrant. Note that for a given total bandwidth, B, only B/4 need be analyzed by each module (neglecting some required overlap); and recirculation will provide four times as many total channels as compared with the Single-Band mode applied to bandwidth, B. This is therefore the highest resolution mode but has the additional complexity (and perhaps, errors) of four overlapping bands.

e. <u>Polarization Spectral-Line Mode</u>

Two orthogonally polarized I.F. bands at identical center-frequencies are correlated including cross-products. Each quadrant will compute one of the four polarization products and the channels and resolution will be equal to the one-quadrant values given in Table III-2(a).

<u>Examples</u>. Suppose a total bandwidth of $\geq$1.0 MHz is needed, and the observer desires the maximum possible spectral resolution. By choosing 1.56 MHz total bandwidth and operating in single-band mode, he finds from Table III-2 that the correlator is capable of computing 128 x 4 = 512 channels at 14.7 kHz/4 = 3.7 kHz resolution. However, it is expected that later data processing will be limited to a maximum of 256 channels. The observer can either ignore the outer 256 channels, retaining the 3.7 kHz resolution and reducing the total bandwidth to 0.78 MHz; or he can compute only 256 channels, degrading the resolution to 7.4 kHz but retaining the 1.56 MHz bandwidth. As a third alternative, he could operate in double-band mode, putting both bands on the same frequency but with opposite polarizations; at 1.56 MHz bandwidth, he gets 256 channels per band. By adding together the oppositely polarized bands, he can keep the total number of channels to 256 while improving the signal-to-noise ratio.

As another example, consider an observer who uses the line system to reduce bandwidth-smearing in continuum mapping of large fields (cf. Appendix A). By using the single-band mode at the maximum bandwidth of 50 MHz, he obtains 16 channels with 3.78 MHz channel bandwidth. In double-band mode at 25 MHz per band, he obtains the same total bandwidth of 50 MHz, but now has 32 channels (16 per band) and 1.89 MHz channel bandwidth; alternatively, at 50 MHz per band, he has 100 MHz total and 7.56 per channel. He could also operate in four-band mode, which at 25 MHz per band gives 100 MHz total and 3.78 MHz per channel; but because of the front end arrangement (Fig. III-3), the four bands must include two of each polarization.

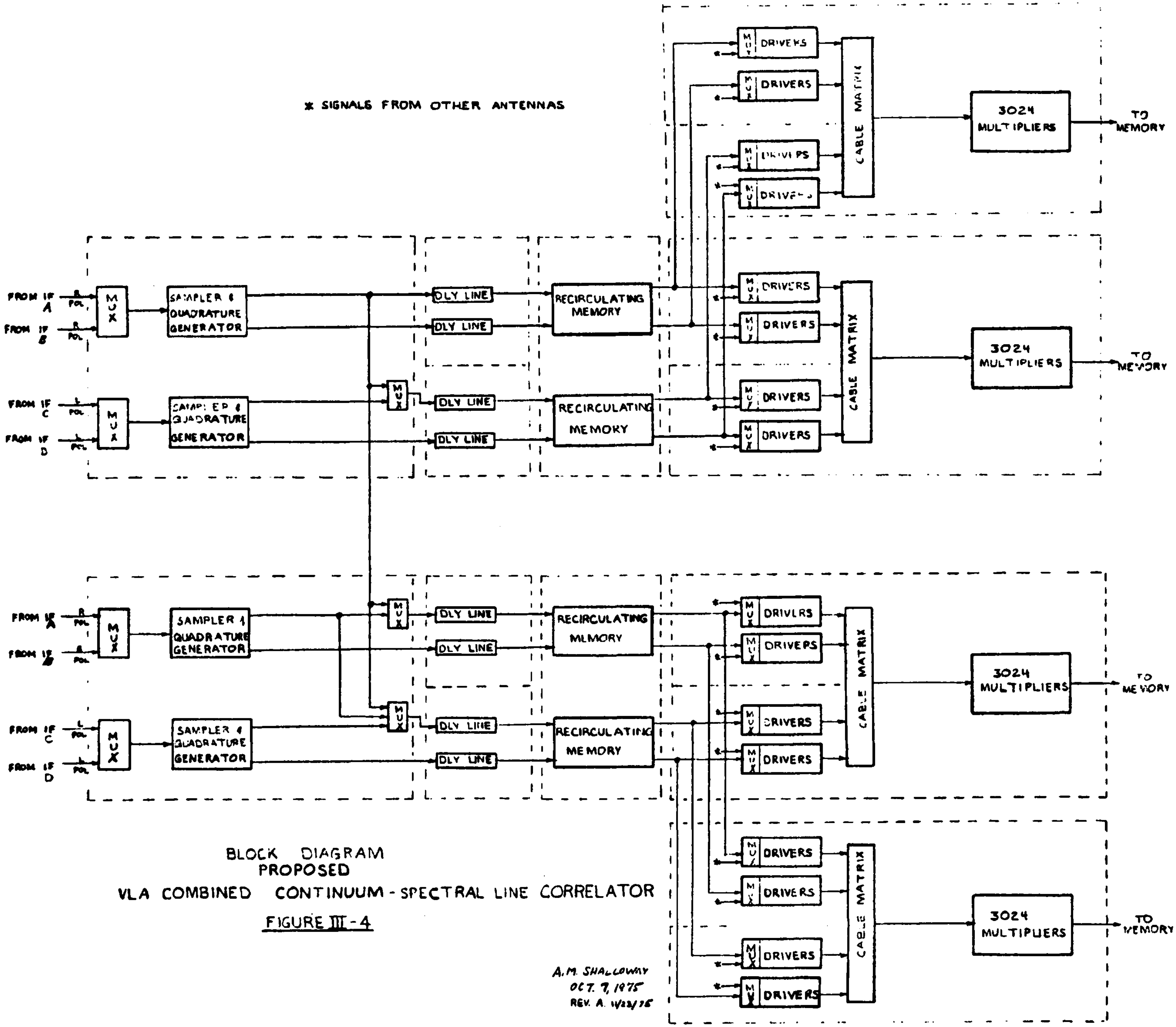
Several engineering studies are presently in progress and will be completed in the next several months. Questions which have been identified and the present state of solution are summarized below:

1) How much modification is needed to the present delay and multiplier boards?

No modification is required on the delay boards or delay mother boards. We must add two multiplexers (24 pin chips) to the multiplier boards. This may be possible without an artwork change - but obviously an artwork change would be much better. No change is required in the multiplier mother board, as these signals (extra counter outputs) can be taken from the connectors with wirewrap.

2) What is the system layout? Where should recirculators and line-continuum switching be located?

See Figure III-4 (Block Diagram) and Figure III-5 (Rack Layout).



BLOCK DIAGRAM
PROPOSED
VLA COMBINED CONTINUUM - SPECTRAL LINE CORRELATOR

FIGURE III - 4

A.M. SHALLOWAY
OCT. 7, 1975
REV. A. 11/23/75

I= INTEGRATORS
M= MULTIPLIERS
DR=DRIVERS
D= DELAYS
DT=DELAY TEST
S= SAMPLERS

NUMBER PRECEEDING ABBREVIATION
INDICATES NUMBER OF
MOTHER BOARDS IN RACK

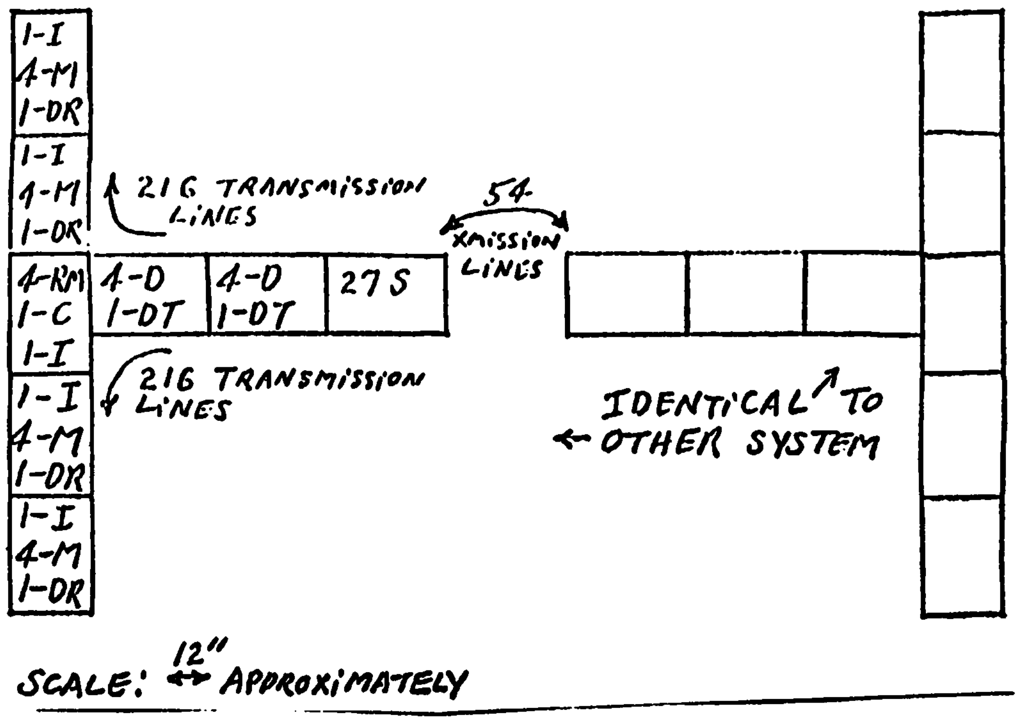


FIGURE III-5. VLA CONTINUUM AND SPECTRAL CORRELATOR, RACK ARRANGEMENT

2) (cont.)

Three sets of multiplexers - one each preceding the samplers, delay lines and drivers - provide complete electronic switching.

3) What is the optimum trade-off between recirculator memory capacity and multiplier accumulator readout rate?

Many factors influence the above: cost of memory used in recirculator memory, speed of memory which takes into account the number of parallel inputs and outputs per IC chip, time available for emptying of accumulators - this is affected by size of memory, whether it is random or sequential, and how the cycles of the recirculator memory fit into the 50.48 ms waveguide transmission time. The solution chosen - using Fairchild CCD450A shift registers - is by no means the only solution. However, it has been completely designed and checked, and is a feasible, reasonably priced system. In the final design some other route may be followed. With the one chosen we are able to use the high speed and low speed integrator cards and temporary storage cards with very little change (these are wirewrap cards). With this design, everything is optimized to give the lowest cost and maximum efficiency: 89.26% of the entire 52.0833 ms period of the VLA or 92.09% of the 50.48 ms waveguide transmission time available. In this design, we throw away five of the least significant bits of the accumulators and save 25 bits. 17.64 μsec are required to empty the high speed accumulators after each recirculator pass.

4) What logic family should be used for the recirculator memory?

As discussed in 3 above, for the present design a Fairchild charge controlled device - CCD450A - was chosen. This was based on projected prices as compared with shift register and RAM prices. A 4K or possibly 8K RAM (or maybe 16K in the future) may be a possibility if the projected prices are incorrect.

5) Should the sampler input spectra be bandpass or low pass? A pre-sampler, switchable-filter, system design is needed, but has not been initiated.

6) How can we most efficiently reduce the available data points at narrow bandwidth to $\leq$256x351 per 10 seconds for further processing?

This has not been investigated at present.

7) What is the system timing? How is recirculation timed with I.F. transmission transmit-receive and with integration periods? Can time bihs for pulsar observations be accommodated?

VLA cycle $= \frac{1 \text{ sec}}{19.2} =$ 52.0833··· ms.

VLA observation time = 50.48 ms per cycle

The 52.0833 ms cycle period has been divided into 70 recirculator memory cycles. 64 of these cycles occur during the 50.48 ms integration time.

Memory cycle time = 737.28 μsec

Dead time between memory cycles = 6.768 μsec

Accumulator dump time = 17.640 μsec (this consists of 6.740 μsec of the dead time plus 10.900 μsec, of the memory cycle time.

Efficiency as described in 3 above.

This provides 64 memory integration cycles of 726.38 μsec integration +17.668 μsec dead time during the 50.48 ms observation time of the receivers. Any group of the 64 results can be separated from the remainder to provide pulsar observations giving a resolution on the pulsar of 726.38 + 17.668 = 744.048 μsec. Obviously, if the VLA timing system based on 19.2 μsec is not changed, occasionally the pulse can occur during "data invalid" time and this one cycle would be thrown out.

A cost estimate for the sytem is given in Table III-3.

TABLE III-3. COMBINED SPECTRAL-LINE/CONTINUUM COST

ITEM	UNIT COST $K	ONE QUADRANT QUANT	ONE QUADRANT TOTAL $K	TWO QUADRANT QUANT	TWO QUADRANT TOTAL $K	FOUR QUADRANT QUANT	FOUR QUADRANT TOTAL $K
I.F. Filters, 10-bandwidth	2.5	28	70	56	140	108	280
Dual Samples, 90° Phase Shifter	1.0	56	56	108	108	108	108
2-Bit Delay Lines	.81	112	91	224	182	224	182
2-Bit Recirculating Memory	1.58	28	44	56	88	112	176
2808-Multiplier Quadrants	115.8	1	116	2	232	4	463
2808-Output Memory for x32 Recirc.	50.0	1	50	2	100	4	200
Controller/FFT Device	70	1	70	1	70	2	140
Racks, Power, Cables, and Clock	27	1	27	2	54	4	108
Development, Testers	90	1	90	1	90	1	90
Non-Recoverable Costs of Interim Continuum System	50	1	50	1	50	1	50
TOTAL			664		1114		1797

4. Rejected Option: Spectral Filtering Before Cross-Multiplication

Recent advances in both digital and analog implementations of the discrete Fourier transform (DFT) make it attractive to consider separating the IF signals into narrow-band channels before cross-correlation. The attractiveness arises because when the DFT is implemented by the chirp-z transform (CZT) algorithm (see Rabiner et al. 1969), as can be done very efficiently in some new technologies (Whitehouse et al. 1975), a serial-in, serial-out structure arises which allows us to construct the spectral line correlator with only two real multipliers per baseline.

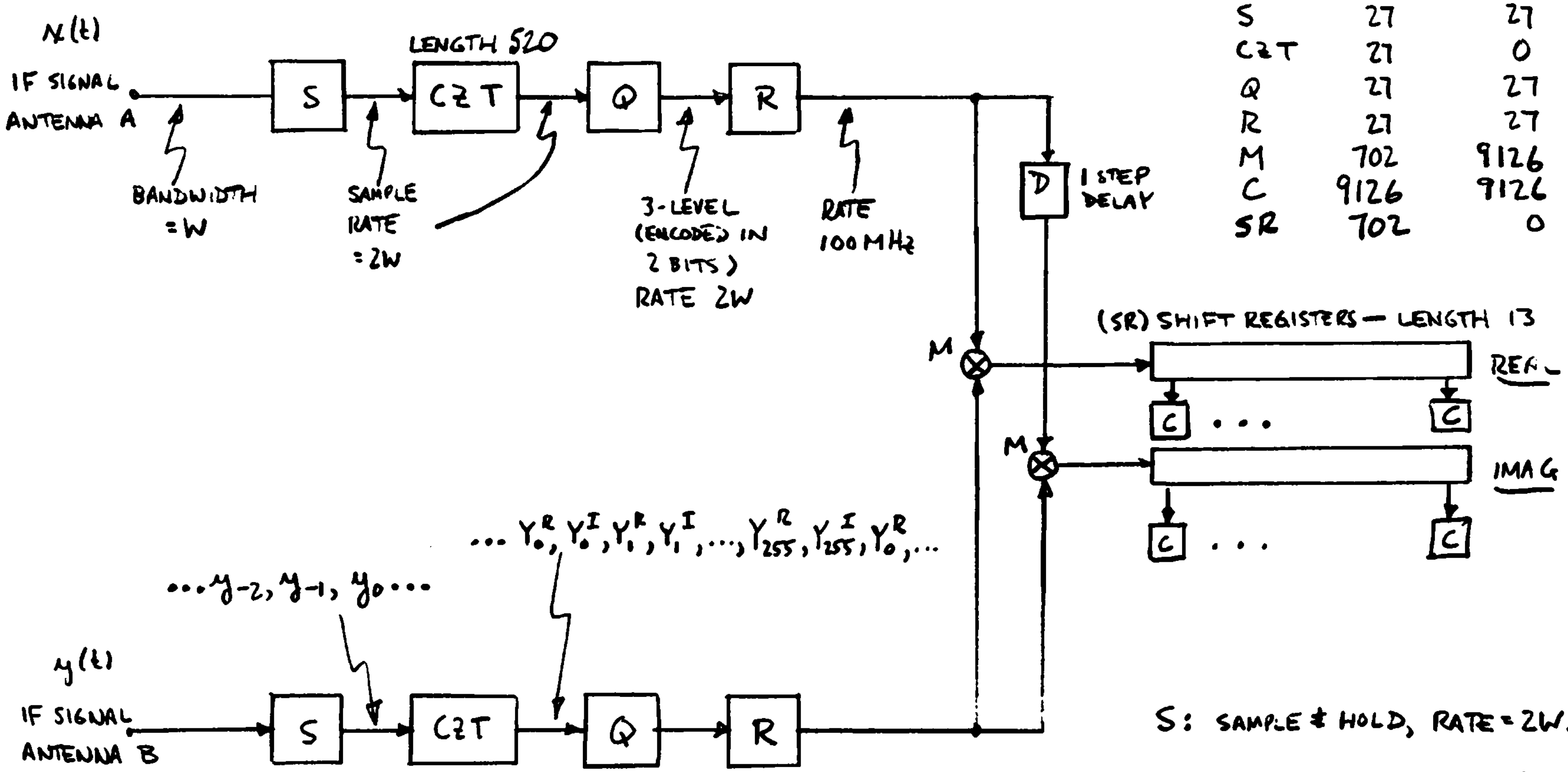
Specifically, consider the block diagram of Figure III-6. The incoming IF signals at bandwidth W are sampled at rate 2W and applied to CZT filters of length 520. The output of the CZT, like its input, is a discrete-time signal at sample rate 2W; at any particular sample time, the output is either the real or the imaginary part of one element of the DFT of the last 520 input samples; i.e.,

$$\begin{matrix} Y_t \\ Y_{t+1} \end{matrix} = \sum_{k=0}^{519} y_{t-k} \begin{matrix} \cos \\ \sin \end{matrix} (-2\pi kt/260)$$

where $\{Y_t\}$ is the output sequence and $\{y_t\}$ is the input sequence for the CZT. The output samples are then quantized so that further processing can be done digitally (we envision an analog CZT implementation, as discussed below); three-level quantization is used, just as in the other correlator designs. Since the various narrow-band frequency channels come out sequentially, the cross-multiplication for each baseline requires only two multipliers (real and imaginary) for all frequencies; furthermore, each multiplier need only operate at rate 2W. We must, however, separate by frequency the sequentially-generated products, and accumulate a count for

FOR 27 ANTENNAS:

	THIS SCHEME	CROSS-CORRELATION
S	27	27
CZT	27	0
Q	27	27
R	27	27
M	702	9126
C	9126	9126
SR	702	0



S: SAMPLE & HOLD, RATE = 2W.

CZT: CHIRP-Z TRANSFORM (DFT)

Q: QUANTIZE TO 3-LEVEL

R: RECIRCULATOR, RECIRC. FACTOR = 100 MHz/2W.

C: COUNTERS (ACCUMULATORS)

M: 3-LEVEL MULTIPLIERS, 100 MHz.

FIG. III-6: SCHEME FOR LINE INTERFERMETER HAVING 260 FREQUENCY CHANNELS AT 2.5 MHz BANDWIDTH, BASED ON CZT.

each frequency separately. By operating the multipliers at a rate much faster than 2W and using the recirculation concept discussed earlier, the number of counters required can be reduced by the recirculation factor. Thus, in Figure III-6 we show 13 real-part and 13 imaginary-part counters per baseline; at W = 2.5 MHz and with 100 MHz multipliers we have a recirculation factor of 20, so that there are enough counters for 13 x 20 = 260 frequency channels, corresponding to the CZT length. At larger bandwidths, up to the CZT's speed limit, the number of available counters limits the number of channels.

The component count for this scheme is compared with the cross-correlation schemes of the preceding subsections in the upper right corner of Figure III-6. Note that while the number of counters is the same, those for the present scheme can be a factor of 13 slower.

In spite of the great simplifications which such a scheme appears to offer, we have decided not to pursue it for the VLA because of several disadvantages, enumerated below. The first of these is rather fundamental, and affects any arrangement in which the spectral filtering preceeds cross multiplication; the others are technical difficulties with available CZT devices, and may be alleviated as the technologies develop.

a. Dynamic range problem. Each input sample and each output sample of the CZT is a zero-mean, Gaussian random variable. The input samples all have the same variance, which can be held to a known value by an automatic level control on the IF signal. The variance of an output sample, however, depends on the average power in its narrow-band frequency channel, so the samples can have different variances. If the channel-to-channel variation is substantial, the three-level quantization with fixed switching voltages may not be accurate enough to represent all channels.

Such variations have two causes: non-flat receiver temperature-gain product across the passband, and the non-flat spectrum of the source being studied. The former can probably be kept to ± 1 db (26 percent), a variation which can be ignored after astronomical gain calibration of each narrow-band interferometer. The effect of the source spectrum can also be ignored if the source is weak, contributing no more than a few percent of the system temperature in any narrow-band channel, or if the source spectrum is essentially flat, being mostly continuum radiation. In these cases, three-level quantization would still be adequate. A difficulty arises for very strong sources with deep absorption features. Here there may be a ratio of as much as 10 in the power in different channels, requiring quantization to 4 bits or more, with a considerable increase in multiplier complexity. Although there are only a few sources known which would thus require finer than three-level quantization, we did not want to impose this limitation on the astronomical capabilities of the VLA, nor did the increased complexity of multi-bit multipliers seem justified.

b. CZT device speed and flexibility. Two technologies for analog computation of the CZT have been successfully demonstrated by various laboratories (see Whitehouse et al. 1975), namely surface acoustic wave (SAW) devices and charge-coupled devices (CCD). Both utilize a transverse-filter architecture which propogates a signal down a tapped delay line having pre-programmed tap weights. Some difficulties are that (1) while SAW devices are capable of very large bandwidths (hundreds of MHz), the technology does not readily allow variable bandwidths, because of the fixed sound speed in the medium, although some bandwidth selection capability could be provided by multiplexing several SAW devices together; and (2) the CCD devices, which allow variable bandwidth by adjusting the clock rate, are at present only fast enough for total bandwidths up to about 5 MHz.

c. CZT device stability and accuracy. While it appears that the devices just mentioned can be fabricated for computational accuracies better than 1 percent, and perhaps as good as 0.1 percent, it is uncertain whether this is good enough for some critical experiments. It is also not known whether the temperature and time stability of these devices is adequate for our purposes.

D. THE SYNCHRONOUS COMPUTER SUBSYSTEM

1. General Considerations

The synchronous computer subsystem is entrusted with the real-time control of the observations. It must accept and interpret observation requests from the user; send appropriate commands to the front ends and other telescope electronics and to the correlator for proper setup of the observation; maintain accurate clocks; compute necessary ephemerides; control the antenna pointing, the lobe rotators, and the delay lines; keep track of all monitor information and sound alarms or set error flags as needed; and do some preliminary processing of the visibility data before passing it to the asynchronous subsystem. Of course, all of these functions are being implemented for the continuum system. The principal difference with the line system is that the number of measurements being made by the correlator is much greater. In addition, the control of the correlator is somewhat more complex. In this subsection, therefore, we concentrate on the handling of the spectral line correlator data by the synchronous computer, and assume that most other functions of the subsystem will be very similar, if not identical, to the continuum case.

Logically, all processing which must be done in real time (as the observation progresses) should be considered part of the synchronous subsystem. But here we shall take the names "synchronous" and "asynchronous" loosely and allow some equipment allocated to the asynchronous subsystem to operate synchronously, if this proves convenient. Further, we shall allow some tasks which in principle could be executed asynchronously to be assigned to the synchronous subsystem if there is good reason to do so. The tasks which we shall consider here, calibration and data flagging, are both in the latter category.

2. The Calibration Task

To convert the accumulated complex cross-power spectra from the correlator into astronomically meaningful complex visibilities, it is necessary to divide each one by the corresponding narrow-band interferometer's complex gain. Each complex gain is controlled by many factors, including the two antenna gains, the system noise temperatures, the local oscillator phase offsets, and the frequency responses of the front ends and the IF transmission systems. Some of these, e.g. the system noise temperatures, are measured frequently by the monitor system; the estimation of others requires astronomical calibration.

In most existing synthesis telescopes (including the Green Bank interferometer), the analysis of calibration source observations to determine the complex gains and the dividing of these gains into the correlator measurements to obtain calibrated visibilities are done asynchronously, usually after the observing session is completed, so that all calibration data can be considered simultaneously. For the VLA line system, however, there are compelling reasons to perform the calibration in real time.

First, the time required to post-calibrate using a general-purpose computer would be excessive; for example, the DEC-10 CPU time required merely to multiply each elementary measurement in a 12-hour observation by a complex number is 2.1 hours (not including any overhead for indexing or input-output)*. Second, real time calibration allows elimination of temporary storage for uncalibrated data (up to 1.2×10^{10} bits in 12 hours), a substantial saving. (The complex gains could still be stored, at small incremental cost, so that the uncalibrated data could in principle be recovered if necessary.) Finally, it allows immediate re-ordering of the data to a form suitable for mapping (as discussed more thoroughly later, in connection with the asynchronous subsystem), possibly saving additional storage and making possible the production of one or a few real-time images to assist the astronomer in controlling the observation.

The accuracy of real-time calibration is limited by the fact that only past calibrator observations, and not future ones, can be used to calibrate current data. This can, of course, be overcome by more frequent calibrator observations; whether or not this requires too high a frequency of calibration depends on the stability of the front ends, the local oscillators, and the IF transmission path. We are expecting that the drift on a time scale of several hours can be kept negligible.

Real-time calibration leads to the concept of the current <u>state of calibration</u> of the telescope, which is defined by the set of complex gains currently stored, and which is determined by means of a <u>calibration algorithm</u> from the recent history of calibration source observations and from current values of appropriate monitor system measurements.

---

* It is possible to reduce substantially this computation time by using special array processors under the control of the DEC-10. But with I/O, a single calibration pass on 12 hours of data can still be expected to take more than an hour.

The calibration algorithm attempts to provide the best possible estimates of the current complex gains; the algorithm can be quite complicated, and is not seriously limited by the fact that its results will be applied in real time. The details of the algorithm are beyond the scope of this report, and will be dependent on our experience with the telescope.

It is expected that no additional equipment will be necessary in the synchronous subsystem to accomplish line-system calibration, beyond the equipment already on hand for continuum operation (recall that the data rate across the correlator-synchronous interface is the same in the line system as in the original full-scale continuum system). The calibration algorithm can be implemented in software and executed as needed to update the complex gains, using either the Boss CPU or the Cora/Corbin CPU's. The reciprocal gains can then be stored by antenna in the core memory of Cora and Corbin, requiring at most 14K words (for 256 spectral channels, all antennas). Cora and Corbin, which interface to the correlator, can then handle the calibration of incoming data in a pipeline fashion, passing the results one baseline at a time to the asynchronous subsystem interface.

3. <u>The System Monitoring and Data Flagging Task</u>

Automatic fault detection and flagging of data which is probably not usable is an important and necessary feature for both the continuum and line VLA. For the line system especially, the possible data base size will make post-observation review of the entire data base very difficult. We expect that some facility for selective review and manual editing will still be necessary (in the asynchronous subsystem), but the amount of bad data handled in this way must be kept very small. We regard this as a software problem for the synchronous subsystem, where again the details of the implementation will depend heavily on our experience with the telescope.

4. <u>Hardware and Software Requirements</u>

As mentioned above, no new hardware is needed to handle the line system tasks, which are for the most part not substantially different from the continuum case. However, we do propose to purchase two high-density (6250 bpi) tape drives on which to write a backup copy of the data sent to the asynchronous subsystem. In addition to data protection, this allows the telescope to be kept in operation even if some part of the asynchronous subsystem is down. About five tapes will be written per 12-hour period with the maximum line data rate. Our current cost estimate for the drives, controller, and interfacing is $100,000.

Development of the software necessary to handle the full line system, beyond that needed for continuum, is expected to require one man-year.

E. THE ASYNCHRONOUS COMPUTER SUBSYSTEM

1. Overview

In the system block diagram (Figure III-1), the asynchronous subsystem is shown as performing synthesis, storage, and display tasks under the control of an astronomer through a central computer. Such a system can, we believe, be made adequate for producing accurate three-dimensional maps, displayed in a variety of convenient forms. However, as discussed in Chapter II, we expect that most observers will require considerably more facilities in order adequately to analyze their results. The analysis task can be performed partially in the mind of the astronomer, assisted by the interactive display devices which we expect to provide; but numerous important tasks, such as the separation of continuum and line radiation and the estimation of optical depths, require extensive machine computation. We shall show in the last part of this section that the extent to which such computational tasks can be supported with a general purpose computer is extremely limited; in particular, any computations involving the whole of a maximum-size map (1024 x 1024 x 256 points) are probably impractical. It may be that such computations can eventually be supported by employing optical processors or multiple high-speed digital array processors; our investigation of this will continue.

We conclude that the deliverable product of the VLA spectral line system, at least initially, will be a three-dimensional map for each source observed, displayed both interactively and in permanent recordings in any of a large variety of forms at the astronomer's option (as described later in this section). We expect that this will allow many useful experiments to be performed, but we recognize that extensive additional computational support will eventually be needed if the VLA is to approach its full potential for spectral line astronomy. Although we are unable now to specify just what equipment will be needed (especially since the image processing field is changing rapidly), we recommend continued development of the computational facilities after an initial system* -- designed only to produce and display accurate maps -- becomes operational. Computations for map analysis in the initial system are expected to be limited to what can be implemented in DEC-10 software; significant analysis might be possible to the extent that the astronomer is willing to limit it to a small subset of his three-dimensional map. It is only this initial system which we propose to discuss in this chapter; it alone presents formidable data processing problems.

However, the ultimate need for much-expanded analysis facilities suggests a design option which we have not pursued, but which should be mentioned. It may be that a large computing facility not located at the VLA site will eventually support (1) all map analysis, and (2) as much of the routine processing (e.g., editing, re-calibration, synthesis)

---

* By an "initial system" we mean one which will handle data from the full VLA, but which has limited analysis capabilities. We shall also make occasional reference to "interim systems", by which we mean equipment and software for handling part of the telescope before all antennas become available.

as we choose to locate there. This might be cost-effective, for example, if the off-site facility supports not only the VLA but also other work of the NRAO and/or other organizations. In that case, the equipment at the site need only do as much processing as is necessary to ensure the quality of the data, and then prepare it for transmission to the off-site facility. If it were clear that the VLA is to be operated in this way, we expect that the design of the on-site equipment would be quite different from what we shall recommend here; in fact, hardly any additional on-site investment (beyond that already committed for continuum operation) is likely to be necessary.

The discussions which follow assume that the initial system will be implemented entirely at the site; this will require considerable on-site investment. There are numerous advantages, including that (1) it allows us to make full use of existing equipment; (2) it is consistent with the continuum asynchronous subsystem, where all currently-planned computational support will be at the site; and (3) it allows the astronomer more flexibility in deciding what processing to carry out in near-real-time, since all facilities are in one place.

In the subsections below, we discuss several alternatives for each of the major data processing requirements in the initial system. We do not make definite recommendations as to which alternatives should be selected. Our studies are not complete, and this report attempts only to summarize our best information and ideas as of November 1, 1975.

2. Synthesis

In this subsection we consider several alternative implementations of the block labeled "Synthesis Processor" in Figure III-1. This processor is required to accept visibility data in the order supplied by the synchronous computer (frequency-baseline-time order, abbreviated f-b-t, where the first parameter changes fastest), and to produce the needed three-dimensional map. If the discrete Fourier transform is to be used, the processor must include the necessary sorting and gridding. For either the discrete or direct Fourier transform, provision for user-selected weighting of the uv-plane measurements is highly desirable. Ths processor is expected to be capable of handling up to 12 hours of data at a time, and it must finish computing and storing the resulting maps in less than 12 hours; it is highly desirable that this processing time be a very small fraction of 12 hours. With these specifications in mind, possible designs for synthesis processors are discussed below.

a. A design based on optical processing

General-purpose digital computers usually process one number at a time. When the data base becomes as large as it will be in the VLA line system we must look for more efficient data processing procedures. Optical analog computers are capable of performing convolutions, cross-correlations, and Fourier transformations on large two-dimensional arrays while acting, essentially, on the full array at one time. In this section we will examine the principles of optical computers and the nature of currently available devices. We will describe an optical computing system capable of handling the spectral line synthesis problem and attempt to answer the fundamental questions of speed, accuracy, and cost.

We may provide a general summary of optical processor characteristics by noting that they are:

1) Capable of Fourier transforming 1000 x 1000 data arrays with high accuracy.
2) Capable of other mathematical operations on large arrays including addition, multiplication, convolution, and cross-correlation.
3) Inherently fast - requiring about one second for each 1000 x 1000 Fourier transform.
4) Suitable as display devices having very high spatial resolution and dynamic range - useful for baseline time, u-v, and map data displays.
5) Suitable in interactive systems because of their speed and display qualities.
6) Capable of noise levels due to scattered light more than 40 db below the brightest map point.
7) Capable of 3° - 6° (rms) optical phase abberation. This abberation is not time-varying, and can be calibrated and removed.
8) Capable of 1000 x 1000 space-bandwidth products in transducer-based systems and 4000 x 4000 products in film-based systems, and are being developed to 4000 x 4000 in transducer-based systems.

At the heart of the optical computer is the input modulator or transducer which acts as the interface between the digital or analog electronic signal and the optical wave front. We will review a considerable number of such active elements during the detailed design stages. However, at present, we will discuss only the three representative and promising devices described in Table III-4. Two of these are electro-optic transducers: the Coherent Light Valve ("CLV") produced by the Electronics Laboratory of the General Electric Company and the Pockel's Readout-Optical Modulator ("PROM") produced by Itek Corporation. Such transducer systems can only hold one two-dimensional array at a time and, hence, require an external memory for the full three-dimensional data base. The third possible system uses a film based modulator ("FM") and usually, but not always, uses photographic emulsions. In this system no external memory is required. Instead, a spatial array of u-v planes, one for each frequency, is recorded (by analogue methods) as the observation proceeds. FM systems suffer from the inability to access the data in baseline-time order for re-calibration and re-editing and from logistical difficulties. Since electro-optical transducer systems are no more expensive than FM systems, we will devote most of our attention to the CLV and PROM. However, if optical memories (which are film based) are found to be attractive as storage media, then the use of film-based modulators must be reconsidered.

This section reflects an early decision to use optical methods only for asynchronous data processing. In principle, it would be possible to construct optical computers which operate on the IF signals from the antennas to produce radio maps. Because such computers would not allow baseline-time data to be examined and calibrated and because current electro-optical technology is probably not able to support such computers, we will not consider them further.

TABLE III-4

	C.L.V.	P.R.O.M.	Photosensitive Films
Modulation	Phase	Amplitude or phase	Amplitude or phase
Optical Aperture	Now 20 x 20 mm Near future 26 x 26 mm	30 x 30 mm 40 x 40 mm	very high unlimited
Resolution (linear)	Now 50 c/mm Near future $\simeq$ 125 c/mm	coherent recording 500 c/mm incoherent imaging 100 c/mm	$>10^3$ c/mm
Dynamic Range	50 dB	60 dB	30 - 40 dB*
Linearity	Good	Good	Fair
Input	Electron beam (electronic deflection)	Laser beam (mechanical or acousto-optical deflection)	Electron beam or laser beam
Carrier frequency Modulation	2.5 to 3 cycles/samples	3 to 4 cycles/sample	10 cycles/sample
Aberrations and Correction	Oil film + substrate + vacuum windows $\simeq\lambda$ rms can be corrected by servo control of the phase $<\lambda/60$ predicted for tests in November and December '76.	Crystal + coating + polarizer $\simeq\lambda/20$ rms can be reduced by polishing of the crystal $<\lambda/60$	Film + substrate + liquid gate with 2 liquids $\simeq\lambda/20$ rms. Film coating and substrate dependent
Lifetime	1000^+ hours	No known limit	Unlimited

* The optical computers using a liquid gate in radar processing have a dynamic range for all components about 70-80 dB.

We turn now to the basic physics and a short mathematical description of an optical Fourier transform device. Although a number of optical configurations are possible, the one given in Figure III-7 is representative.

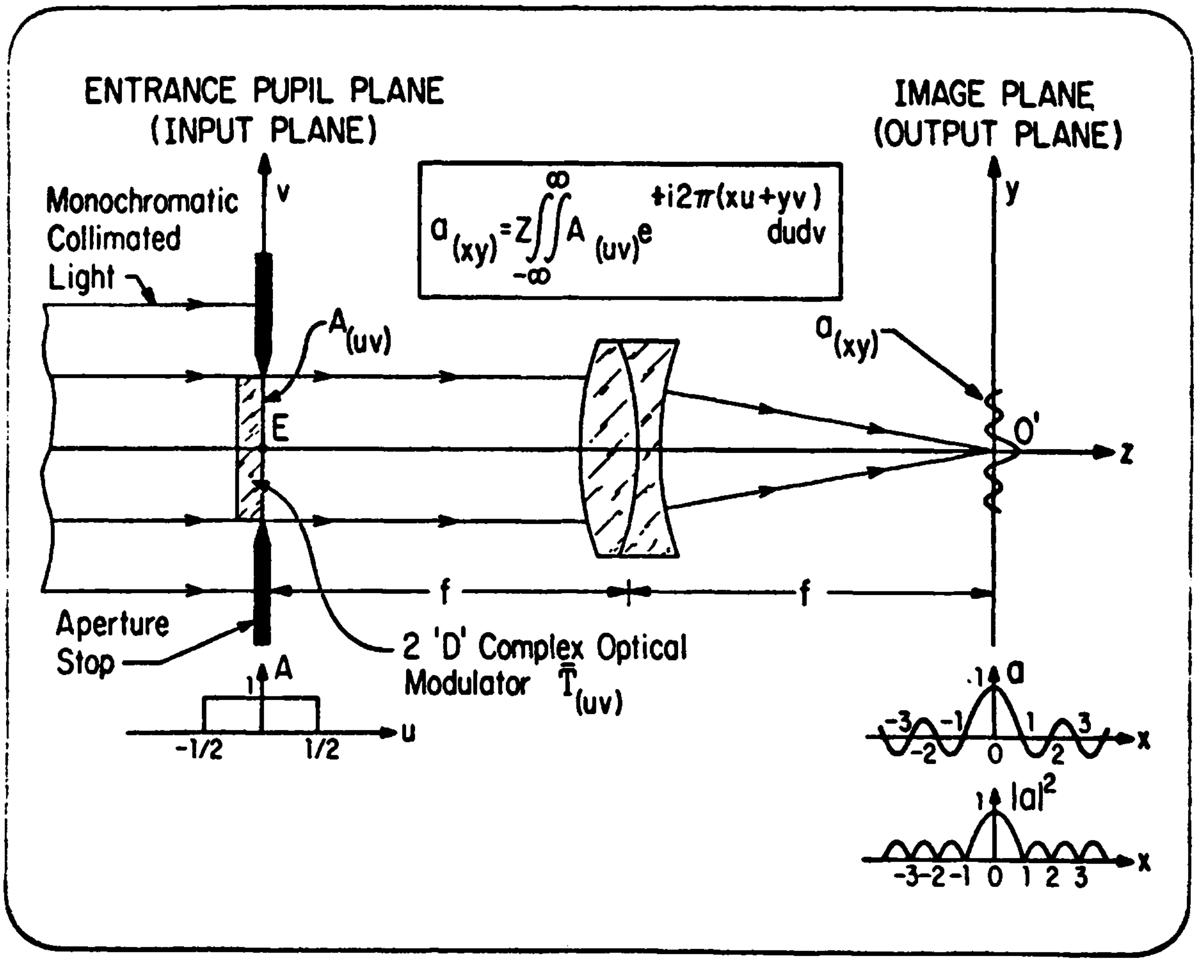


FIGURE III-7

In this system, the complex optical field a(x,y) in the output plane is related to the complex optical field A(u,v) in the input plane by

$$a(x,y) = Z \iint_{-\infty}^{\infty} A(u,v)\, e^{i2\pi(ux+vy)}\, du\, dv, \tag{1}$$

where Z is a complex constant. The input plane contains a two-dimensional modulator or transducer capable of controlling, at each point in the plane, the amplitude and phase of an incident optical plane wave. Although reflection is sometimes used, most transducers operate by transmission and are characterized by their complex transmittance function $T(u,v) \equiv |T(u,v)|e^{i\tau(u,v)}$.

It is possible to modulate the amplitude transmission $|T(u,v)|$ and optical phase delay $\tau(u,v)$ independently. However, it is more practical to generate a quasi-sinusoidal spatial signal of the form

$$D(u,v)\cos[\omega_c u + \phi(u,v)] \tag{2}$$

where ω_c is called the spatial carrier frequency and where D and ϕ are real functions such that D exp $i\phi$ is proportional to the measured complex visibility function, and then to use this signal to modulate either $|T(u,v)|$ or $\tau(u,v)$ but not both. As we shall show shortly, the Fourier transform of (2) is then computed by the optics, and it contains the desired transform of the visibility function.

Optical sensors respond to the square of the modulus of the output optical field. But it is the field itself which is proportional to the desired Fourier transform. Therefore we add to the output plane a uniform optical reference wave (equivalent to a spatially uniform local oscillator) whose phase can be changed by π radians. The sensor output is then proportional to

$$S_{\pm}(x,y) = |a(x,y) \pm R|^2 \tag{3a}$$

where R is the (uniform) complex amplitude of the reference wave, and the sign depends on the reference wave phase. To obtain a linear output we record the difference between S_+ and S_-:

$$M(x,y) = S_+ - S_- = 2\,\mathrm{Re}(R)\,a(x,y) \tag{3b}$$

To provide a specific example, let us consider a CLV processor in detail. In the CLV device, a deflected and modulated electron beam is used to generate a charge distribution on a deformable plastic ("oil") surface. The equilibrium among electrostatic forces, viscosity, and surface tension results in a surface deformation which is a function of the charge distribution. If we arrange the deflection and modulation of the beam so that the surface deformation is proportional to (2) then we will have generated a two-dimensional, quasi-sinusoidal surface of frequency ω_c, spatial phase $\phi(u,v)$, and amplitude $D(u,v)$. When light passes through this surface in a direction perpendicular to the u-v plane, the net optical path difference is given by

$$(n-1)\,D(u,v)\cos\left[\omega_c u + \phi(u,v)\right] \tag{4}$$

where n is the refractive index of the oil. Note that for such surface deformation modulators the amplitude transmission is constant. Only the relative phase of the optical wave is altered by the surface profile on the oil. The complex transmission of this modulation may then be written as

$$T(u,v) = C \exp\left\{ i k (n-1) D(u,v) \cos\left[\omega_c u + \phi(u,v)\right]\right\} \tag{5}$$

where C is a complex constant and k is the optical wave number.

We may now describe the overall operation of a CLV direct Fourier transformation system. A plane wave $A\, e^{i(\omega t-kz)}$ traveling along the z axis is incident on the complex modulator T(u,v) located at z = 0. The exit wave is given by

$$A(u,v) = T(u,v)\, A\, e^{i(\omega t - kz)}\Big|_{z=0} \tag{6}$$

The lens performs a spatial Fourier transform, which results in a field in the xy-plane given by

$$a(x,y) = Z \iint_{-\infty}^{\infty} T(u,v)\, A\, e^{i\omega t}\, e^{-i2\pi(ux+vy)}\, du\, dv. \tag{7}$$

Substituting for T(u,v) from (5) we find

$$a(x,y) = Z C A\, e^{i\omega t} \iint_{-\infty}^{\infty} e^{ik(n-1)D(u,v)\cos[\omega_c u + \phi(u,v)]}\, e^{i2\pi(ux+vy)}\, du\, dv. \tag{8}$$

The identity

$$e^{ir\cos\theta} = J_0(r) + 2\sum_{N=1}^{\infty} i^N J_N(r)\cos N\theta$$

may be substituted to obtain

$$a(x,y) = Z_1'\iint_{-\infty}^{\infty}\left\{J_0(r) + 2\sum_{N=1}^{\infty} i^N J_N(r)\cos N[\omega_c u + \phi(u,v)]\right\} \cdot e^{i2\pi(ux+vy)}\,du\,dv \tag{9}$$

where $Z' = ZCAe^{i\omega t}$ and $r = r(u,v) = k(n-1)\,D(u,v)$.

The right-hand side of (9) is the Fourier transform of an infinite series. Note that each term in the series is quasi-sinusoidal at a harmonic of the carrier frequency ω_c; by Fourier transform theory, the various terms will be spatially separable in the output plane, appearing at intervals of $\omega_c/2\pi$ along the x axis. The N = 1 term contains the information of interest to us:

$$a(x,y)\Big|_{N=1} = Z_1'\iint_{-\infty}^{\infty} 2i\,J_1(r)\cos(\omega_c u + \phi)\,e^{i2\pi(ux+vy)}\,du\,dv. \tag{10}$$

For small values of r, we have

$$J_1(r) = \frac{r}{2} - \frac{r^3}{16} + \frac{r^5}{384} - \ldots \approx \frac{r}{2} \tag{11}$$

so that (10) becomes

$$a(x,y)\Big|_{N=1} \approx iZ'\iint_{-\infty}^{\infty} r(u,v)\cos(\omega_c u + \phi)\,e^{i2\pi(ux+vy)}\,du\,dv \tag{12}$$

which may be rewritten as

$$a(x,y)\Big|_{N=1} \simeq i\xi_1' \iint_{-\infty}^{\infty} \left[r(u,v)\, e^{i(\phi(u,v)+\omega_c u)} + r(u,v) e^{-i(\phi(u,v)+\omega_c u)} \right] \cdot e^{i2\pi(ux+vy)}\, du\, dv. \tag{13}$$

Recalling that $r(u,v)e^{i\phi(u,v)}$ is proportional to the visibility function, we see from (13) and Fourier transform theory that a(x,y) contains an image proportional to the desired brightness map displaced by $\omega_c/2\pi$ along the x-axis, and it also contains a complex conjugate map displaced by $-\omega_c/2\pi$.

In the operation of practical devices r(u,v) is kept very small. This insures the linearity of $J_1(r)$, as assumed earlier. If r(u,v) is limited to 0.089 radians, then $J_1(r)$ is linear to about 1 part in 1000. Typically, however, linearity of 1 part in 100 is used and r may be correspondingly larger.

Let us now consider an integrated digital-optical processor capable of carrying out the VLA spectral-line synthesis task. A general block diagram for such a device is given in Figure III-8. The data from the synchronous subsystem are sent to a fast-access memory in frequency-baseline-time ("f-b-t" where frequency varies most rapidly) order. This memory must be able to accept data either from current or from archived observations. To be compatible with electro-optic Fourier transformers, the memory must have an output rate on the order of 2-200 x 10^5 words/second in a t-b-f or a b-t-f order. The memory technology and the reordering of the data are significant concerns, and at this writing possible designs have not been extensively explored. the most attractive

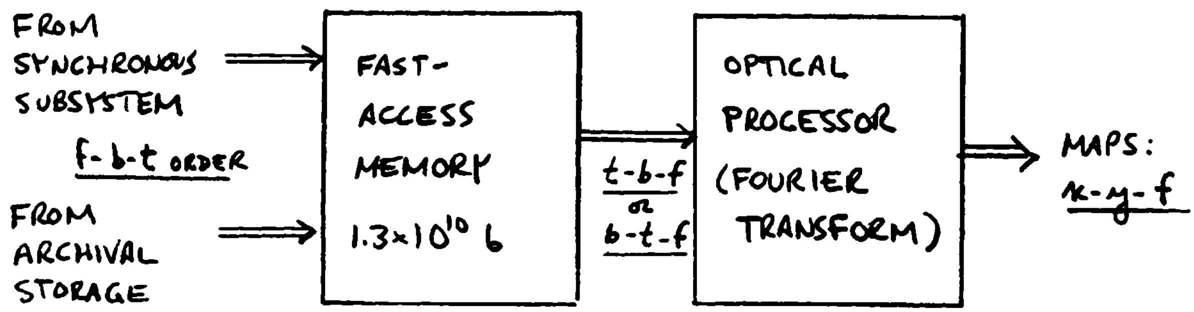


Figure III-8

in terms of speed and cost effectiveness are filmbased; conventional magnetic materials are also possible but less attractive.

The electronic and optical portions of the hybrid system are shown in more detail in Figure III-9. The input electronics block controls the deflection and modulation of the signals sent to the light modulator (transducer). The optical block contains the transducer (shown here as a CLV) and lenses needed to perform the direct Fourier transform of the input u-v plane into the output x-y map plane. Tapering may be done with filters in this optical system. A second set of optical elements may be used to form a direct image of whatever is written on the transducer. The output electronics block consists of diode arrays and A/D converters to produce digital output for later processing plus immediate display devices such as film and television.

Figure III-10 is an enlarged detail of the input electronics block. For Fourier transformation the input data consist of the real and imaginary parts of the visibility (immediately converted to amplitude and phase), the time and baseline, and a measure of the local density of data samples N. For display operations, the input data can also include calibration source and digital array data. The deflection generation block operates either in raster (rectangular displays) or elliptical (u-v plane generation) modes. The latter may be implemented by using transformation matrices or, preferably, by computing and using a dc bias plus two sine waves to generate elliptical Lissajou patterns (baseline tracks). In the modulator block, the visibility amplitude and phase are used to modulate an internal local oscillator. Weighting (multiplication by 1/N) and tapering functions further alter the amplitude of the LO before it is used to modulate the electron beam of the CLV. The LO is also used to synchronize the full system and to insure that the electron beam modulation and deflection are phase coherent.

In all the examples above, we have described CLV-based systems. PROM transducers are also very promising. The PROM is basically an electro-optical crystal surrounded by transparent insulators and electrodes. Using incoherent blue light (~430 nm) an image is stored on the device as a charge and electric field distribution in the crystal. Linearly polarized red light (e.g. 633 nm from a HeNe laser), after reflection or transmission through the PROM, carries the full image in the cross polarization. Dynamic ranges as high as 60 db are obtained, but are currently limited by residual strain in the crystals and the contrast limitations of the polarizers.

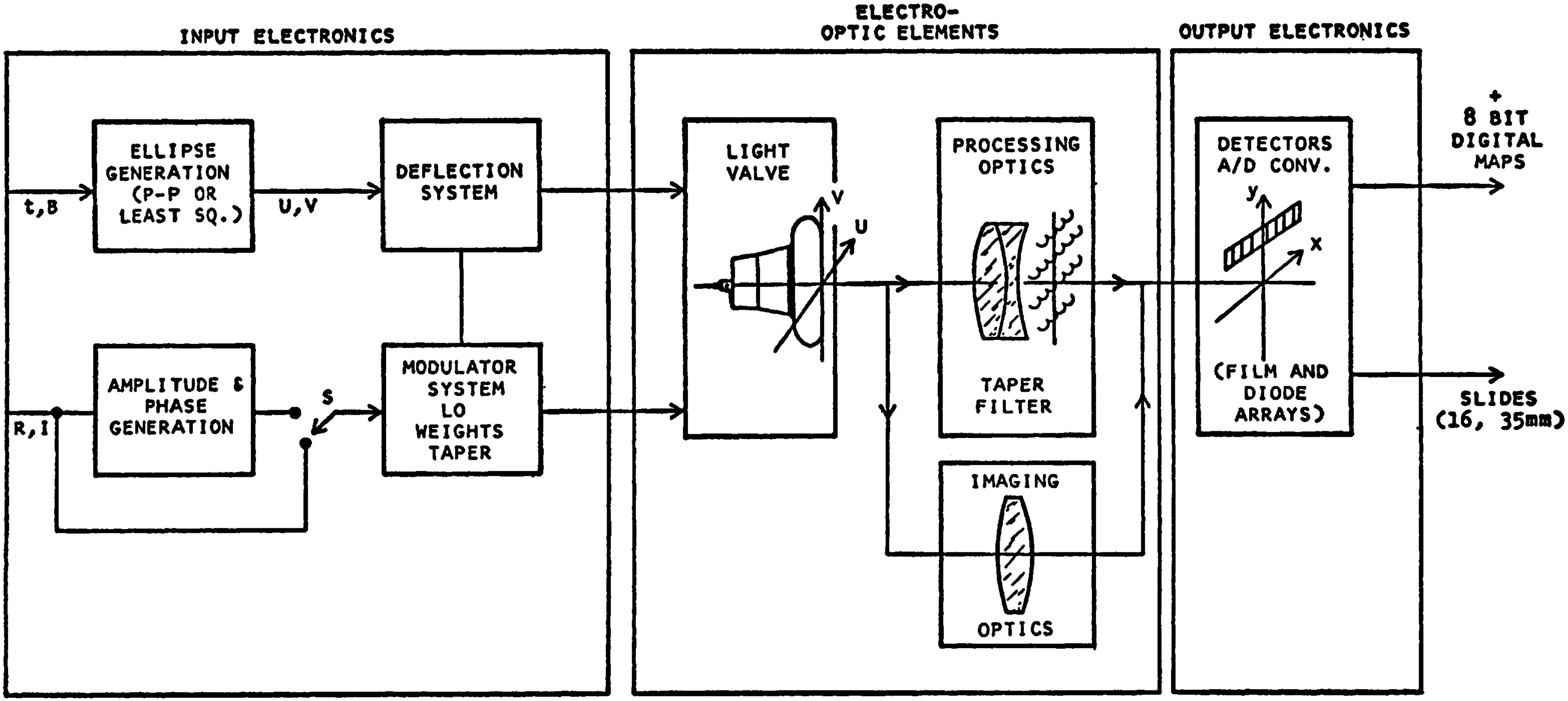

INPUT ELECTRONICS
ELECTRO-
OPTIC ELEMENTS
OUTPUT ELECTRONICS
t,B
ELLIPSE
GENERATION
(P-P OR
LEAST SQ.)
U,V
DEFLECTION
SYSTEM
R,I
AMPLITUDE &
PHASE
GENERATION
S
MODULATOR
SYSTEM
LO
WEIGHTS
TAPER
LIGHT
VALVE
V
U
PROCESSING
OPTICS
TAPER
FILTER
IMAGING
OPTICS
DETECTORS
A/D CONV.
y
x
(FILM AND
DIODE
ARRAYS)
+
8 BIT
DIGITAL
MAPS
SLIDES
(16, 35mm)

Figure III-9

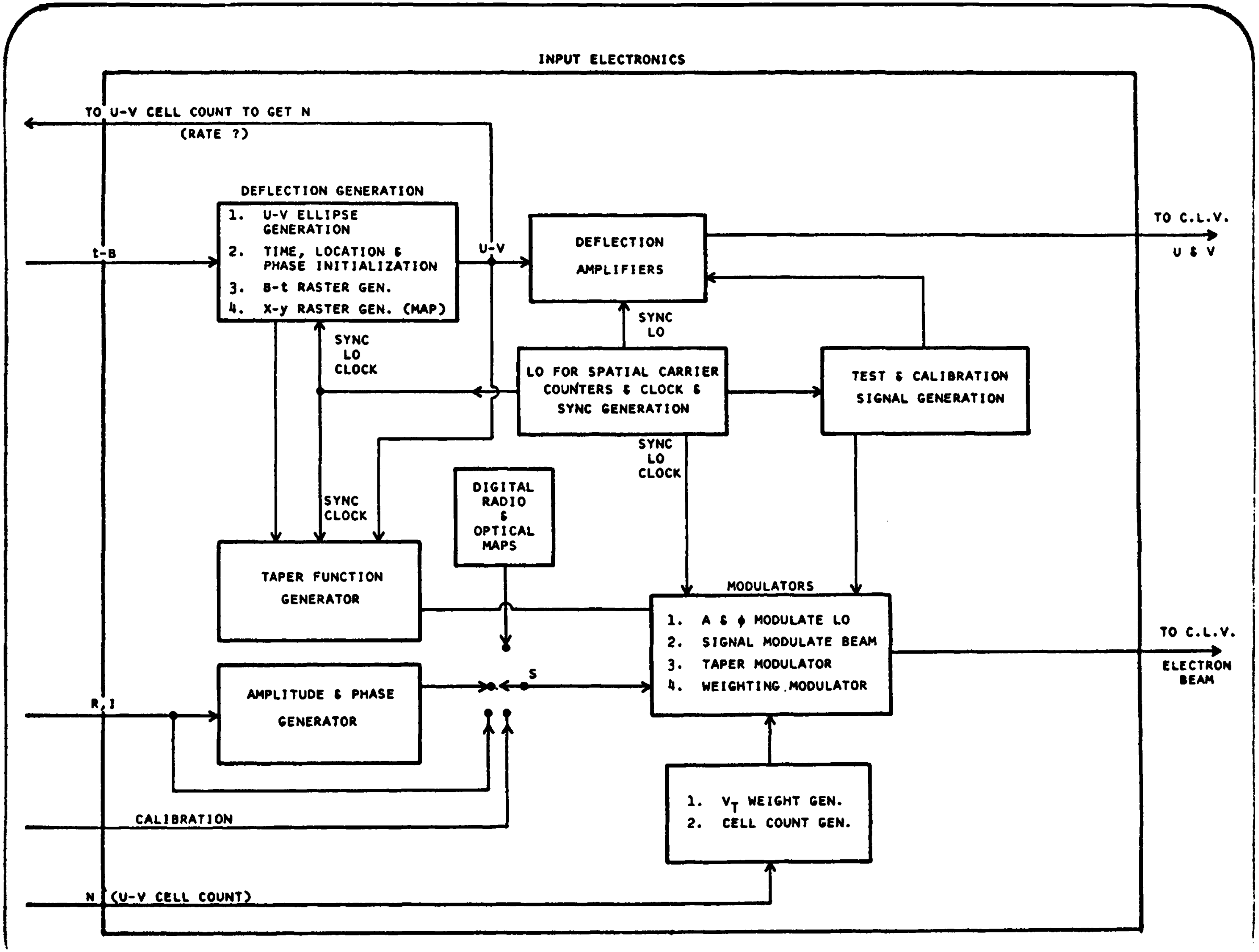


Figure III-10

A conceptual schematic of a PROM system is shown in Figure III-11. PROM can be used both with and without the spatial carrier. The latter requires separate processing of the real and imaginary parts of the visibility but has the advantage of a four-times increase in space bandwidth. The u-v track generation, spot deflection, weighting, tapering, reference-wave linearization, and Fourier transform and display outputs of both systems are very similar. The local oscillator used as the time reference for writing the spatial carrier in the CLV is replaced, in the PROM system, with a spatial reference grid. This grid may be a ruled grating or it may be generated interferometrically by the superposition of two optical plane waves. This grid acts as the spatial phase reference and must be accurate to one part in 150,000. Because the signal written on the PROM does not decay, the high data rates ($\sim 2 \times 10^7$ words/second) required by the CLV are not needed in the PROM system. A feedback system to improve deflection accuracies is also illustrated in Figure III-11.

Film based processors would be virtually identical to the PROM processor, with the PROM replaced by film. Although film processing and transport would be completely automated, a few seconds delay would be required to develop the film completely. The film would be mounted in a liquid gate to minimize optical path variations arising from variations in emulsion thickness. The limits to the accuracy to which this can be done have yet to be determined; $\lambda/20$ rms is typical. Film processors are somewhat unattractive because of the logistical problems inherent in chemical processing and film consumption.

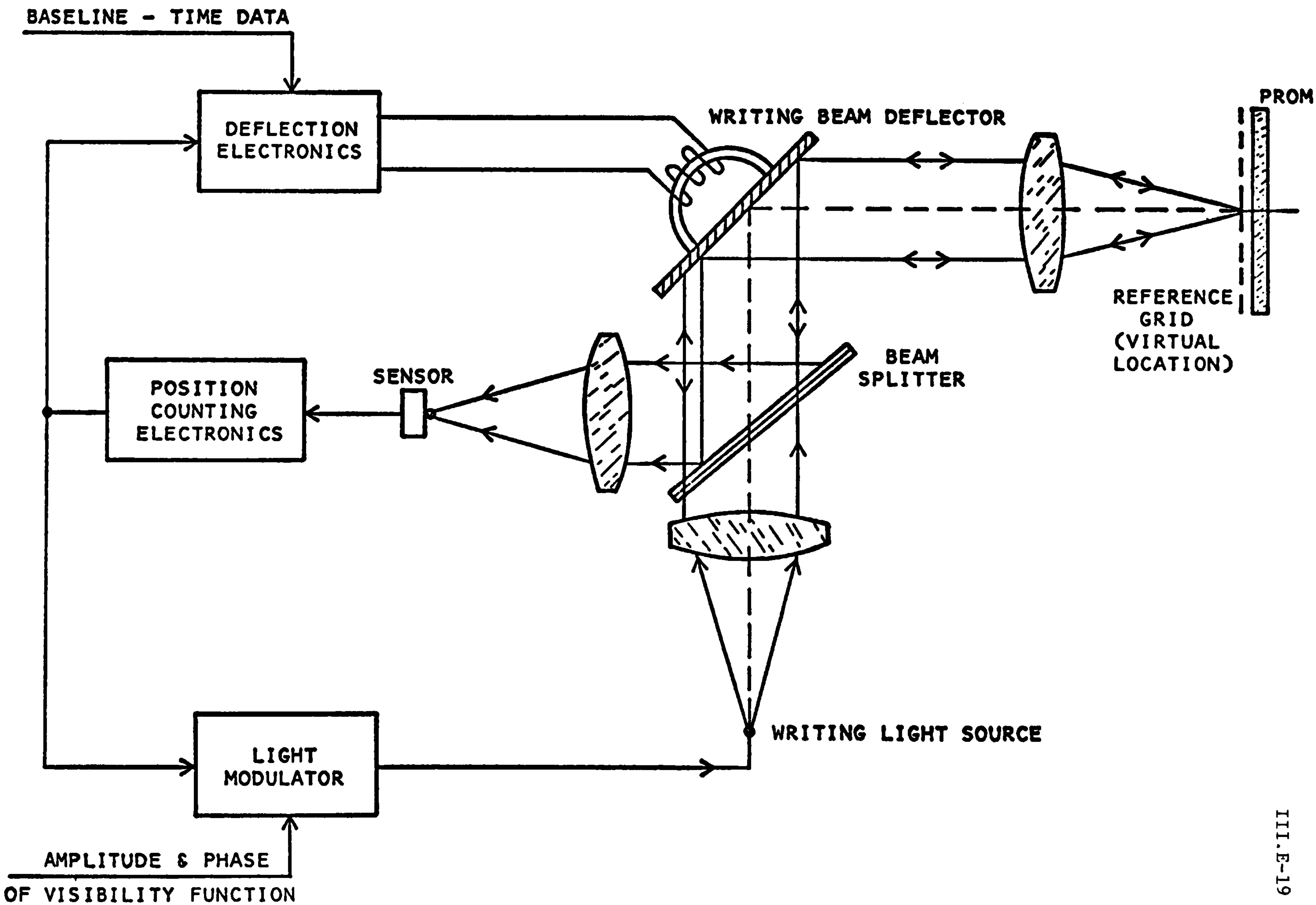


Figure III-11

For a more detailed technical treatment of optical processors, including the expected performance of various optical configurations considered for the VLA, the reader is referred to a study by Bulabois (1975).

b. Digital Processing

For digital computation of the required three-dimensional map, we consider first the use of the two-dimensional discrete Fourier transform (implemented by the FFT alogrithm) on classically gridded and apodized visibility data for each frequency channel. This procedure will produce low-distortion maps under most circumstances, provided that the (at most) 1024 x 1024 points of each channel-map are used at two or more points per synthesized beamwidth. However, at $\lambda/21$ cm and with arrays A and B only, the outer portions of such maps will be seriously distorted because the procedure ignores the three-dimensional locations of the baselines (by projecting them all into the uv-plane); considerably more computation is needed to avoid such distortion. We begin by considering only the situations where the classical procedure is adequate.

The two-dimensional discrete Fourier transform implies the need to arrange the data on a rectangular grid in the uv-plane; this is easily accomplished (by cell averaging or similar schemes) if the visibility measurements can be accessed in u-v-f order. However, the measurements are supplied by the synchronous subsystem in f-b-t order. One way to arrange for access in u-v-f order is to physically re-order the data storage (sort the data). The general problem of data sorting has been well studied (e.g. Knuth 1973), and efficient algorithms are known. Nevertheless, the time required to accomplish such a sort for

12 hours of VLA spectral line data, using currently available off-the-shelf digital hardware, strongly dominates the synthesis processing. The sorting time depends mainly on the amount and kinds of memory available to the processor and on the amount of data to be sorted.

For the VLA, the amount of data produced in 12 hours depends on the correlator setup, in particular on the integrating time and the number of spectral channels. This leads to the idea that, for a given amount of digital sorting hardware (principally memory) and a given number of channels (say 256), there is a critical integrating time below which the time required to sort a 12 hour block exceeds 12 hours, resulting in a growing backlog. For example, a particular design for a sorting processor which we have investigated has a critical integrating time between 30 and 40 sec; it uses a PDP-11 minicomputer and eight DEC RP-04 disks, and would be easily interfaced to the DEC-10. A more extensive sorter, designed to achieve a critical integrating time of 10 sec or less, requires three minicomputers (Varian V72's) and ten 400-megabyte disk spindles (see Appendix B).

The maximum integrating time which may be used without significant smearing of the visibility function depends on the baseline and on the source size, but we may take as a lower limit the time required for the longest projected baseline to move through one antenna diameter. This number is tabulated in Table III-5 for each VLA configuration. (Also given in the table is the number of data bits produced in 12 hours if the integrating time is equal to the limiting value rounded to the nearest multiple of 10 sec.) For many experiments it may be that the use of integrating times somewhat longer than these limits will produce no significant degradation of the map, but the exact effects are not well understood. In particular, a 40 sec integrating time is clearly adequate for the C- and D-arrays, and probably also for B, but not for A.

Table III-5: Time For Longest Baseline To Change By 25 Meters

Config.	Longest Baseline	t (25m)	bits/12hours
A	35 km	9.8 sec	12×10^9
B	10.7	32	4×10^9
C	3.25	106	1.1×10^9
D	0.98	350	0.3×10^9

Depending on costs, we feel that it may be worth considering the construction of a sorting processor which can keep up with 256-channel data only if the integrating time is $\gtrsim$40 sec, on the assumption that shorter integrating times will be needed only occasionally, in which case more than real time could be devoted to sorting those observations. Calculations indicate that about 80 hours might then be required to sort 12 hours of data with 10 sec integrating time.

Another option, not yet fully studied, is to avoid a complete sorting of the data on u and v by immediately gridding the visibilities; that is, in one pass through the data, we add each measurement into its rectangular cell in the uv-plane, keeping track of the number in each cell. This requires specifying the cell size in advance, and is an irreversible operation. Sorting by frequency is still required. It is not yet clear whether a significant amount of processing time can be saved in comparison with the general sort; this depends on the detailed characteristics of the hardware used. Also, if numerical fields of view as large as 2048 x 2048 are allowed (in order to cover the single-antenna beam in the A-configuration), much more memory will be needed than for the general sort.

Once the visibilities have been sorted and gridded, or just gridded directly, we must multiply by an apodizing function and compute the FFT for each frequency channel. Applying the apodizing function would require several hours of CPU time in a general purpose computer like the DEC-10 (for 12 hours of data at 10 sec), but the use of a peripheral array processor such as the DEC-10 will have available should reduce this time considerably. The two-dimensional FFT for each frequency channel requires a substantial amount of computation. The array processor just mentioned can compute FFTs in 1.3 μs $\cdot$ N $\log_2$ N, where N is the number of complex points being limited by the DEC-10's core memory. The two-dimensional transform is efficiently computed by a series of one-dimensional transforms on the rows, followed by the same on the columns. This requires that the array, viewed as a matrix, be transposed between the row and column trnasforms. Since a 1024 x 1024 word array cannot fit in the DEC-10's core memory, external storage and possibly an external processor are needed to accomplish the transpose. Calculations indicate that about 1.2 min is needed to compute a 1024 x 1024 word two-dimensional FFT, including transpose, using the DEC-10 CPU, its RP-04 disks, and the array processor; thus 5.1 hours are needed for 256 of them.

We now consider the problems introduced when the out-of-plane baseline components are significant. This happens only when the relatively large field of view accessible at 21 cm (34 arcmin single antenna HPBW) is mapped to high resolution. It has been proposed (Clark 1973) that the three-dimensional locations of the baselines be accounted for by gridding the visibilities into uvw-space and computing a three-dimensional transform, evaluating the result only on the surface of the unit sphere in xyz-space. For a 1024 x 1024 map at two points per synthesized beamwidth, a uvw grid of dimensions 1024 x 1024 x 3 should be adequate for both the A and B configurations at 21 cm. The visibility measurements can be fully sorted on u, v, and w with the same effort required for a u and v sort (it depends only on the amount of data). Direct gridding would also require the same computing effort, but would require a factor of 3 more memory, most of which would be empty. On the other hand, the Fourier transform would require at least a factor of four more effort, so that additional equipment would be needed in order to finish in less than 12 hours.

The above results are derived and further discussed in Appendix B. Also considered there is the fact that to map sources which extend to the single-antenna half-power point in the A-configuration, or to beyond this point in B-configuration, more than 1024 x 1024 map points are needed. Although it exceeds our specification, a design is developed which will keep up with (1) the complete uvw-sort of 12 hours of data at 10 sec integrating time, and (2) the Fourier transforming of a 2048 x 2048 x 4 grid, for each of 256 channels. This and other designs which have been studied are summarized in Table III-6.

TABLE III-6: SUMMARY OF DESIGN STUDIES FOR DIGITAL SYNTHESIS

(In all cases, 12 hrs, 256 channels, and 351 baselines are assumed)

Task	Equipment	Cost Est.	Time Est.
Sort, t_{int} = 10s	3 V72's(1) 10 400 M byte disks(2)	\$497K	10 hrs
Sort, t_{int} = 20s	2 V72's 4 400 M byte disks	293K	8.8
Sort, t_{int} = 40s	1 V72 1 400 M byte disk	200K	7
FT, 2048x2048x4	4 V72's 4 400 M byte disks 1 Fast FFT	797K	7.4
FT, 1024x1024x3	1 V72 1 400 M byte disk 1 Fast FFT	340K	5.6
FT, 1024x1024x1	(use DEC-10/array proc.)	--	9

Notes: (1) Varian minicomputer system, with 660ns core.
(2) e.g. Calcomp 1035/235-4 disk.

One synthesis processor has been designed in some detail, based on the philosophy that it should be easily added to the asynchronous computer hardware and sortware already on hand or on order for the continuum system. The hardware additions consist mainly of a minicomputer (PDP-11/45) with 128K 16-b words of core and 6×10^9 b of disk storage, which would be inserted between the synchronous and asynchronous systems and would handle the visibility sorting in real time. Complete sorting with 256 channels should be possible for integrating times ≥ 40 sec. Gridding, weighting, and Fourier transforming would be accomplished using the DEC-10, its disks, and the array processor, as outlined above, taking ~5 hours for 256 1024 x 1024 transforms. Figures III-12 and III-13 show the asynchronous computer subsystem before and after the proposed additions for spectral line work. The equipment mentioned above forms an "initial sorting system", for which a possible equipment list is given in Table III-7. It is suggested that this equipment be duplicated to form an "off-line sorting system" which would provide backup, handle some overflow work of the initial sorting system, and allow some non-real-time operations such as the merging of data taken on different days. (Other equipment shown in Figure III-15, including a "high density output medium", "mass store", and a minicomputer-based set of graphics equipment, will not be discussed here.) The current price of the equipment in Table III-6 is $400K, but equivalent equipment can be obtained for $200K to $300K; thus a cost estimate for the initial and off-line sorting systems is $400K to $600K.

1976 Asynchronous Hardware

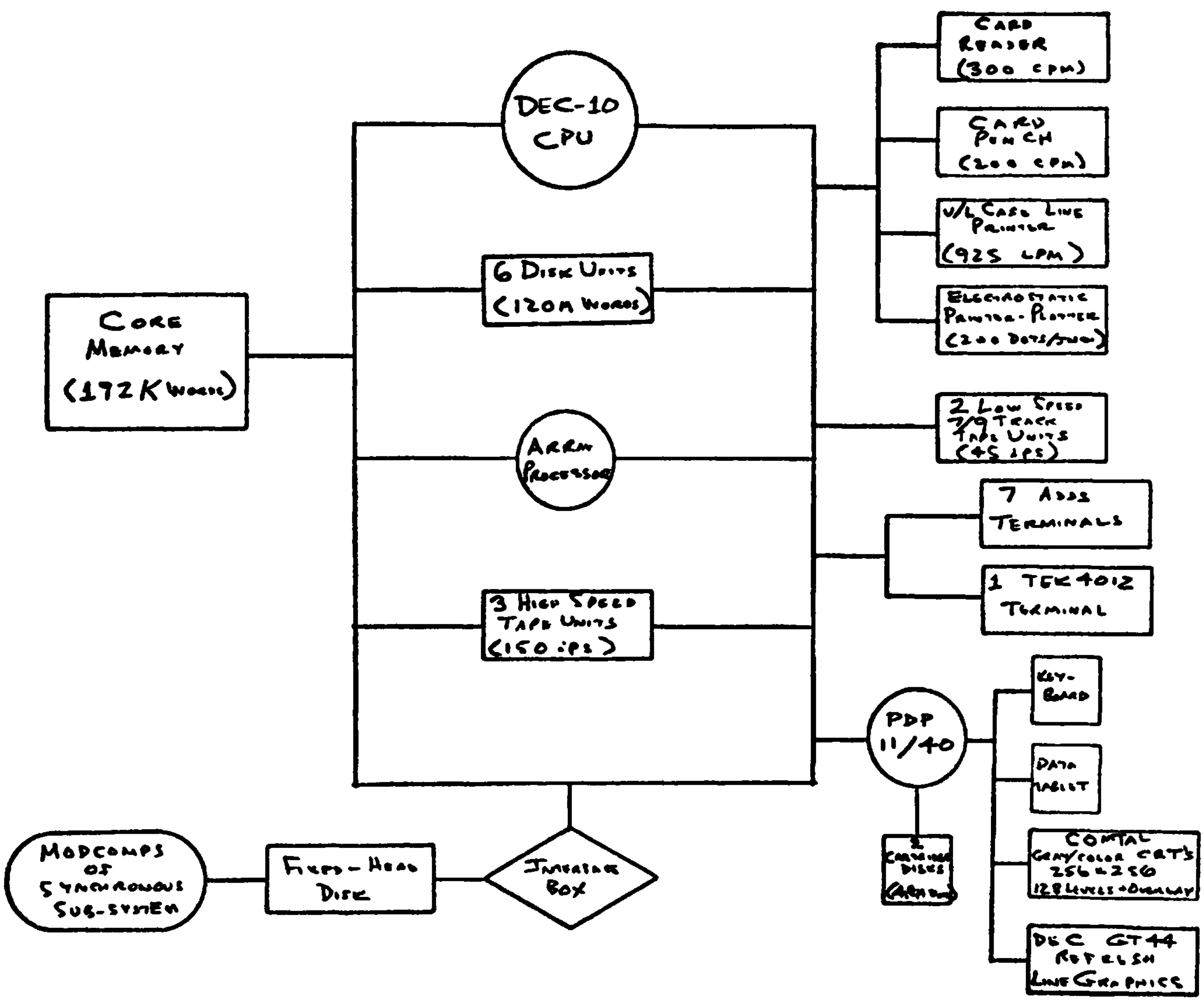


Figure III-12

1980 Asynchronous Hardware

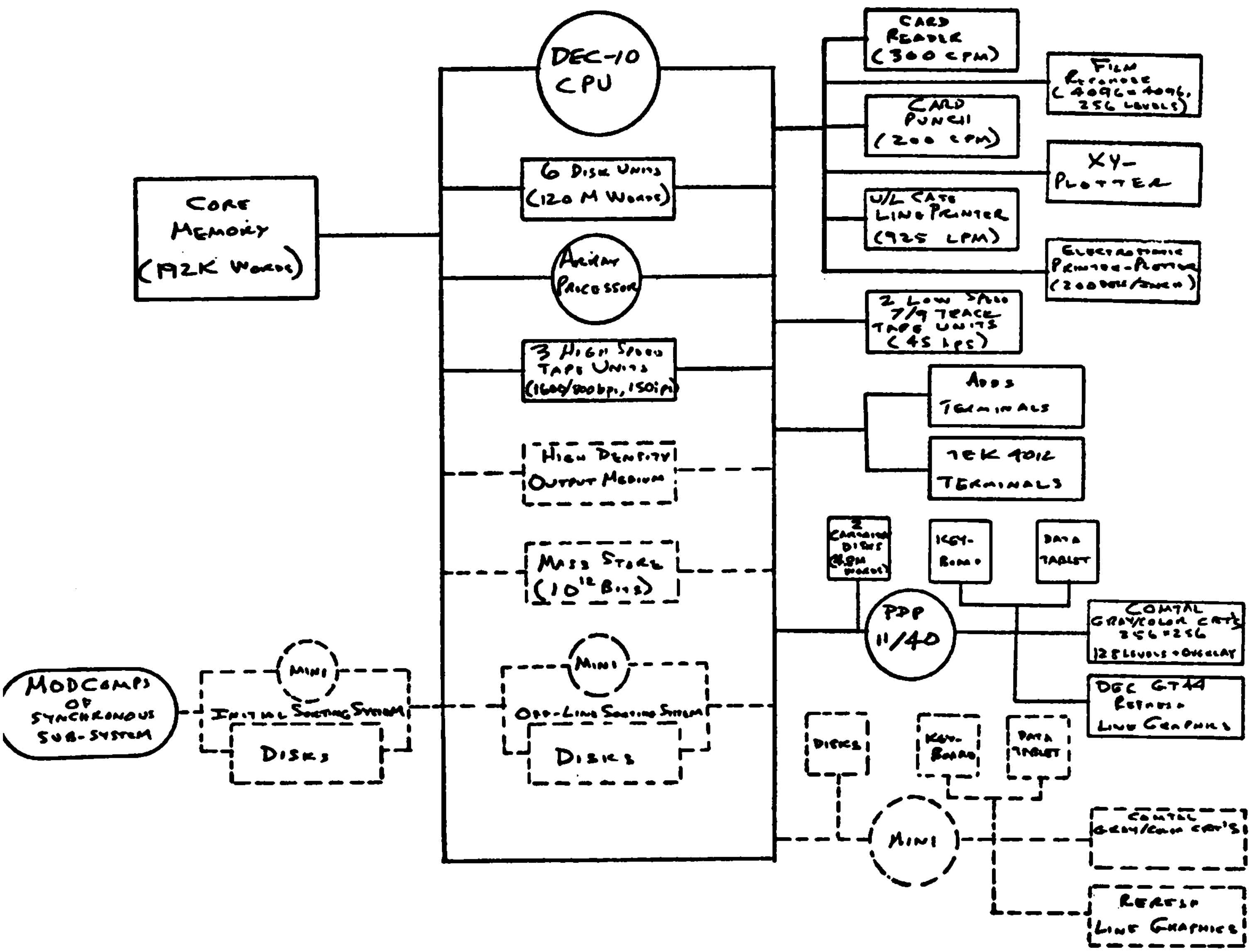


Figure III-13

Table III-7: Possible Equipment List For Initial Sorting System

1 DF10 Data Channel	connecting the disks of the sorting system to a DEC-10 memory port.
1 PDP 11/45 CPU	with 128K words (16 bit) of core to take the data from the interface to the shnchronous sub-system, carry out the sort, and format the results in the mass storage device.
1 RH10 Disk Controller	which can handle up to 8 RP04 disks.
1 RJP-04 Massbus	for communication between the PDP 11/45 CPU and the disks of the sorting system.
8 RP04 disks	with dual ports accessed by DEC-10 or PDP 11/45 on to which the data is sorted.

3. Active Storage

a. Requirements. In Chapter II we defined "active storage" as that information which is readily accessible to the VLA computers. In the asynchronous subsystem, such storage is required for several purposes, each of which imposes different requirements for storage capacity, access time, and access organization. In order to make specific estimates, we shall assume that most storage in the subsystem is contained in a central set of files maintained by the DEC-10. The files would be kept on a variety of media, in order to satisfy the requirements of the various tasks. The synthesis task is an exception; in view of the results of Section II.E-2, it appears worthwhile to have a large mount of storage - enough for slightly more than one 12-hour data block* - dedicated to the synthesis processor, rather than shared in the central files. This arrangement of the asynchronous subsystem's active storage is depicted schematically in Figure III-14.

* Define a VLA 12-hour data block as the maximum amount of visibility data to be generated in 12 hours, which occurs for 256 spectral channels, 351 baselines, and 10 sec integrating time. With 16-b words it equals 1.242×10^{10} b.

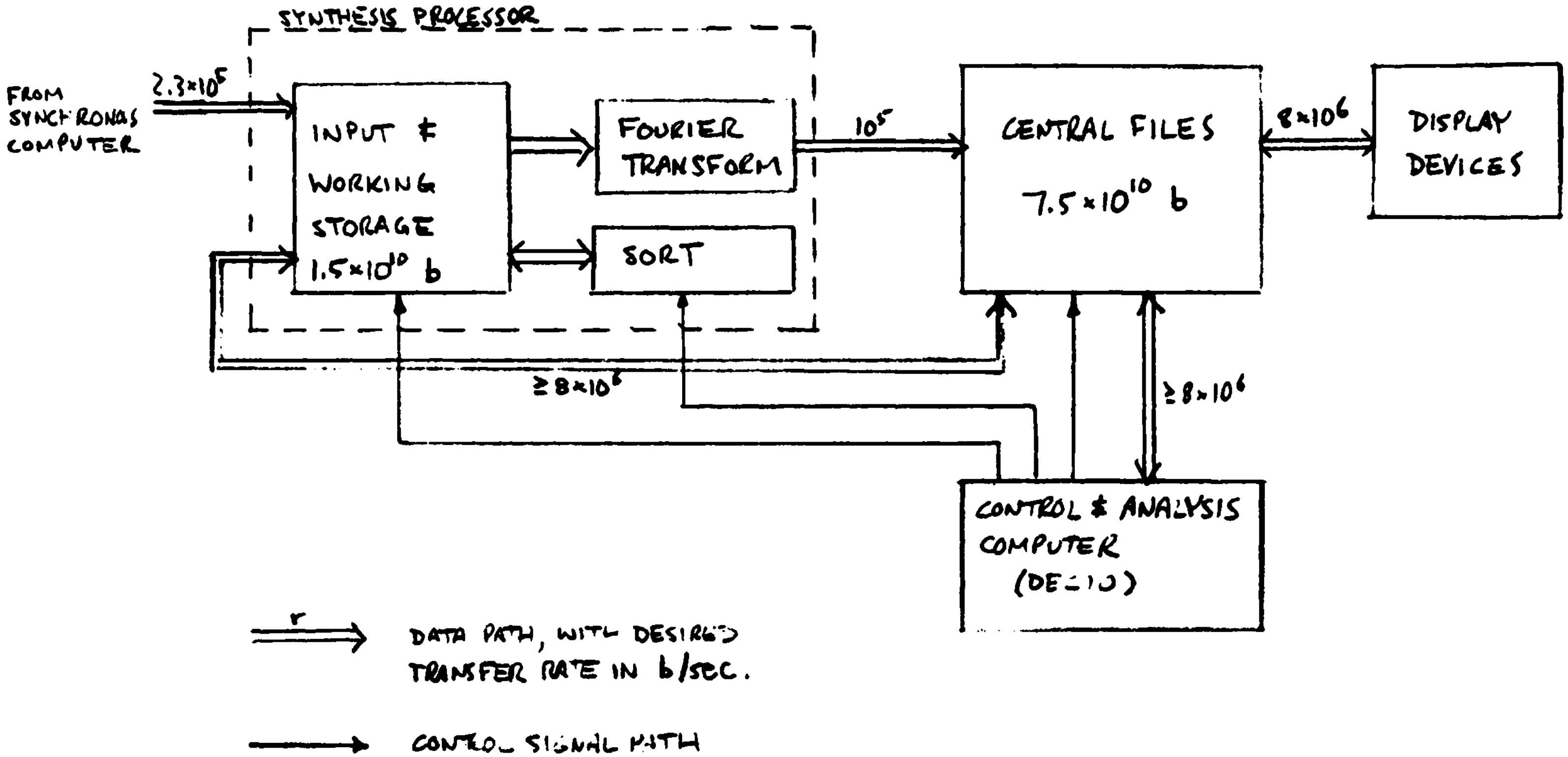


Figure III-14

Asynchronous Subsystem Configuration Assumed for the Purpose of Estimating Storage Requirements

The display task imposes severe requirements on the active storage. In order to provide even simple flexibility features (such as the ability to display the spectrum at a particular map point as well as the map at a particular spectral point) which operate in a reasonable time, it is necessary to have at least one complete three-dimensional map in random-access storage in such a way that the average access time to any point is ~10 msec. (Suppose that a 256 x 256 display is to be created; at 10 msec per point, this takes 11 minutes.) If the user is to be able to display and possibly modify his visibility data, then it too must be in such a direct-access store.

The analysis task may ultimately impose the most severe storage requirements, in terms of access capability. However, as we indicated earlier, the initial system will support only as much computational analysis as can be implemented in software, using the DEC-10 and its associated peripheral processors, and this is not expected to be storage-limited. To allow some analysis, however, we must at least provide the user with working storage for partial results; enough direct-access storage for one complete three-dimensional map should be adequate.

If we assume that an astronomer with a 12-hour observing run will not complete his display and analysis jobs within the 12 hours immediately following the observation, then the system must support the storage described above for more than one user. In principle, not everyone needs to have his data on the high-speed, direct-access media at once; those who are not currently "logged on" could have their data moved to sequential media, such as tape, thereby freeing the direct-access media for others. But in practice, because of the large amounts of data involved, the time required for such data movement may be excessive (41.7 minutes for 2.5×10^{10} b at 10^{7} b sec^{-1}). Thus, depending on the costs of storage media and devices, it may be reasonable to provide direct-access storage for several users (or 12-hour runs) simultaneously. Here we assume that enough for three users, each with one 12-hour data block, one maximum-size map, and one map work space (a total of 2.5×10^{10} b per user, including 20% overhead) is provided.

The above discussions are summarized in Table III-8. For three users, this suggests a total of 7.5×10^{10} b of direct-access storage in the central files, plus 1.5×10^{10} b dedicated to the synthesis processor. Also indicated in the table are estimates of the desired transfer rates for each task, and the desired random access times for records which are (1) anywhere in storage, (2) within one three-dimensional map of 12-hour data block, and (3) within one two-dimensional map or spectral channel.

Note that Figure III-17 is not a design proposal, nor is Table III-7 a definite set of specifications. They serve merely to allow some reasonable estimates of our requirements to be made. We turn now to a brief survey of the state-of-the-art in data storage, in view of these requirements.

Table III-8: Summary of Storage Requirements

Task	Total Storage bits	Transfer Rate bits/sec	Random Access Time		
			Full Storage	Within 12-Hr Blk	Within f-Channel
Synthesis, input and work space:					
Optical Processor	1.5×10^{10}	10^7			
Digital, Compressed Data	5×10^9		50 msec	50 msec	--
Digital, Full Data	1.5×10^{10}				
Display: one 12-hr block of visibilities & one 3-D map	2×10^{10}/user	10^6	1-10 sec	10-100 msec	10-100 μsec
Analysis: one 3-D map for work space	5×10^9/user	10^6	1-10 sec	10-100 msec	10-100 μsec

b. Survey of storage technology, 1975.

(1) Magnetic tape. Recent advances have pushed the recording density on standard half-inch computer tape to 6250 b/inch, which means that, at 9 tracks and 2400 feet per reel, 10 reels are needed to hold a 12-hour data block. Drives capable of 6250 b/inch are now or will soon be available from several manufacturers. Much higher densities have been demonstrated, using wideband instrumentation records; more than 25,000 b/inch with 28 tracks is possible, which allows a 12-hour data block to fit on one reel. While such equipment is not yet available off-the-shelf, JPL and the NRAO will soon be obtaining specially-built drives of this type for use in VLBI.

(2) Magnetic disks. Disc storage provides a reliable and cost effective means of storing random access data in the range of 10^6 to 10^{11} bits. For large scale storage (10^7 bits and above) the most common units presently in use are the 800 Mb (single density) and 1600 Mb (double density) spindles of the IBM 3330 type. Technical specifications include rotational position sensing, average access times of ~30 msec, average latency times of ~8 msec, transfer rates of ~1.6 Mb/sec and usually a capability for dual controllers. A number of manufacturers make such units.

Units having larger capacity have recently been announced by Calcomp and IBM. The Calcomp unit allows two spindles per drive with a capacity of 3200 Mb per spindle. Combining such units one can expect to achieve a capacity of 7.7×10^{10} bits (12 units - 3 controllers) without the physical space required being too unreasonable. Such a system would cost $312 K (12 x $17K + 3 x $36K). IBM's dual spindle units have a combined capacity of 5080 Mb. Most of these high capacity units involve non-removable disc packs, which is not considered here a severe limitation for use with the VLA.

Disc units of the type described here would almost certainly be used as a conventional solution to the VLA spectral line sort problem, if the system were to be implemented today. Their use would also be of interest for staging into any mass store which itself did not have a random access capability.

For special problems discs and drums having unusual specifications can be obtained, but usually at a significant cost increase per bit. Channel rates of 10^8 b/sec are available as are multiple head drums allowing parallel access. Most likely, such special devices would be used only in connection with an analog device having some peculiar interface specification.

(3) Mass Storage Devices. In this category we place several recently developed devices which allow very large amounts of data - up to $\sim 10^{12}$ bits - to be accessed without operator intervention. Random-access times are typically 5 to 15 seconds, with transfer rates of 3 to 10 Mb/sec. Some devices allow random access to large portions of the data within times comparable to disks, ~100 msec, while others use a "staging" disk to achieve the same result. Data on specific devices now available on the general-purpose computer market are presented in Table III-9.

It may also be possible to obtain custom-designed devices which give a closer match to our needs than those available on the general-purpose market.

Table III-9: Mass Storage Devices, 1975

Manufacturer/Device Name	Min. Configuration			Add-on Units		Random Access Times			Transfer Rate	Notes
	Size	Price	Price/Size	Size	Price	Full	In Block	Blk Size		
	bits	$	¢/bit	bits	$	sec	sec	bits	b/sec	
Ampex/TBM	$9x10^{10}$	461K	$5x10^{-4}$	$9x10^{10}$	100K	15	Staging disk		10^{7}	fast video tape
CDC/38500	$1.2x10^{11}$	431K	$4x10^{-4}$			5	Staging disk		$6x10^{6}$	8M byte tape cartridges
IBM/3850	$2.8x10^{11}$	657K	$2x10^{-4}$	$1.4x10^{11}$	70K	16	Staging disk		$7x10^{6}$	50M byte tape cartridges
Precision Inst./ System 190	$1.3x10^{11}$	440K	$3x10^{-4}$	$1.3x10^{11}$	165K	10	.22	$1.6x10^{9}$	$3x10^{6}$	Permanent recording by laser beam. 200 M byte strips.

c. Projection of Technology to 1980.

(1) Magnetic tape. Further increases in the recording density are likely to occur, although not by an order of magnitude. Drives which operate reliably at >30,000 b/track-inch, with 20 to 50 tracks on 1- or 2-inch wide tape, could become off-the-shelf items by 1980 if there is a large market for them; they should certainly be available as specially-built devices.

(2) Magnetic disks. Again, some increases in recording density are likely, but no revolutionary changes are expected. However, continued price reductions should occur due to competition and production improvements.

(3) Optical memories. In this category we include devices which use light for reading and/or writing digital or analog data. Most development efforts seem to be aimed at digital storage, and fall into two categories: bit-by-bit and holographic recording. The Precision Instruments System 190 (Table III-9) represents the 1975 state-of-the-art for bit-by-bit recording, with a density of 3×10^4 b/mm^2. The fundamental optical limit to the density is ~10^6 b/mm^2 at visible wavelengths, and for technical reasons this can be better approached by holographic recording. Various difficulties have so far kept holographic recording from progressing beyond the laboratory, and the prospects for 1980 are not clear. Optical storage is advantageous compared with magnetic methods not only in the achievable recording density, but more importantly in the ability to randomly address very large blocks of data; for example, an entire 12-hour data block can be written on a single sheet of film with access to any word possible in ~1 msec. Various materials into which

data can be stored optically have been studied. Some, such as silver halide film and laser-etched surfaces, record permanently; others are alterable, including photochromic materials (whose transmission or reflection spectrum is altered by exposure) and magneto-optic materials (whose optical polarization transmission is altered by changing magnetization, usually using photon-induced heating). (Chen and Zook, 1975.)

(4) Other technologies. Several other memory technologies are on the horizon, and we mention some of them here without much discussion.

A large industrial effort to produce "video disks" for the consumer market (plastic, phonograph-size records containing ~1 hour of television) may produce valuable spinoffs in the mass storage field; each disk can hold at least 10^{11} bits.

Another emerging technology being pursued by several laboratories is that of magnetic bubbles. Memory modules can be fabricated on chips, externally similar to LSI circuits, but with potentially higher packing densities than semiconductor memories (a 65K b chip is now under development) and with non-volitile, read-write-erase storage. Bubble memories are not expected to be economical for mass storage of the size we require, but it is hoped that they will eventually fill a gap between core/semiconductor fast RAMS and disk, providing storage capacities of 10^6 - 10^8 bits with random access times ~5 msec. However, it remains to be seen whether the cost per bit can be made competitive.

One additional technology of interest is the beam-addressed semiconductor memory. The memory element is an unstructured MOS chip in which data is stored as small islands of charge which can be written and read by an electron beam. About 3×10^7 bits can currently be stored in a single module, with a random access time of 30 μs and a transfer rate of 10^7 b/s. The memory is non-volitile, at least on a time scale of several days.

A summary of the estimated cost per bit for various storage technologies, both now and in 1980, is given in Table III-10. (Much of the data for this table is from papers in Proc. IEEE 63, Aug. 1975.) From the preceeding discussion and the table it should be apparent that memory technology is undergoing rapid development, so that large advances can be expected over the next 5 years.

Table II-10: Cost Per Bit of Memory - Various Technologies

	1975		1980	
	System Size bits	Cost Per Bit ¢	System Size bits	Cost Per Bit ¢
Semiconductor	10^8	$3x10^{-2}$	10^8-10^{10}	$3x10^{-2}$
Disk	$7x10^9$	10^{-3}	$7x10^8$	$5x10^{-4}$
Mass store, tape based*	10^{11}-10^{12}	10^{-4}	10^{11}-10^{12}	10^{-4}
Mass store, optical	10^{11}-10^{12}	$2x10^{-4}$	10^{11}-10^{12}	$\leq 10^{-4}$
Magnetic bubbles	$<10^6$**	?	$\sim 10^8$	?
Beam-addressed MOS	$3x10^7$**	$2x10^{-2}$**	10^9	10^{-2}

* e.g., IBM 3850

** Laboratory prototypes only

4. Off-Line Storage

a. Archives. If we assume that one copy of the calibrated visibilities and one, standard-parameter, three-dimensional map will be archived for each observation, then we find that we must store

1.67×10^{10} b	per 12 hours
2.17×10^{11} b	per week (6.5 days)
1.09×10^{12} b	per year (50 weeks)

in the worst case. It may be that these numbers can be reduced by as much as a factor of 10 if most observers (a) use fewer than 256 channels (e.g., continuum observations), (b) use integrating times longer than 10 sec (easily possible for C- and D-arrays), or (c) compute maps smaller than 1000 x 1000 points (although the latter saves only a little storage). Nevertheless, the required storage is substantial. It would take up about 5000 reels of 6250-b/inch tape per year, or would fill any of the mass storage devices of Table III-9 (minimum configurations) in about one month. Most of the mass storage devices, on the other hand, have removable media on which data is stored at substantially higher density than 6250-b/inch tape; for example, the Precision Instruments System 190 would store one year's data in a volume of 0.3 m^3 (including protective cannisters). Furthermore, advanced optical storage devices (film or video-disk based), not yet commercially available, will be capable of data densities at least a factor of 10 greater.

Besides the storage density, important parameters of the archival storage are its access time and transfer rate. Access time can of course be much slower than for active storage; we assume that one hour is a satisfactory time to transfer a random data block of interest to active storage. This allows operator intervention. If we further assume that we usually want to transfer blocks of 2.5×10^{10} b (visibilities and two maps for a 12-hour observation), then storage media containing this much data should be easily handled by a human operator, and a reading rate of at least 10^7 b/sec is needed. To store 2.5×10^{10} b, we need about 17 reels of 6250 b/inch tape; 2 reels of 28-track, 25,000 b/inch tape (Mark III VLBI tapes); 63 cartridges for IBM's 3850 mass memory; two "packs" of Precision Instruments data strips (20 strips); or half of a video disk.

b. User output. It is clearly necessary to be able to store data in a form that the astronomer can retain for his own use, independent of VLA facilities. We assume that this will include both human-readable and machine-readable outputs.

For human-readable output, it is possible to produce photographic images (slides or prints) for each spectral channel which preserve the full space bandwidth and most of the dynamic range of the numerical maps, and moreover to do so on a reasonable time scale (a few hours with commercially available recorders, much less with a specially-built device). Such photographs might be considered the primary output of the telescope. The only other human-readable outputs which we expect to be needed are printed text (line printer) and line drawings (pen plotter), although the amount of data which can be usefully recorded in these ways is very limited. Further discussion of human-readable output, including devices like CRTs which provide no off line storage, is given in the next subsection.

For machine-readable output, some considerations similar to those for archival storage apply. One user's data should fit on a medium which he can easily handle and transport, and it should be capable of being written and read in an hour or less. Additionally, since the data must be readable at many user-operated facilities, well-standardized media and formats should be used (unless special-purpose readers can be made very inexpensive). For the user who wishes to retain a full 12-hour data block plus one or two 1024x1024x256 maps (2.5×10^{10} bits total), these requirements are at present contradictory; the highest density "standard" medium is 6250 b/inch tape (although not many units are yet installed), of which 17 reels would be reqsired. Higher-density media may be usable in the 1980's; video disks and other optical devices are promising, as are higher-density magnetic tapes. Meanwhile, it is apparent that most users will be able to retain only a portion of their maps or visibilities in machine-readable form.

5. Display Devices

In this subsection we review presently available display equipment and comment on its usefullness to the VLA observer. Both "hard-copy" devices and "interactive" (CRT screen) devices will be considered. It should be kept in mind, especially in connection with interactive displays, that the capability of a display device is strongly influenced by the storage to which it has access. For example, a CRT device may be capable of accepting data at 200,000 bytes/sec, so that a 256x256-pixel display could be set up in .33 sec (1 byte/pixel); but if the display is complex so that the 256x256 = 65,536 pixels are not in contiguous storage, then, allowing 20 msec average access time, 22 minutes are needed to assemble the display.

a. Basic types of computer graphics. One type of computer graphics produces a hard copy, e.g., a piece of paper or film that the user can put in his pocket or send to his publisher. Another type of computer graphics produces a picture on the face of a CRT screen.

Another way to classify computer graphics is according to the way that the picture is produced. In one type, the picture consists of lines, for example, a pen is moved to draw a line on a piece of paper, or an electron beam is deflected to produce a line on a CRT screen. Some line drawing systems can produce lines of several different intensities or several different colors. It should be noted that most line graphics devices can also easily produce alphanumeric characters. In other types of graphics the picture is produced by a raster scan. This works in much the same way that your TV at home works. Most raster scan systems allow several gray scale (intensity) levels, and some will produce color pictures.

Graphics systems which produce pictures on a CRT screen can be further subdivided into storage-tube systems and refresh systems. With a storage-tube system, the electron beam draws the picture once and then the CRT screen stores the picture. With a refresh system, the pictures quickly fade from view and must be continuously redrawn.

The obvious advantage of devices which produce a hard copy is the fact that they produce a hard copy. Also, line drawing devices, e.g., plotters, can easily produce large pictures which have high resolution and show the data in great detail. A disadvantage of hard copy devices is that the picture is of course static. No part of the picture can be moved or changed without producing a whole new picture. Also hard copy devices are frequently relatively slow.

Devices which produce line graphics on a refreshed CRT screen have the advantage that the picture can be easily moved or changed. For example, a three dimensional data plot can be displayed and rotated and moved as the user requests. A disadvantage of this type of device is that there is a limit to the number of lines that can be displayed on the screen. Thus, in theory a CRT screen cannot display pictures which are as large and detailed as the pictures produced by some hard copy devices. However, as a practical matter, plotters are rarely used to produce plots which are more detailed than those displayable on some refresh line drawing CRT systems. Also, it is important to keep in mind that this potential disadvantage is counterbalanced by the fact that pictures on a CRT screen can very rapidly change and respond to the user. (It should be noted that most line graphics devices can also easily produce alphanumeric characters.)

CRT devices which use a storage tube are not limited to the number of lines that can be displayed. However, with a storage tube CRT a part of the picture cannot be changed or moved without erasing the screen and redrawing the picture.

Raster scan CRT devices can produce much more complex pictures than can line drawing CRT devices since each spot on the screen can have its own intensity or color value. However, with a raster scan device it is more difficult to move or change a part of the picture. Although lines and characters can of course be formed on a raster scan CRT, the resolution will generally not be as good as with line drawing CRT's.

Since no single type of computer graphics is best for all applications, the question is not which type is needed, but rather what collections of types will produce the balance of capabilities which will enable the user to make the most effective use of the VLA data processing system. It should be emphasized that this is a critically important question in the overall design of the VLA software. This is due to the fact that graphical portrayal of data will necessarily be a major part of the interface between the user and the enormous complexity and data rate of the VLA. If the user cannot comprehend his data or if the operators cannot comprehend the functioning of the instrument then the VLA project will, to put it midly, have problems.

Without going into the details of the needed graphics system, some general requirements are apparent. There will clearly be a need for hard copy equipment. This is due to the need for preparing publishable results. Also, there will probably be occasions when it will be necessary to have the large and very detailed display of data which only hard copy can provide. Line drawing hard copy equipment (such as a Calcomp plotter) will be useful for such things as plotting one variable as a function of another variable. Gray scale hard copy equipment (such as a Dicomed device) will be useful for producing pictures of radio intensity maps which can easily be compared with optical photographs.

There will also be a need for line drawing and raster scan gray scale CRT devices. One use of these will of course be to provide a "quick look" capability so that the user can see what a picture looks like before taking the time to have a hard copy made. In addition, an important use of CRT grpahics results from the fact that refreshed CRT devices can present dynamic pictures in which all or part of the picture moves or changes with time. This arises because not only are human beings marvelously well adapted to processing visual information, they are also extremely good at processing dynamic visual information. It is not uncommon for a person to look at something and not immediately comprehend what he is seeing. If the object does not move, the person will move himself so that he sees the object from a constantly shifting angle. By seeing the view of the object smoothly change with time, the person usually quickly comprehends the nature of the object.

In order for the user to effectively use a dynamic graphics display he must be able to interact with the picture. Thus, there must be a means for the user to supply input to the system. One input device is of course the keyboard. Another is the lightpen, which the user points at the CRT screen to indicate a position or a part of the picture. The data tablet is another device which allows the user to input position information. Control knobs can be conveniently used for some types of input. For example, the position of a knob can control the amount or rate of rotation of a three dimensional plot.

In summary, an effective computer graphics system will necessarily consist of more than one type of device. Choosing the proper set of devices will require careful consideration of the strengths and weaknesses of human beings as well as the strengths and weaknesses of the different types of computer graphics devices.

b. Hard-copy devices. The line printer has been the most fundamental hard-copy display device. Currently, typical impact printers achieve about 1000 lines/minute printing rates with full character sets. At this rate, the worst case map (1024x1024x256) could be printed with a two-digit representation in just under 70 hours not counting the time to add paper. Some printers such as the Telex 6721 are capable of about twice this speed with 16-character print sets. Non-impact printers are capable of speeds of 12,000 to 18,000 lines/minute. There is also a wide variety of microfilm and microfiche printers currently on the market. They are available at printing speeds from 5000 to 27,000 lines/minute. Many of them also have line plotting capabilities. We should consider the use of microfilm or microfiche because of the large volume, weight, and cost of computer paper and because of the greater speed of the film printers. The drawbacks of such printers include their (currently) ienfficient interfaces to the computer systems and the need for potentially costly viewing devices.

Traditional line plotters come in three general types; flat bed and drum plotters use pens and ink in drawing lines, while electrostatic plotters use a point plotting method and chemical processing. The electrostatic plotters can be very fast. For example, the Versatec 1200 can plot a contour map with spatial resolution of 2112 x 2112 points in about 11 seconds. A full set of contour maps would then take less than one hour to print. The quality of the resulting plots would be satisfactory for user examination, but not (directly) for publication. Drum plotters come in a range of sizes and costs and are capable of plotting accuracies in the range 0.002 - 0.01 inches. Flat bed plotters are capable of higher accuracy but tend to be, depending on size, very expensive. Such plotters tend to be very fast in terms of maximum pen speed. However, properly annotated contour maps consist of very large numbers of quite short lines and hence require plotting times, depending on map complexity, from 1 to 10 minutes or more per map. Pen plotters also suffer from problems of maintaining proper ink flow over long periods of time.

Several conclusions may be drawn from the preceding discussion. The traditional display media - impact printers and pen plotters - are too slow for displaying full size three-dimensional VLA maps. Even 256x256x256 maps would require inrodinate amounts of time on these media, except that electrostatic printers offer sufficient speed for contour plots. Since the output of such devices, used heavily, would be very voluminous, we must consider the use of microfilm or microfiche printers. Printers and plotters are very useful hard copy display devices, but they will have to be used sparingly in the VLA line system.

Among film recorders, the principal one currently in use is the Dicomed Color Image Recorder. The device is capable of 4096x4096 spatial resolution with 256 gray-scale levels. The device requires about 10 minutes to record the average 4096x4096 picture with annotation. However, with 1024x1024 picture size (and sensitive, high resolution film), it should be possible to record 256 channel maps in under 4 hours. A good quality photo-processing laboratory located on site and staffed by VLA personnel would be required to support the Dicomed device.

c. Available technical options for a raster-scan CRT display.

Addressable pixels. Systems vary in the number of resolvable pixels. Typical resolutions are 100x100, 256x256, and 512x512. For reasons we will shortly discuss the basic cost of a system is largely determined by the number of resolvable pixels available.

Bits per pixel. The number of bits per pixel used to store the information about each pixel determines the number of different gray scale (intensity) levels or colors which can be contained in an image displayed on a raster scan CRT.

Number of gray scale levels or colors. The number of different gray scale levels or colors which can be displayed does not necessarily correspond to the number of bits per pixel. For example, there may be two bits per pixel, yielding four possible values for each pixel. However, the system may be capable of displaying a great many different colors, any four of which can be assigned to the four possible pixel values at any one time.

Memory type. In refresh raster scan CRT systems the stored pixel values are continually read out of the memory to refresh the CRT. In some systems, the memory is a disk. Others use a solid state memory.

Some systems use a storage tube CRT instead of a refresh CRT. In these systems, the memory is the face of the CRT itself. Some other systems use a plasma panel instead of a CRT (we include these with the raster scan CRT systems because they are used for the same purpose, i.e., to display an image which is a rectangular array of pixels). A plasma panel consists of an array of tiny "spots", each of which can be either glowing or not glowing. Like the storage tube CRT, the plasma panel itself serves as the storage medium for the image.

Memory size. Some raster scan systems require a large memory. For example, a system which has (512x512) resolution has 262,144 pixels. If there are eight bits per pixel (256 intensity levels or colors), then over two million bits of storage is needed. This would, for example, use up about one third of the 192K words of core memory in the DEC-10 (one would never use the DEC-10 core for this purpose).

Numbers of images stored. Some systems can store more images than are displayed at one time. This feature would be very useful for such things as comparing the appearances of a radio map at two or more frequencies. For example, two map images could be stored and the user could be supplied with a button to select which map is to be displayed on the CRT. By depressing and releasing the button the user could instantly switch back and forth between the two maps. This feature could be invaluable for purposes of displaying and understanding maps for different spectral line channels.

Mapping hardware. On some systems there is a mode of operation where the stored pixel value is not directly used to determine the displayed intensity level or color. Instead, the stored pixel is used as an index into a mapping table and the corresponding entry in the table determines the displayed intensity or color. This allows changing the displayed appearance of an image by just changing the mapping table rather than changing the entire stored image. In a 512x512 system with 8 bits per pixel, the number of pixels is a thousand times greater than the number of entries in the mapping table. This mapping feature could be used for a large variety of things. For example, constant intensity contour "regions" could be put into an image of a radio map by mapping selected stored pixel values to zero or full intensity. Or, the displayed mapping could be rapdily changed from linear to logarithmic. Another use would be to increase the contrast in the image to more clearly show a feature of interest, e.g., bring out a very weak bridge between two strong radio galaxies. All of these things could be done as instant responses to the pushing of a button or the turning of a knob.

Line generator. Since a raster scan CRT can display any image (subject to the pixel resolution limit), it can of course display images of line drawings. However, putting a line on the display would require changing a great many stored pixel values. Some systems include line generating hardware that automatically changes the appropriate pixels when given only the coordinates of the endpoints of the line.

Character generator. Some systems have hardware character generators to make display of alphanumeric characters (labeling, etc.) easier.

Cursor. Some systems have a cursor which the user can move around on the screen without affecting the picture. This could be used to point to a position in a radio map to allow the user to indicate the points on the map where he wishes to obtain hard numerical information such as position, intensity, etc. The most useful way in which this capability would be implemented would be as a way of giving input to programs that would: (1) give intensity, position, etc. for the current position of the cursor; (2) cause the line graphics CRT to display a plot, with correct numerical information on abscissa and ordinate, for any line drawn on the screen of the raster scan CRT under control of some pointing device such as a joystick, lightpen, or data tablet.

Graphic overlay. Some systems have a graphic overlay capability. This is essentially a one bit per pixel image which is non-destructively superposed on the main image. This would be useful for such things as showing contour lines or coordinate grid lines.

Video camera overlay. In some systems the image from a video camera can be combined with the stored image. This feature could be used to enable comparison of any two "maps" of the sky, e.g., an optical photograph and a corresponding radio map.

Number of independent CRT's. As with line drawing CRT systems, some raster scan controllers can handle multiple CRT's.

Memory read back. In some systems the computer can read back the stored pixel values. This would allow one to retrieve a map image that has been produced with image modification under the interactive control of the user. This would greatly speed up the process of allowing the user to obtain exactly the pictures he wants from the high quality hard copy device.

In some systems the pixel values resulting from hardware mapping can be read back in this way. In others the memory can be read and written in a random access mode and thus can be used as a bulk digital storage medium when not being used to put an image on the CRT.

d. Available options for line-drawing CRT systems. In order to provide a general purpose graphics system to be used in connection with a system of applications programs, display system hardware can be divided into three basic categories: storage tube CRT systems, refresh CRT systems that have hardware transformations and refresh systems that do not. Storage tube CRT systems will not be discussed in detail here. For systems that do not have full hardware transformations, the transformations can be done by software. The result produced by the software is a transformed display file, which is used to refresh the CRT. The "transformed display file" can be composed of several segments, each corresponding to a part of the image to be displayed on the CRT screen. A segment contains instructions which are interpreted by the display processor. These instructions are a description of the two-dimensional image on the screen and are expressed in terms of the screen coordinate system. When hardware transformations are available, a "structured picture definition" is created instead of a transformed display file. It is "structured" in the sense that a total picture can be composed of various "instances" of sub-picture elements placed at various locations in the main picture. The sub-picture instances can also be at various scalings and rotations. The picture can be the protrayal of some data, and is expressed in a coordinate system which is not necessarily the same as the screen coordinate system. The display hardware reads this structured picture definition and performs the transformations necessary to produce an image on the screen which is some particular view of the picture.

The simple refresh display is, in general, much less expensive (considering hardware costs only) than is the structured display, at least in terms of extra cost beyond the basic DEC-10 system. However, one should also realize that one is utilizing a powerful and expensive DEC-10 system to do something that could be done with much cheaper, by comparison, special purpose display hardware. Therefore the comparison between these alternatives should be made in the overall context of the load on both the DEC-10 system and the line graphics system. In balancing these considerations one should also keep in mind the relative amount of software development cost for each type of system.

The software execution time needed to build a structured picture definition is less than that needed to build a transformed display file. This is because the production of a transformed display file requires the execution of transformation software. The difference between the two types of display is even more dramatic when a new view (zooming, rotation, etc.) is to be displayed on the screen. For the structured display, all that is necessary is to calculate a few new parameters; for the refresh display, each new view of a picture requires just as much software execution time as the first view (except that most refresh displays allow rapid two-dimensional translation of the entire picture).

e. Equipment already on hand. Much of the above discussion was considered in obtaining display equipment for the continuum system. That already available is summarized in Table III-11.

Table III-11: Display Equipment Already On Hand
(costs include interfaces to DEC-10 and some software.)

Manufacturer	Model	Description	Cost
Tektronics	4012	Storage tube line-drawing display	$5K
DEC	GT44	Refresh line drawing system, including CRT and display processor, PDP11/40 minicomputer with 28K core and two disk drives (no color, no transformation hardware).	$44.5K
Comtal	--	Raster scan display, 256x256 pixels, 8 b/pixel, b/w and color monitors.	$25K
Summagraphics	--	Data tablet (position input device), 11x11-inch surface, 100 pts/inch.	$4K
Versatec	--	Electrostatic printer, 200 dots/inch, 0.75 inch/sec.	$9K

6. Analysis of Maps

We present here the results of some calculations on the time required for the DEC-10 and its array processor (Spectra Data model 900) to perform a few map manipulations of interest. We assume that a 256-channel, 1024x1024-pixel map has already been synthesized, and we consider the following operations on it:

a. Find continuum: Average together 20 selected channel maps and write the resulting map to a disk.

b. Subtract continuum: Subtract a given 1024x1024-pixel map from each channel map.

c. Mean velocity map: For each pixel in the map, calculate the brightness-weighted mean velocity, and write the resulting mean velocity map to disk.

d. Find continuum point source response: Compute response to a continuum point source over a field twice as large in each dimension as the map (2048x2048 pixels).

e. Subtract continuum point source: Given a flux and position, subtract the point source response from all channels.

These operations are thought to be representative of those which line observers will find useful. Time estimates are given in Table III-12, and were computed using the following assumptions:

-- The 1024x1024x256 map is stored with two pixels per DEC-10 word in fixed-point format; it is available on disks, or can be staged to disks from a larger storage device, so that transfers to and from DEC-10 core memory can be accomplished at 200,000 words/sec plus 50 msec for each access.

-- The Spectra Data processor includes the Array Multiplier option, allowing vector additions and multiplications at 5μs per element and ratios at 10μs per element, plus memory access time (0.61μs per DEC-10 word).

-- The packing and unpacking of two pixels per word is done in the Spectra Data - DEC interface hardware.

-- A large amount of DEC-10 core (up to 100K words) can be devoted to array storage for these analysis tasks.

-- The task under discussion is the only one competing for system resources, or has the highest priority. This is probably not realistic, since real-time mapping tasks (see section III.E.2.b) will probably have higher priority and will make heavy use of the disk channels, core, and the array processor; thus the times in Table III-12 are processor times, and real times may be a factor of 3 to 10 larger.

From these results we conclude that only a very limited amount of map analysis will be possible with the DEC-10/Spectra Data processors. Each observer may have time to do one or two operations like those in Table III-12 before the system becomes oversubscribed. There is some likelihood that an observer can do more sophisticated analysis by restricting operations to a small subset of his three-dimensional map.

But the required preliminary operation of extracting the desired subset will take a time comparable to those in the table.

Table III-12: Use of the DEC-10/Spectra Data For Map Analysis

Task	Time Estimate in seconds if Core Buffer in DEC-10 Words is		
	100K	30K	10K
a. Average 20 channels	210	230	310
b. Subtract continuum	3300	3900	5700
c. Compute mean velocity map	4200	4500	5400
d. Find continuum point source response	470	470	470
e. Subtract continuum point source	3300	3900	5700

F. REFERENCES

Ball, J. A. 1973, "The Harvard Minicorrelator", IEEE Trans. IM-22, 193.

Bulabois, J. M. 1975, "Optical techniques for image synthesis in radio astronomy". VLA Computer Memorandum No. 128.

Chen, D. and Zook, J. D. 1975, "An overview of optical data storage technology", Proc. IEEE, 63, 1089.

Clark, B. G. 1973, "Curvature of the Sky", VLA Scientific Memo. No. 107.

Kunth, D. E. 1973, Sorting and Searching. Volume 3 of The Art of Computer Programming (Reading, Mass.: Addison-Wesley).

Rabiner, L. R. et al. 1969, "The Chirp-z Transform Algorithm", IEEE Transactions AU-17, 86.

Whitehouse, H. J., R. W. Means, and J. M. Speiser 1975, "Signal processing architectures using transversal filter technology". Report, Naval Undersea Center, San Diego.

APPENDIX A

VLA SCIENTIFIC MEMORANDUM 121

Use of the Multichannel (Spectral Line) System for Continuum Observations with the VLA.

Frazer N. Owen

Although most of the planning of the VLA multi-channel correlator system has thus far involved its use for spectral line work, it is quite likely that an equal amount of useful science will be generated by the system's application to continuum observations. For this reason it is important that the scientific requirements of the relevant continuum projects be carefully considered in the design of this system. In this memo I will try to outline these requirements and give examples of some of the types of experiments which would be done with the system.

The primary reason for using the multi-channel system for continuum work is to allow high sensitivity observations over a large field of view by reducing the effects of finite bandwidth (or delay beam). This subject has recently been discussed by D'Addario (1975). Examples of experiments with these requirements are as follows,

1. Determination log N - log S

2. Determination of the flux density, angular size relation for extragalactic sources.

Both of these experiments are classical radio astronomy problems which are relevant to cosmology especially at very low flux densities. Both experiments are also extremely important to the interpretation of other VLA experiments since they provide an estimate of the expected contribution of random background sources to any observation.

3. Determination of the radio galaxy luminosity function in clusters of galaxies.

4. High resolution observations of nearby galaxies.

Both of these experiments have wide spread importance to our general understanding of the radio properties of galaxies and rich clusters and certainly will be among the first experiments proposed for the VLA.

5. High resolution mapping of compact structures contained within low surface brightness radio galaxies, supernova remnants or HII regions.

High resolution mapping of radio galaxies such as Cygnus A and 3C 83.1 are of extreme importance to the physics of extragalactic sources. However many of the prime sources are too large for full mapping in the A configuration and thus could not be observed at full sensitivity and resolution without the multi-channel system. Similar work is also important for galactic sources.

6. Observations of radio star activity in nearby star clusters.

The high sensitivity of the VLA could be used most effectively if a large number of stars could be monitored simultaneously producing an n-fold increase in the efficiency of such searches.

7. Studies of transient sources (X-ray, etc.) with initially poorly known positions (± 2' - 15').

Often the most interesting phase of the radio history of transient X-ray outburst are lost because of initially poorly known positions. With a multi-channel system the VLA would be able to monitor the entire error box of such events at almost full sensitivity. Otherwise, especially when outbursts occur when the VLA is in the A configuration, observations will have to wait days or weeks for refined positions.

The experiments listed above could all be performed with narrower bandwidths or by repointing the VLA to sufficiently many positions in the area of interest. However, since all of these experiments are sensitivity limited, these procedures effectively reduce the VLA sensitivity and make some of the projects impossible in reasonable periods of time.

The variety and significance of the experiments described should emphasize the necessity for the VLA continuum system to deal with the delay beam problems. Unfortunately as shown by D'Addario (1975) this would require channels with bandwidths no wider than 1.7 MHz for the A configuration at 1400 MHz. Even this specification does not completely eliminate the problem; however, it would require 240 such channels per baseline to correlate the full 100 MHz bandwidth and all four polarization channels. Clearly such a system is quite far beyond the possibilities presently being considered. The much modest system presently under consideration with four independent correlator banks each with 64 channels over a bandwidth of 3.125 MHz (4 I.F. channels, 11232 multipliers) is potentially quite useful, but would need to be somewhat flexible to make up for some of its limitations. A similar system with, say, half the total number of correlators proposed above would be of use for some experiments but is really inadequate for the full VLA.

The 4x (64 channel, 3.125 MHz) system must be flexible in the polarizations, bandwidths and number of telescopes which it can handle while using its switching rate and all of its correlators to full advantage. Several examples of common configurations which might be used are given in Table 1. Operation with all four polarization channels (RR, LL, RL, LR) or with only the parallel hands (RR, LL) is an important requirement. Also operation using a subset of the full array with a resulting narrower channel bandwidth should also be possible. Such a mode would be very useful either in bad weather or for special combinations of U, V coverage and field of view.

Table 1 summarizes examples of some configurations which should be possible.

Configuration	Number of Telescopes	Polarizations	MHz Channel Bandwidth	# of Channels per Polarization	Total Bandwidth
1	27	RR, LL	6.25	16	100
2	27	RR	6.25	32	200
3	27	RR, LL	1.56	32	50
4	27	RR, LL, RL, LR	3.125	16	50
5	27	RR, LL, RL, LR	12.5	8	100
6	27	RR, LL, RL, LR	0.781	32	25
7	19	RR, LL	1.56	32	100
8	13	RR, LL, RL, LR	1.56	32	100

Of these eight possibilities configuration 2 is the least useful.

In summary, the multi-channel spectral line system should be of considerable importance to continuum work. However, the 4x (64 channel, 3.125 MHz) system is of minimal capacity to handle the continuum problems and must be designed with the continuum work in mind in order to be particularly useful.

D'Addario, L. R. 1975, VLA Scientific Memorandum 120

APPENDIX B

NATIONAL RADIO ASTRONOMY OBSERVATORY

SOCORRO, NEW MEXICO

November 10, 1975

VLA COMPUTER MEMO #127

SPECTRAL LINE MAP MAKING BY DIGITAL COMPUTERS

A DESIGN STUDY

B. Clark

I - INTRODUCTION

In the process of spectral line system design, I have made an independent study of the spectral line sort-merge problem, based on disks competing with the IBM 3350. Specifically, W. R. Burns suggested the consideration of the Calcomp 1035/235-4 disk, which stores 400 M bytes on a nonremovable spindle.

I have not attempted to optimize any procedure in any way nor have I searched for any sophisticated algorithms. I merely took reasonable approaches to every problem until I found one which worked without having a great deal of unused capacity. For this reason, this is not the usual conservative computer design study. The reader will note safety factors of the order of 30% instead of the 100% usual in real-time computer design. The system, bought and programmed as described here, would very probably not work. However, it is about right, and a more careful look should pick up about as much from optimization as is lost to items I have omitted.

I have evaluated three approaches:

1) the classical disk-based sort,
2) storing the spectrum at an address on disk so that it lies in the appropriate (u,v,w) cell, adding the spectrum to any previous one at the same (u,v,w) address (an approach suggested by L. D'Addario),
3) forming the (u,v,w) solid, in large part, in a large memory area in the DEC-10.

I have presumed we are trying to handle 256 frequency channels, each channel output (in u,v,w space) consisting of two 16 bit numbers. For simplicity, the information passed for every baseline every sample time is assumed to be exactly 1024 bytes. The not-very-useful end channels are dropped and replaced by the values of u,v,w, baseline, time, status information, etc. at an early stage in processing.

For each approach, I have looked into sample times of 10 seconds, 20 seconds, and 40 seconds. The maximum usable numerical fields of view (NFOV's) are about 2048, 1024, and 512 respectively. In the last case, where the chore is not too onerous, one might wish to use a 1024 field of view to better suppress aliased sidelobes of distant sources. One should also be aware that the outer parts of the field of view are corrupted by various instrumental effects due to gridding and aliasing, and, depending on the desired accuracy, the usable NFOV may be significantly less than the computational NFOV, which is all I consider here. If these fields of view are used at two points per beam at 21 cm wavelength, we can calculate the angular field of view in degrees.

NFOV \ Config.	A	B	C	D
2048	.57	1.20 (ant beam)	-	-
1024	.28	.92	-	-
512	.14	.46	1.20 (ant beam)	-
256	.07	.23	.72	1.20 (ant beam)

From this follows the required number of steps in the third dimension ≈ (numerical field of view)x(angular field of view)/4

NFOV \ Config.	A	B	C	D
2048	6	11	-	-
1024	2	5	-	-
512	1	2	3	-
256	1	1	1	2

Further, one may use the Hermitian property to cut the required number of w samples by a factor of about 2; more specifically, to (N+1)/2. As is shown below, handling an appreciable number of w-samples gets very expensive, and it is well to take a closer look at the requirement of 6 samples for the B array and 2048 NFOV. If the mapping is made to the first null of the antenna pattern, we can have, with 2048 NFOV, about 3 points per array beam. If we go back to 2, we are really using only about 1400 NFOV, and the number in the above table drops from 11 to 7, or to 4 samples, using the Hermitian property. This number is the one used in the discussions below. Specifically, I design, for the three sampling intervals, a system for the u,v,w dimensionalities given in the following table.

Sample Interval	Sample Data Rate Bytes/sec.	uvw Dimensionality	u,v,w Data Rate (12 hr obs) Bytes/sec.	x,y Data Rate (12 hr obs) Bytes/sec.
10s	36 K	2048 x 2048 x 4	398 K	50 K
20	18	1024 x 1024 x 3	75	13
40	9	512 x 512 x 2	13	3.2

It should be emphasized that, with proper connection of the special purpose hardware to a mass store, the choice of "pipeline" sample time does not preclude utilization of shorter sample times, at the price of not being able to reduce spectral line observations in real time.

At this point, also, the properties of the Calcomp 1035/235-4 disk can be summarized:

1600 cylinders/spindle
19 tracks/cylinder
400 M bytes/spindle
1200 K bytes/sec. peak transfer rate
789 K bytes/sec. long record mean transfer rate
4 ms single track seek
16 ms max latency
30 ms mean access time
55 ms max. access time

Cost $17.5 K per spindle.

Approx. $20 K controller and interface.

The Calcomp interface to the Varian minicomputers is already designed, so these computers have been costed. The compute times do not appear to be a limiting factor, so the high numbered members of the family are not required. On the other hand, memory capacities beyond 32 K words are required, so the V71 may not be used. The V72 is, therefore, used in these estimates. The 1.6 microsec per word maximum disk transfer rate appears to me too high to try to run into 1.2 microsecond core memory, so I have used 660 ns core in the cost estimates.

I have estimated no programming costs.

II - SYSTEM BLOCK DIAGRAMS AND COSTS

A. Ten Second Sampling

In all of the possible ten second sampling systems, much of the cost of the system lies in components occurring after the data is first given to the DEC-10. These components depend only on the map size, and so are in variant to the sort algorithm. The cost variation of the different algorithms is thus small compared to the total cost, but the minimum seems to be attained by the classical disk based sort, and a system based on it is described below.

Doing the sort really classically would mean making up little records with keys of frequency, u,v,w and contents of baseline, time as well as real and imaginary parts. This is a little too classical, as the overhead in storage required is about 300%. The approach I take here is to do a classical sort on records of 256 frequency channels, and, at the end, commutate the frequency channels into individual partitions. It is likely that the optimum procedure is somewhere between the two, to move some of the work of the final commutation pass into one of the other passes. I have not investigated this at all.

The technical details of the processes taking place in the various system processors are given in the technical appendix. The system block diagram is given in Figure 1 and the rough costing in Table I. A more detailed cost breakdown is in the technical appendix. It will suffice to give a brief summary here.

The system is a pipeline processor, which pumps the data through the various stations at a sufficiently high rate that no backlog is accumulated. The pipeline processor is also presumably associated with a mass store which ideally can be accessed from several points in the pipeline, so that, if something breaks down, the data flow may be diverted into a reservoir in the mass store without interfering with observation. When operating in the pipeline mode, the DEC-10 CPU and array processor are heavily subscribed, and other programs will slow down by a large factor. If twelve hours of data on a single source are run into the pipeline, the first map appears on the output display about 4½ hours after the end of the observation. After this, the maps appear about every two minutes.

1) The Sort Station. This device accepts, from the existing Modcomps, 351 records of length 1024 bytes every 10 seconds, and, after receiving about 8000 such records, puts them in u,v,w order and outputs this sorted string (about 8 million bytes) to the next station.

2) The Merge Station. This device accepts the 8 million byte sorted strings and combines them into a final output sorted string through four phases of four input merges.

3) The Commutation Station. This device converts the sorted string of 1.5 million records of 256 channels per record into 256 sorted strings with one channel per string. This would be an appropriate place to archive data.

4) DEC-10/Spectradata System. These existing components grid and perhaps apodize the u,v,w data, do the Fourier transform in the third dimension, and generally control the data flow from the sort system through to the display system.

5) Very Fast FFT. The Spectradata is not sufficiently fast to handle the Fourier transforms needed in this system. An external FFT, of which there are several on the market of sufficient speed (2048 complex vector transformed in 5 ms), is required.

6) Transpose Station. A two dimensional Fourier transform is most conveniently handled by conventional devices by a transform of the row vectors followed by a transpose of the half transformed matrix and a second transform of the row vectors. I suspect that the least costly device for performing this transpose is a large CCD shift-register memory (1.4×10^8 bits). However, a system constructed from existing half megabit cards is more expensive than the alternate system priced here of four minicomputers, each with a disk.

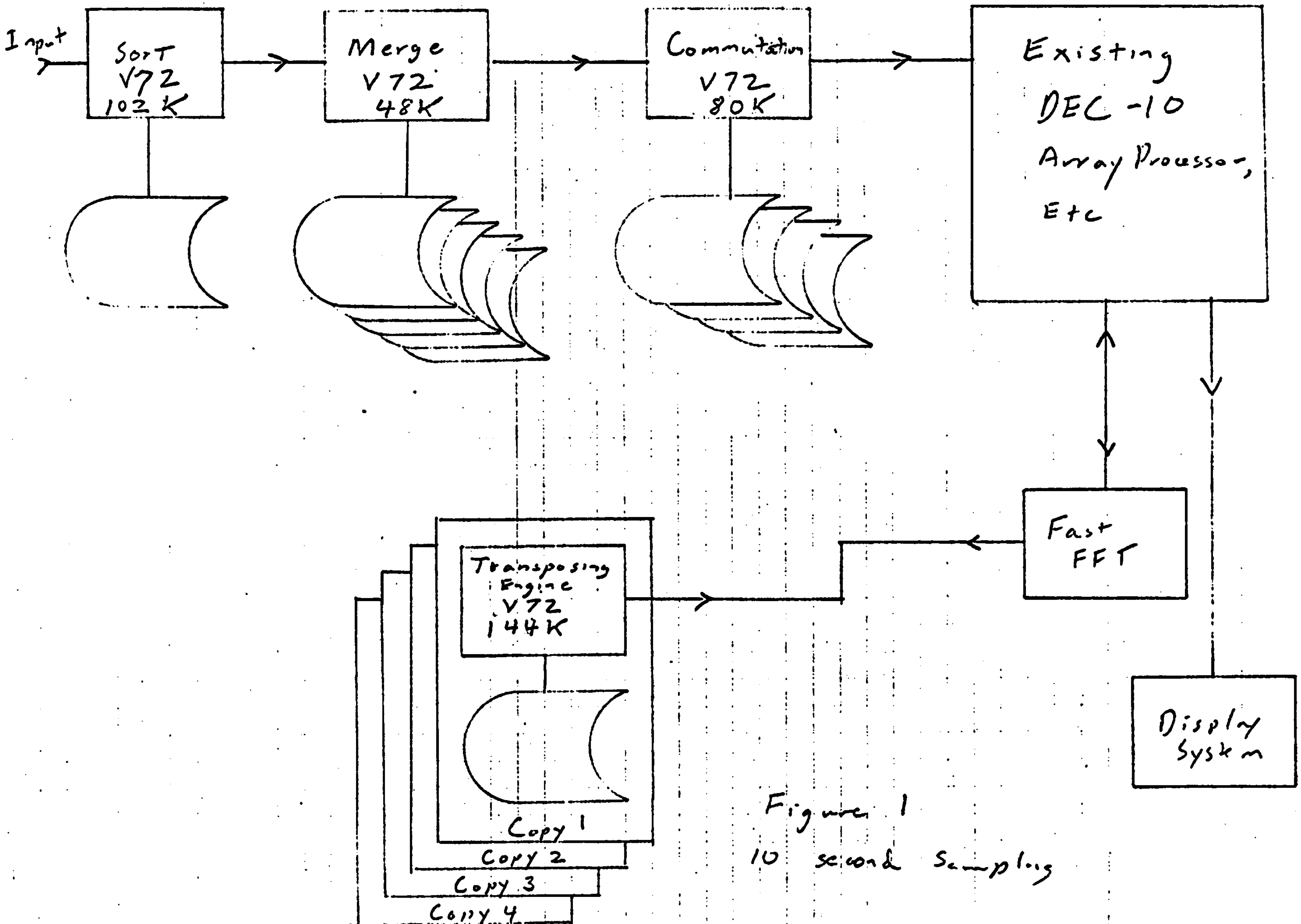


Figure 1
10 second Sampling

Table I
Cost of 10 Second Sampling Classical Sort System

Component	
Sort Station	$ 107 K
Merge Station	147
Commutation Station	144
Transpose Station	488
Fast FFT Station	150
Spares @ 15%	155
System Integration, Documentation, Purchased Software @ 10%	119
Total	$1,310 K

B. Twenty Second Sampling

As stated in the introduction, the map size appropriate for this sampling is 1024 x 1024 x 3. The most significant savings occur in the resulting smaller data handling problem in the DEC-10 and beyond, due to this smaller map size, rather than to the decreased data rate in the sort machine per se.

1) Description of a 20 Second Data System.

The most effective algorithm for the 20 second sort involves some degree of direct sorting into DEC-10 memory. That is, data is input to a minicomputer at the rate of 351 baselines every 20 seconds, and is then run into a radix sort - the sort used by a card sorter. The computer has time to do three passes of a four 'bin' radix sort. At the end of the three passes, there are 64 'bins', each containing, say, the data for all v's and w's and a range of 16 (out of 1024) in u. The records are then commutated (as in the 10 second system) to convert their organization from having a value from a given baseline/time contiguous, to having frequencies grouped together. Then, in the DEC-10, the 48 rows of the u,v,w (16 u's and 3 w's) solid are formed in place. The data can then be apodized by the Spectradata processor.

With the 20 second sampling, the map making process becomes just easy enough to do entirely in the DEC-10/Spectradata system, but both CPU's and the disk channel would be running at saturation, and time sharing would cease. This is intolerable. I think it is necessary to install an aide to both CPU's. This means an external FFT, and a single transposing machine of the sort discussed with 10 second sampling.

The costs of the system are given in Table II, and the block diagram in Figure 2.

Table II
Cost of 20 Second Data System

Sort/Commutate Station	$ 127 K
64 K Words Core for DEC-10	66
Transpose Station	122
Fast FFT Station	150
Spares @ 15%	70
System Integration, Documentation, Purchased Software at 10%	53
Total	$ 588 K

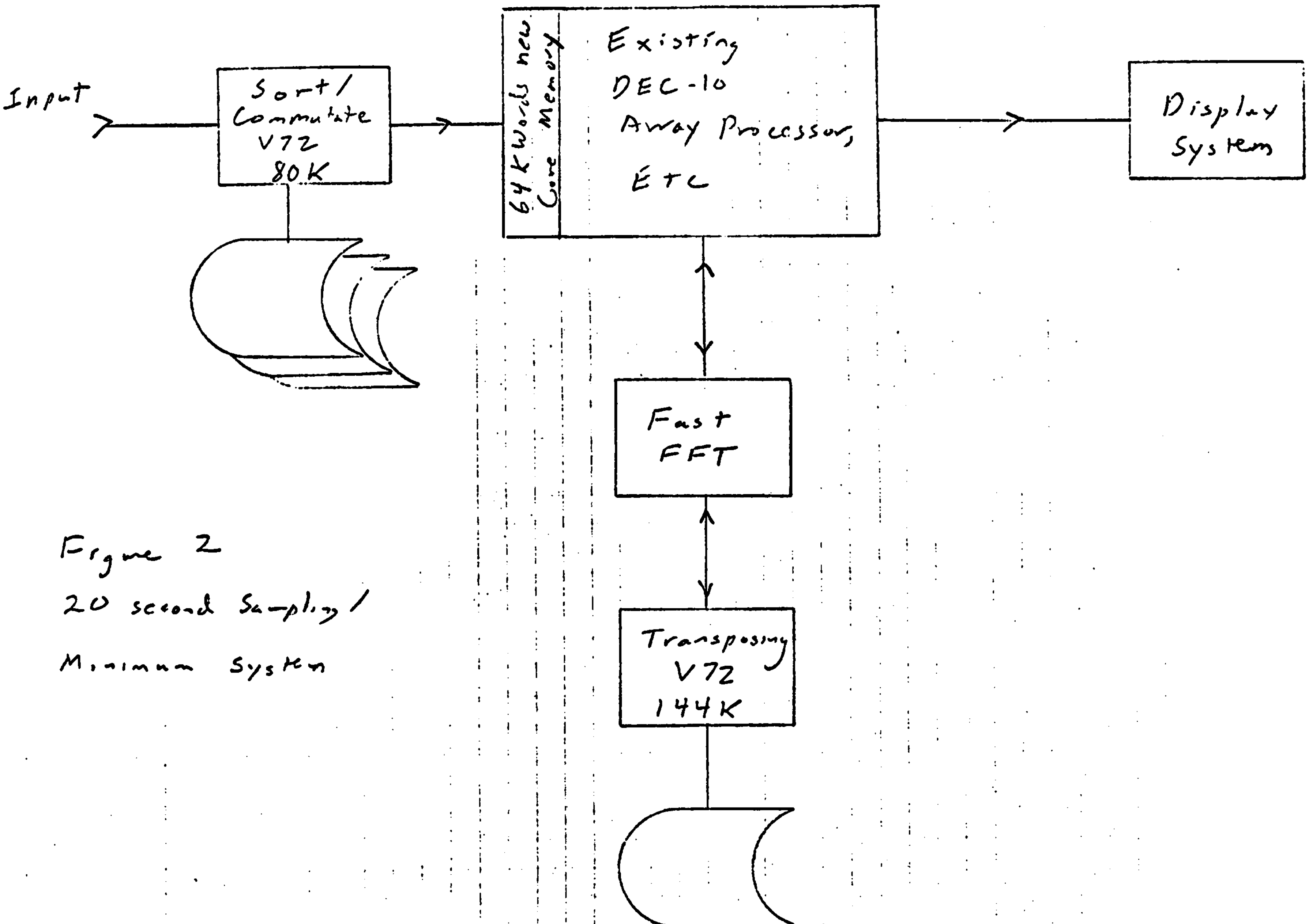


Figure 2
20 second Sampling/
Minimum System

2) Use of the System for Processing 10 Second Data.

The system described above can clearly be used to emulate the full ten second system at rates less than real time. The only minor hang up is that the three spindles of disk on the sort/merge/commutate station can hold only about 9 hours of 10 second data; this can presumably be resolved by appropriate use of the mass store. The rates data can be processed are nearly proportional to the number of CPU's involved. Therefore, this system will require about three times real time on the sort/merge/commutate computer and four times real time on the transpose computer to process full 10 second data.

3) An Enhanced 20 Second Data System.

If the four times real time is felt to be too much, an intermediate step is possible, which can process 10 second data in twice real time.

This machine differs from the one above by employing the solid state memory transposing engine rather than the minicomputer. Also, the sort/merge/commutate engine consists of two minicomputers, which may do the entire sort for 20 second data, eliminating the additional core for the DEC-10.

When used in the 10 second emulate mode, during real time the two minicomputers do the sort and merge phases, and output completely sorted but uncommutated data to the mass store. Then, during the second pass (also requiring about real time) the sort/merge engine is split apart, and the computer with three disk spindles is used for commutation, and the other is used in conjunction with the solid state memory for transposing. The solid state memory can be interfaced to provide sufficient throughput to transpose four 1024 x 1024 quadrants of the 2048 x 2048 plane in 30 seconds, and in a similar time the minicomputer can combine them to form the final transpose plane.

The cost of this system is shown in Table III, and the block diagram in Figure 3.

Table III
Cost of Enhanced 20 Second Data System

Sort/Merge Station	$ 107 K
Merge/Commutate Station	127
Solid State Transpose Station	299
Fast FFT	150
Spares @ 15%	102
System Integration, Documentation, Purchased Software @ 10%	79
Total	$ 864 K

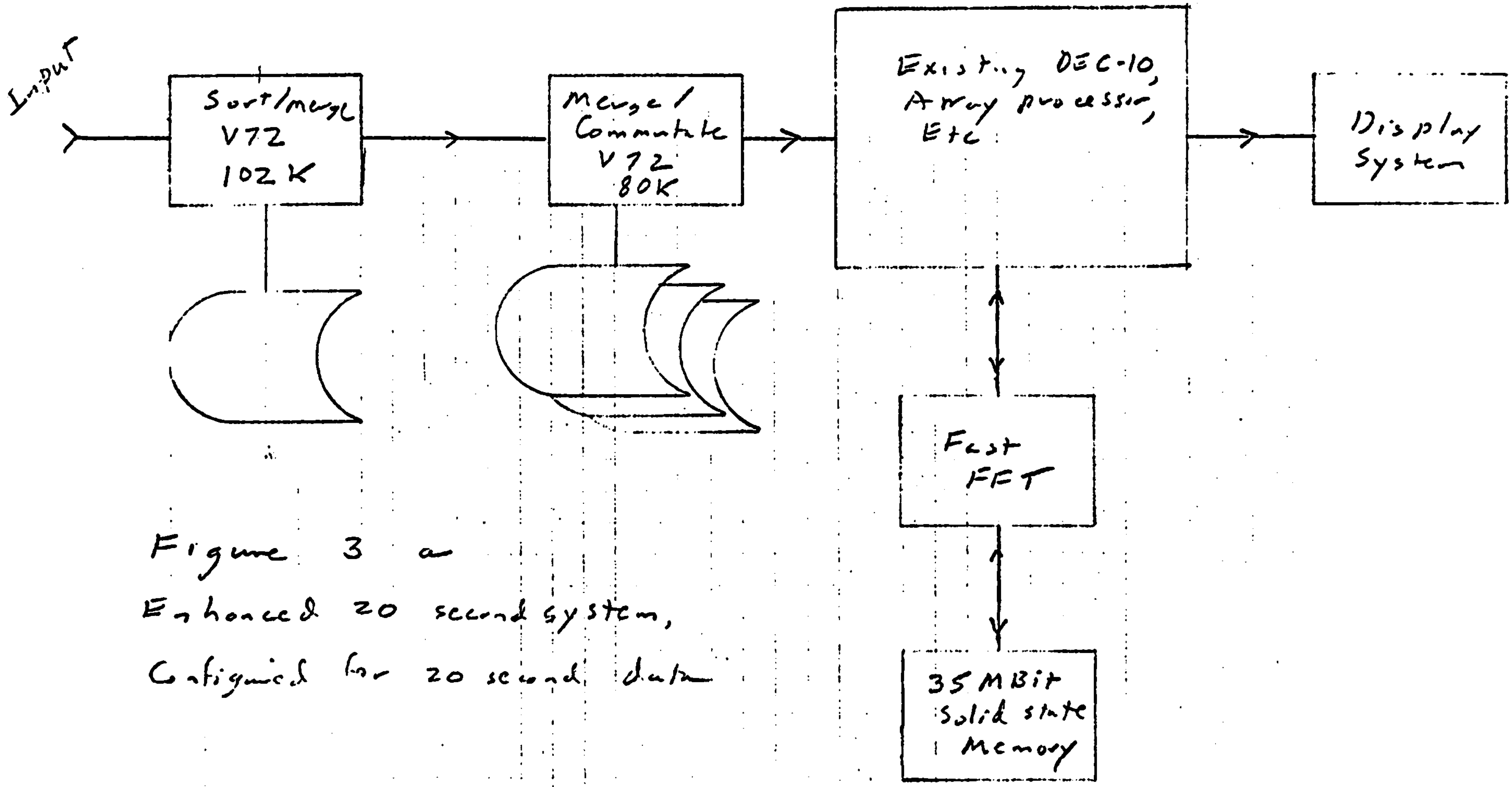


Figure 3 a

Enhanced 20 second system,

Configured for 20 second data

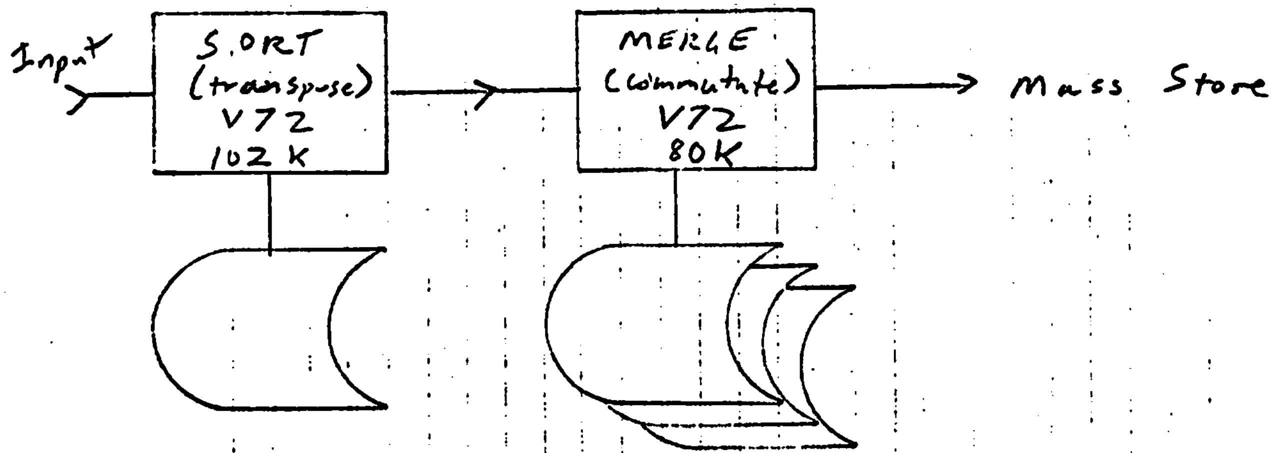


Figure 3b - Enhanced 20 second system during observing of 10second data.

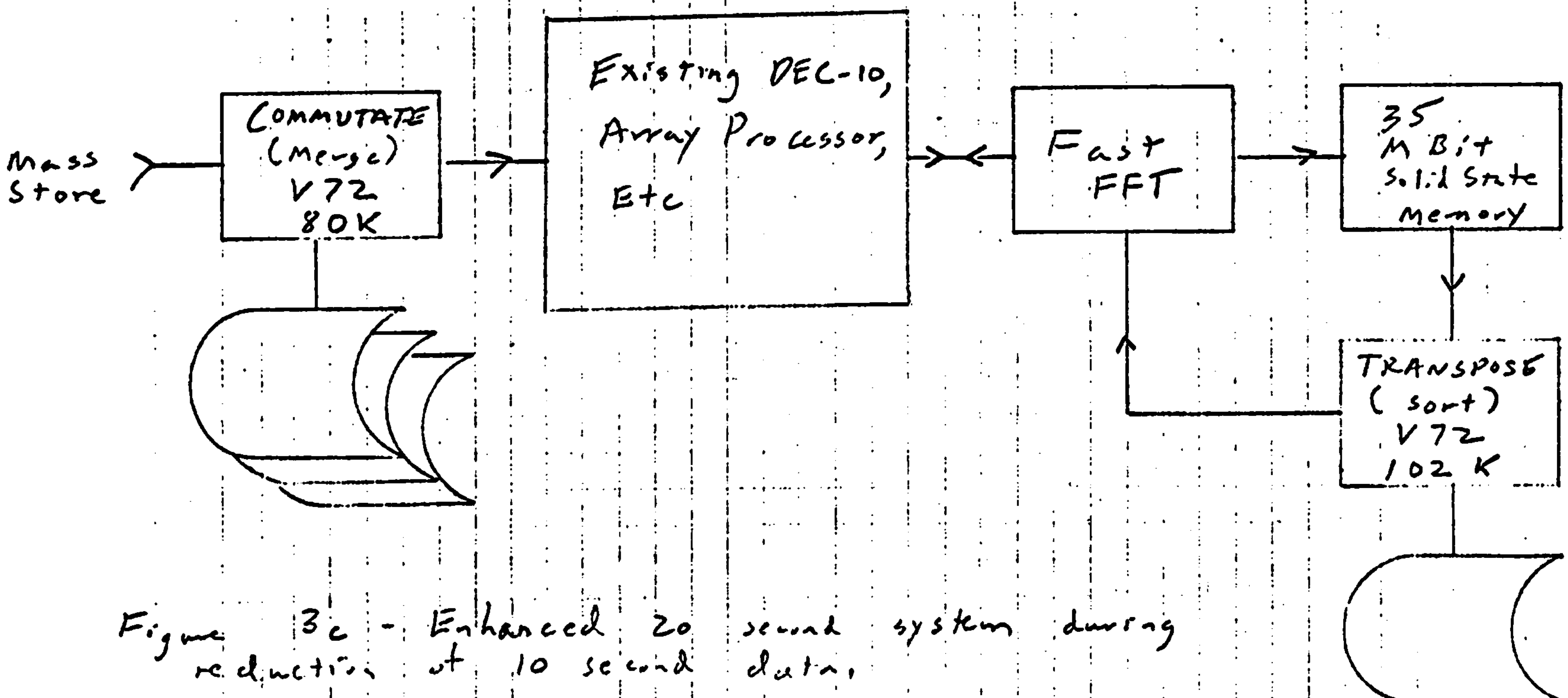


Figure 3c - Enhanced 20 second system during reduction of 10 second data.

C. Forty Second Sampling

With this sampling time we may go to a NFOV of 512 x 512 x 2. This is yet another reduction of a factor of 5 in the amount of map data it is necessary to process. The DEC-10/Spectradata now may process the entire mapping problem alone. A small assist from a sort/merge device is still required at the front end, and this comprises the entire addition to the existing system. The cost is $160 K, including spares.

However, when one of the faster sampling rates is wanted, the impact falls heavily, and most undesirably, on the DEC-10. The time taken by the Spectradata processor to do the FFT's alone is:

512 x 512 x 2	1.7 hrs.
1024 x 1024 x 3	10 hrs.
2048 x 2048 x 4	54 hrs.

There is likely to be a comparable impact on the DEC-10 CPU, even if all sorting and transposing are done by the single minicomputer.

With this system, processing 20 second data will take about three times real time, and 10 second data about seven times real time.

III COMMENTS AND OPINIONS

Although the 40 second sampling system is adequate (barely) to map the full beam of the C configuration at two points per beam, it impresses me that it has much too little margin for all the little variations of processing we might want to try. On the other hand, considering the small range of problems for which it is necessary, the full ten second system impresses me as overkill. I would, therefore, recommend either the twenty second system or enhanced twenty second system.

At the moment, the high price of the solid state memory makes the enhanced system look rather undesirable. However, the price I have estimated for the memory is based on integrated memory boards whose price is seven times the cost of the memory chips they contain. This cannot be a permanent situation. If the cost of the solid-state memory transposing unit falls from $300 K to $200 K, then the additional cost of the enhanced system over the 20 second system, $165 K, begins to look like a worthwhile investment.

The 40 second system is acceptable, I think, only if it is side-by-side with a good, convenient, and flexible optical processor which has output of sufficient quality that the desire to make digital maps at all is very significantly reduced. If a ten- or twenty-second system is constructed, the role of any optical processor would be essentially to provide a remarkably sophisticated display. The additional advantages of providing displays without the pipeline lag and of being able to inspect maps during observation partake of the same nature.

The costs in this memo are, of course, not perfectly accurate. It is my hope that the necessary items I have omitted would be counter-balanced by the items which could be saved by a more careful investigation and optimization of the systems discussed.

IV THINGS THAT OUGHT TO BE LOOKED INTO

Before a final system is designed, there are several areas which ought to be considered further.

A. I have assumed that the 3-dimensional transform is the computationally easiest method of taking account of sky curvature. The mosaic method should be looked at, wherein small NFOV maps are generated at various phase tracking centers. This method involves much more computation, but the transpose problem (the most severe problem for much of the above system design) is much less severe.

B. Several constants are rather uncertain from theoretical discussion. If general agreement is not reached, simulation studies must be undertaken to discover them. Most important of these are the number of points per beam required, the percentage of calculated map which is 'usable', the upper limit for NFOV*integration time, and the number of w-slices for given numerical and angular fields of view.

C. For the various sample times, I have taken the worst case computationally, Configuration C for 40 second data and B otherwise. Clearly, most of the spectral line observations will not be made at these worst-case configurations. The value of having a 'ten second machine' or a 'twenty second machine' must be looked at very carefully in terms of how often we would use the facility, and the slowdown (beyond real time) when we do not provide the full machine.

D. I did not realize until preparing this report for final typing that I have egregiously misconfigured things for the use of solid-state memory in the transposing system. In this case the Hermitian conjugate property should be invoked in the u or v direction rather than in w. For instance, in the ten second use, instead of transposing four 2048 x 2048 matrices, one should transpose eight 2048 x 1024 matrices. The throughput remains the same, but the amount of memory required at any one time is halved. The impact on capacity calculations elsewhere in the system is, I think, minimal.

V TECHNICAL APPENDICES

Those of you without a considerable interest in computer programming may stop reading at this point. You are to be congratulated for getting this far.

A. Cost of Varian 72. The prices used here are taken from the Datapro Report of April, 1975. In my tables, I have lumped together the following basic items as a single line:

72-1201	V72 CPU + 8K of 660 ns parity core	\$12.5K
72-3100	3 block transfer channels	4.5
72-3300	Memory Map	2.5
70-8301	2 digital interfaces	1.0
70-6402	Console CRT	3.2
	V 72 System	\$23.7K

It is assumed that the 8K of core in these configurations is just sufficient to hold a minimal operating system and program and that buffer space is a separate line in the cost tables.

I assume that communication with all other parts of the system are via the digital interfaces. Communication with the mass store is also probably by such an interface.

Presumably there would be some mechanism for, say, the Boss Modcomp to force a load, and a bootstrap loader for the disk (which, incidentally, I have not priced).

B. Full Ten Second Sampling

1) Classical Disk-based Sort: The system I have considered consists of the three minicomputers whose functions are indicated in Section II A. A more detailed discussion of just what is happening in each machine is given here. A similar discussion for the transpose machines is given in Section V B.4.

a. The Sort Station.

The sort station accepts 351 1024 byte records every 10 seconds and performs a first stage sort on the way to disk. As the data goes by, the sort key (6 bytes) and disk address (2 bytes) are preserved in core. Eight thousand pointers are accumulated (representing 8 million words of data) and sorted by a partitioning sort. The data are accessed from disk and output in sorted, 8 million byte strings (corresponding to 220 seconds of observing time) to the merge station.

i) Input Stage.

This stage requires a 72 K byte buffer divided into eight areas holding nine records (the 9th record area is an overflow area used to hold a new input during writing of the previous eight). The incoming record is directed to one of the eight areas, depending on its key (with equal probability over the ensemble of VLA observations). When the area has eight records, they are written to disk. The disk time required to write the eight records is 71 ms (two accesses are figured, since the heads must be returned to the operation in progress). A 71 ms access time every eight records corresponds to an accumulated time of 3.6 hours for the data taken in twelve hours.

ii) Partitioning Sort Stage.

Eight thousand pointers of eight bytes each stored in a 64 K byte buffer. A partitioning sort of eight thousand items typically requires 90 thousand comparisons and 50 thousand exchanges. A comparison on the V 72 computer typically requires about 200 memory cycles (including the overhead of getting to unit item partitions) and an exchange 25. The partitioned sort will run in about 14 seconds. This corresponds to a total time of 0.7 hours in a twelve hour observation.

iii) String Assembly and Output Stage.

This stage is allocated a 64 K byte buffer, requiring 64 more or less randomly located records to fill it. Because of the input strategy, however, these records will lie in only one-fourth of the whole 8 million word area - on only 12 cylinders of the disk. In a typical case, most of the information will fall on five cylinders of the twelve. Making a 30% allowance for CPU overhead and track switching, about ten records per track can be written, so, in a given record slot and cylinder, there is an everage of 1.6 records to be recovered. The mean worst case among the ten record slots has four records in the slot, so the disk must rotate through four turns to pick up all the data. The dwell time on the other seven cylinders is assumed to average 1½ turns. The total disk time is thus

Initial access		.030^s
Cylinder access	.008	
Dwell 4.6 turns	.077	
Times 5		.425
Initial access to second area		.020
Cylinder access	.008	
Dwell 1.5 turns	.025	
Times 7		.231
		.706^s

The whole 8 million byte record goes in 91 seconds, representing 4.8 hours for a twelve hour observation. The time of the sort station is thus spent as follows for a twelve hour observation:

Input state	3.6 hours
Partitioning sort	0.7 hours
Output stage	4.8 hours
	9.1 hours

Note that, although the data storage requirement is small, a 3330 type disk is still required because the high track density and fast cylinder access are required.

Note also that the total buffer memory requirements for this station are about 200 K bytes.

b. Merge Station.

This station accepts the presorted strings of length 8 million bytes and goes into four phases of four input merges. Each phase consists of reading 16 K byte records from each of four input areas and writing a 64 K byte record on the output area. In order to have a completed output record and four partially used input areas, a total buffer area of 128 K bytes is needed. For each phase, the time economy is as follows:

Access time	$.030^s$	
16 K byte input time	.021	
	.051	
Times four		$.204^s$
Access for output		.030
64 K byte output time		.083
		$.317^s$
for 64 K bytes		
or, for 12 hours of data,		2.02 hours
Total for four merge phases		8.1 hours
Input from sort station, output to commutation station		2.0 hours
		10.1 hours for a 12 hour observation

c. Commutation Station.

When the sort is complete, the disk is read into half of the 128 K byte buffer (one-fourth cylinder), commutated into the other half, so that it is organized into 256 records of 256 bytes each, and rewritten on the same cylinder.

The time budget is as follows for each cylinder:

Initial seek		.006 s
Rotational latency	.016	
Input 64 K bytes	.083	
Commutate	.130	
Rotational latency	.010	
Output 64 K bytes	.083	
	.322 x 4 =	1.288
		1.29 s/cylinder
		0.57 hrs/spindle
		2.28 hrs total

Commutation time is figured on the basis that each two bytes take a load, store, and miscellaneous overhead, each of two 660 ns memory cycles.

The records can be stored so that the four records pertaining to one frequency channel are reasonably adjacent in rotation. There are 1024 bytes per channel on each cylinder, occupying 0.079 of a turn; suitable spacing between records should occupy about .05 turn. An adjacent cylinder seek takes about 0.3 turn. Therefore, if adjacent cylinders are oriented at 180° (i.e., if channel 1 is at 0° on even cylinders, it is at 180° in odd cylinders), a pack can be searched at the rate of 120 cylinders per second (assuming the accepting device is always ready). The total readout time is thus 13-1/3 seconds per pack per channel, about one minute per channel total, four hours to output the entire half day's data base. The input time to the commutation station is about another hour.

The total time required by the commutation station is thus 7½ hours for a twelve hour observation.

d. Cost estimate is given in Table IV below, from which the numbers in Table I are derived.

Table IV
Cost Estimate, Classical Sorting Scheme, 10 Second Data

Sort Station	
V 72 System	$ 24 K
96 K Words Memory	45
Disk Controller	20
I Spindle	18
	$107 K
Merge Station	
V 72 System	$ 24 K
32 K Words Memory	15
Disk Controller	20
5 Spindles	88
	$147 K
Commutation Station	
V 72 System	$ 24 K
64 K Words Memory	30
Disk Controller	20
4 Spindles	70
	$144 K

2) Store in place in u,v Plane: This sorting system replaces the sort station and merge station by more severe data storage and commutation problems. Holding 256 channels of 2048 x 2048 x 4 u,v,w solids requires 1.7×10^{10} bytes equals 43 spindles. Since the cost of this alone is greater than the classical sort system by a factor of nearly two, it was not considered further. It is possible that it might be more competitive if it were implemented with partial sorting and disk staging onto a mass store, but this line was not pursued.

3) Direct Gridding: In this procedure the commutation stage is very much the same as in the classical sort, but the sort stage is done entirely by indexing in the DEC-10.

In order to minimize cost, only one-eighth of a map need be present in the DEC-10 at one time - this is equivalent to a 32 way radix sort occurring in the commutation device, which does not degrade its performance intolerably, although the safety factor allowed is uncomfortably small (twelve hours data handled in ten hours).

The maximum memory must be installed on the commutation engine in order to minimize the DEC-10 memory required. With buffers approaching 512 K bytes, the commutation described in the classical sort processor can be implemented and the four w-slices may be routed to different disks as well. Then, when data is read sequentially, data is sorted, crudely, into eight v ranges. Only the first is sent to the DEC-10. The other seven are written out to disk again as long records of 64 K bytes. These long records can then be reread and forwarded very easily.

The time budget is, fcr a 12 hour observation,

Input Data	1.7 hrs
Commutation	2.3
Search and Read	4.0
Write and reread unwanted part of (u,v) plane	1.9
	9.9 hrs

Six disks are specified - four are the commutation area, one is reserved for the rewrite process, and the sixth accumulates the first part of the next observation while the commutation of the previous one is occurring.

The DEC-10 must be able to hold one-eighth of a u,v map - half a million words.

The cost of this approach is given in Table V below. This is to be compared with the cost for the entire sorting system of $398 K (sum of the entries from Table IV).

Table V
Direct Gridding to Core Cost, 10 Second Data

Commutation Station		
V 72 System	$ 24 K	
24 K Memory in CPU Chassis	21	
224 K Words External Memory	105	
Disk Controller	20	
6 Spindles	105	
		$275 K
Half million words, DEC-10 Memory		264
		$519 K

4) Map Making: The map making process consists of the following steps: a) grid, average and taper; b) Fourier transform rows; c) transpose x and v directions of the half transformed map; d) Fourier transform the rows of the transposed matrix (i.e., column transform the matrix); e) access the similar rows of the w slices and do the DFT (not FFT) in the third dimension, and store the resulting map for display and analysis.

a. Grid, Average, and Taper.

This step is done by the DEC-10-Spectradata system. The exact division of labor is not clear.

The complex visibility is read from the sorting system, along with the control channel giving the u,v location. The u coordinate is used to index the location within the row of the matrix where the data is to be added. If uniform weighting (independent of integration time within the cell) is desired, a count of data could also be generated. I presume the Spectradata could then be persuaded to divide the two rows. In any event, the Spectradata can easily superpose a Gaussian taper by multiplying by a constant (representing $\exp - \beta v^2$) and by an invariant vector ($\exp - \beta u^2$).

The order of magnitude times involved are:

Input data, (u,v)	3 μs/word
Index data by u and add	10
Count data entries	5
Looping overhead	8
	26 μs/word

Total 2.7 hrs DEC-10 Time

The effort required to divide by the data count and apodize depends on details of the Spectradata machine which I do not have ready to hand. It is possible that it can do the job. I doubt it. However, it might well lie within the capability of the external fast FFT, or, failing that, not too much effort to implement as a separate external special purpose device.

b. Fourier Transform Rows.

A total of 4096 Fourier transforms of length 2048 must be done for each w-slice, or 4 million altogether. The FFT device must have an operation time (total time = setup time + (length)*$\log_2$ (length)*operation time) of 0.3 microseconds to do this in eight hours, not counting setup time and I/O time.

c. Transpose Rows and Columns.

To accomplish the transpose in real time is a remarkably difficult task. To illustrate, let us look at the throughput. To make 256 maps, each with four w-slices and each slice 2048 x 2048, in, say, eight hours, the throughput rate is 149 thousand values per second, or (with 16 bit word length), 4.8 million bits per second. This is about 0.75 times the sustained writing rate of a 3350 disk, so the data cannot be written to a single disk and recovered in real time.

There are two approaches to implementation of a transposing engine. We could have a large (1.4×10^8 bits) CCD shift register memory, at a cost of nearly a million dollars. Alternatively, we may simply set several mini-based disk systems down side by side, and round-robin schedule to get the desired throughput.

For instance, with four minicomputers, each with a 3330 type disk, we can solve the problem as follows: Using a 256 K byte buffer in each machine, and with each machine receiving a quarter of the data, the buffer will hold 128 rows. After transpose, the buffer is written sequentially to disk. When the data is to be output, the buffer can be refilled with 32 columns of length 2048 by 16 accesses. The time adds up as follows:

Initial input and transpose	$3^{s}.0$
Disk write	$6^{s}.6$
Disk read	$14^{s}.3$
Output	2.0
	26^{s}
For 4 x 256 maps	$7^{h}.4$

The cost of a single minicomputer system is given in Table VI below - the cost of the entire transpose station is four times as great, or $488 K.

Table VI
Cost of a Transpose Engine

V 72 System	$ 24 K
128 K Words Memory	60
Disk Controller	20
1 Spindle	18
	$122 K

d. Fourier Transform Columns.

This step was discussed under c.

e. Fourier Transform in the Third Dimension.

With reasonable buffer sizes, the disk access times for the four w-slices can be made decently small and I shall ignore it. The function to be performed is

$$B(x,y) = \mathrm{Re}\{\sum_{n=0}^{3} e^{inz} B(x,y;n\Delta w)\}$$

where $z = \sqrt{1 - x^2 - y^2}\ \omega\Delta w \simeq \frac{1}{2}(x^2+y^2)\ \omega\Delta w$.

The transform is done by the array processor a column at a time, multiplying first by the constant $\exp \frac{1}{2}\, i\, x^2\, \omega\Delta w$ and then by the tabulated vector $\exp \frac{1}{2}\, i\, y^2\, \omega\Delta w$. Finally, the four contributions are summed and the real part taken. Presuming that the array processor takes 3 microseconds for a complex multiply, this consumes 24 seconds per w-slice, 96 seconds per map or 6.9 hours total. If, in fact, the Spectra-data processor requires a significantly longer time, which seems likely, this function must also be farmed out to an external device.

C. Twenty Second Sampling

The data rates in the sort section are lower by a factor of two, resulting in, approximately, a factor of two less compute being needed. The data rates in the map making section are a factor of five lower, which makes the more significant saving in cost.

1) Classical Sort Approach: It is not quite possible to combine all three stations of the 10 second system into a single device. A two station pipeline is still required.

To sufficient accuracy, times can be scaled from the discussion of the 10 second system, with the following results:

Initial Sort Stage	4.4 hours
Merge Phases @ 1.1 hours	4.4
First Commutation Phase	1.7
Second Commutation Phase	2.0

A reasonable split of the effort would be to assign the initial sort and first merge to the first station, and three stages of merge and commutation to the second. The second system needs enough storage for 16 hours observation - 12 hours being commutated and 4 hours to store incoming data during commutation. This is three spindles. It is possible but uneconomic in the long room to replace the 3330 type disk on the input station by a 2314 type. The cost is shown in Table VII below:

Table VII
Cost of Classical Sort System for 20 Second Data

Sort/Merge Station	
V 72 System	$ 24 K
96 K Words Memory	45
Disk Controller	20
1 Spindle	18
	$107 K
Merge Commutate Station	
V 72 System	$ 24 K
64 K Words Memory	30
Disk Controller	20
3 Spindles	53
	$127 K

2) Direct Gridding to Disk: The data storage required to support u,v,w solids 1024 x 1024 x 3 amounts to nine spindles. About 80% of this space is empty. With 128 K bytes of buffer commutation (first phase) can be done (at the rate of 0.6 hrs/spindle) in 5½ hours. During this transpose, the data can be compressed, so that the output phase can be done in the same 2 two hours as in the sort case.

Inputting the data will cause about one hour interference with this process, because it will usually be done on separate access arms, and it will only interface with the commutation process during the times when the record to be written falls at the rotational position on disk where the commutation is currently working.

The data rate of the machine in this mode - one 1 K byte record every 57 ms - is sufficiently slow to permit a mean access, read, add, rewrite operation (47 ms), in a straightforward way.

The cost of the system is shown in Table VIII below.

Table VIII
Cost of Direct Gridding to Disk, 20 Second Data

V 72 System	$ 24 K
128 K Words Memory	60
Disk Controller	20
9 Spindles	175
	$279 K

3) Direct Gridding to Core: The commutation phases require a total of 3.7 hours for a twelve hour observation. There remains time in the same engine to do three 4-way radix sorts. This results in 64 strings, each containing 16 rows by three w-slices of the 1024 x 1024 x 3 u,v,w solid. This last sort, into 48 individual rows, requires very little additional effort on the part of the DEC-10, but about 48 K additional memory.

Table IX
Cost for Final Gridding to DEC-10 Core, 20 Second Data

Sort/Commutate Station	
V 72 System	$ 24 K
64 K Memory	30
Disk Controller	20
3 Spindles	53
	$127 K
64 K Memory for DEC-10	$ 66 K

4) Map Making: The map making process becomes just easy enough to do entirely in the DEC-10/Spectradata system, with both CPU's and the disk channel running near saturation.

I think it is sufficiently undesirable to have either processor saturated that I would include both an external FFT and a transposing station. The transposing station has less work to do than in the 10 second case by about a factor of (2048 x 2048 x 4)/(1024 x 1024 x 3) = 5.3; a single stage of the 10 second minicomputer will handle it nicely, for a cost of $122 K (Table VI).

5) Handling 10 Second Data: It is obviously desirable to be able to handle 10 second sampling times at less than real time rate. A small price premium would attach to being able to do this in the most convenient possible way.

a. Direct Grid to Core System.

The limiting time factor in this case is the transposing station. Working with the full 2048 x 2048 x 4 maps and twelve hour observations, it can transpose at only about 1/4 of real time. At this rate, there is no reason why the sort system should not be used to successively emulate the three stations of the ten second pipeline, sending data to the mass store each time instead of to the next station. This minimizes the heavy load on the DEC-10 and results in a total rate of four times real time for the processing.

b. An Enhanced System.

Choosing the more expensive option in a couple of places in the design considerably enhances its capability of handling 10 second sampling.

i) Sort rather than Direct Grid.

If the sort approach (Section IV C.1) is taken, and a fourth spindle is added to the second station, the sort pipeline can emulate in real time the first two stations of the 10 second pipeline. Then, a second pass through the second station may emulate the third station of the 10 second pipeline. Thus, in twice real time (with a spare CPU the second time through) the 10 second data may be sorted and commutated.

ii) Solid-State Memory Transposer.

A 1024 x 1024 map w-slice, with 16 bit values in real and imaginary parts consists of 32 M bits of information. If we provide a solid state memory of this size, we can write a row at a time and read back a column at a time.

I estimate a price for the solid-state transposer in Table X on the basis of a half M bit RAM board (mounting and driving Intel 2416 CCD IC's). This board sells for $3500.

Table X
Solid State 1024 x 1024 Transposer

64 ½ M Bit Memory Boards	$224 K
Controller, Interface	50
Power Supplies, Packaging	10
Engineering	15
	$299 K

If one uses this concept for transposing a 2048 x 2048 matrix, one must record three quarters of the data and transpose a quarter of the data at a time. That is, of each 2048 value row, 512 entries would be inserted into the 1024 column positions in the solid state memory row (the first value, say, going alternately into columns 0 and 1). Then, when the memory is full, it would be read two columns at a time to extract the 2048 values of a single column of the large matrix. The time budget for a single w-slice is about as follows:

Write 3 quarter rows with 64 K bytes/record	22^s
Read back with 192 K bytes/record	16^s
CCD memory dump time	$\sim 6^s$
	44^s

Processing for four w-slices, 256 channels thus requires 12.5 hours, only slightly more than real time.

D. Forty Second Sampling

The data rate here is sufficiently low that the process could be handled entirely within the DEC-10. However, this would have an extremely severe and probably unacceptable impact on the DEC-10 resource availability.

1) Sort by Minicomputer.

Given a minicomputer front end, either of the two complete sort algorithms (classical sort, direct grid to disk) operate with pleasantly large amounts of surplus time. For a 12 hour observation processed by the classical sort:

Initial Sort	2.2 hours
Merge Phases	2.2
First Commutation Phase	0.9
Second Commutation Phase	1.0
	6.3 hours

or processed by direct gridding to disk:

Access, read, add, rewrite, return access arm for each record	5.0 hours
Commutate	1.9 hours
	6.9 hours.

2) Map Making

Making a 512 x 512 x 2 map is not an extremely painful job for the DEC-10, without any minicomputer help whatsoever. The time is about as follows for a 12 hour observation:

Gridding	0.8 hours
Apodize (Spectradata)	1.5
Fourier Transform (Spectradata)	1.6
Transpose (DEC-10	0.4
(Disk Channel	1.0

3) Use of the Equipment for 20 Second Data.

With the use of 48 K core from the DEC-10, the system can keep up with the sorting and gridding chores, or without the extra core, can run slightly slower than real time. The Spectradata will be the limiting item, requiring about one times real time for apodizing and one for Fourier transforms. The impact on the DEC-10 CPU (not counting memory interference from the Spectradata) is not too bad, about 0.4 real time and the interference with the disk channel is also moderate, about 0.6 real time.

4) Use of the Equipment for 10 Second Data.

The Fourier transforms for a complete set of maps amounts to 54 hours, five times real time. There is another time increment, about twice real time, for apodizing and the third dimension transform.

Since the Spectradata operates at one-seventh of real time anyway, it would seem logical to utilize the single minicomputer CPU configuration to simulate, in turn, the seven stations of the 10 second pipeline discussed above. Thus, without excessive load on the DEC-10 (except memory interference with the Spectradata), the 10 second data may be reduced in about seven times real time. The first map does not emerge from the pipeline until after the sort is complete, about three times real time.