COMPUTER PROCESSING
TECH NOTES & REPORTS
AT/25.1.1/010

OFF-LINE Computing Requirements for the AT

Report of the "data processing meeting", 28-May-84
(part 2)

(mjk, jrf, rpn, ptr, rhw).

summary:

An estimate has been made of the cpu required to perform various processing tasks - an exercise analogous to the 2 VLA reports. We find that a vax 780 with an AP (running at a fair clip) will suffice for a reasonable mix of continuum and modest-field line observations. Some observation programs will be able, however, to overwhelm the machine. Specifically, wide-field observations at high spectral resolution. (eg. 1kx1k maps, 1k spectral channels.

Some possible observing programs are examined below:

A. Continuum

project	: wide field mapping (1kx1k)
days/field	: 4
channels/map	: 16
Stokes/field	: 8
visibilities/12 hour	: 8.3E6
equivalent FFTs (1Kx1K)	: 115
CPU (VAX 780 + AP)	: 1.9 hour/day

Substantial cleaning and processing of 8 maps will require about 8 hrs. total CPU, or 4 hrs/frequency band, with 4 stokes parameters.

B. Line

project	Galactic centre	LMC,SMC, galaxies	masers	molecular clouds	recomb. lines	dark clouds
days/field	8	8	1	4	4	4
freq. bands	2	2	1	2	2	1
channels	512	256	1024	256	128	256
Stokes	2	2	8	1	1	1
Nvis/12 hr	66E6	33E6	270E6			
map size	1024	1024	1024	512	512	256
resolution	1 km/s	2	0.1	.25	1	.1
Clean ?	yes	yes	25%	yes	75%	no
Self-cal ?	yes	yes	yes	yes	yes	no
processing time		10 mins/map			3 mins/map	

In more detail:

A. Continuum

A.1 Wide field mapping.

A probable average choice:

2 frequencies 4 poln. 16 channels 15 baselines 10 sec. integration.
(Full bandwidth at L band requires 32 channels to avoid bandwidth smearing, (80 MHz), but the correlator then allows only 4 products, 1 frequency band; this option would provide the same data rate as examined here, but with fewer maps.)

A.1.1 Data rate.

192 complex visibilities/second, or

8.3E6 complex visibilities/12 hour.

With 4 days/map, the total data set amounts to:

33E6 complex visibilities.

(For reference, 2400 ft. of tape at 6250 bpi holds about 180E6 bytes - adequate for the full data set at 4 bytes/visibility).

A.1.2. Gridding and mapping.

Assuming 1Kx1K maps, and a 6x6 gridding cell.

We need to read the entire dataset, perform 36 calculations, and do an FFT -

36x4.1E6 ops/poln

At 1 Mflop/sec (Vax780 + AP) : 2.5 mins/map . . total: 20 mins.

(Equivalent 1Kx1K FFTs : 20)

A.1.3. Clean.

a. We don't expect to clean all 4 polns. - say 3 as pessimistic average.

b. We must expect several cleans - say 2 as modestly optimistic average.

c. These numbers represent an average over expert analysts (who would use optimum gain and component number), and the cautious less expert (who may require more cpu).

5 major cycles - 10 FFT

1000 components 20 FFT . . 30 FFT/map

(In each minor cycle, scan of order 1/4 map; thus 250 full map scans: number of operations: scale/shift/subtract/store - say 10, giving 20 FFT equivalents).

Total, 6 maps, 2 cleans: 6 hours (360 FFT)

A.1.4. Self-cal.

Self-cal on I only; say 300 components. (Self-cal with the AT may well be computationally more expensive than it is for the VLA, as each day's observations place far fewer constraints on the model).

The load:

Nops = Nvis*(12 arith + 4 trig + (4 arith + 2 trig)*Ncomp)

with each trig calculation ~ 5 arith. we get:

Nops ~ 1E6*(32+14*300) = 1E9 ops/map . . 20 mins/map,

Total, 2 maps: 0.7 hour, (80 FFT)

Thus for each field, we have:

460 FFT, or about 7.7 hour CPU.

Since the observation was spaced over 4 days, the 12 hr. load becomes:

<u>115 FFT</u>, and <u>1.9 hr. CPU</u>

For comparison, Ekers et al (VLA Sci. Mem 150 quote

591 FFTs for a 6 sub-field (mosaic) wide field mapping.

Equally, scaling the VLA canonical task (256 maps, 2Kx2K) to 8 maps, 1Kx1K we get:

T ~ 3.4 hrs. (Note, 1 clean, and no self-cal).

A.2 Snapshots.

These are the continuum tasks which load up the VLA.

I don't see much demand for these at the AT. These continuum maps will, of necessity, be modest in size, and therefore should not represent a serious load.

A.3 Astrometry.

These would add little to the mapping load.

A.4 Pulsars

These could add significantly to the data rate, but not to the mapping load.

B. Line observations.

We assume 1 deep clean/self-cal on the continuum, or a perhaps a strong line (maser ?) source. This would be followed by the production run on the line maps.

thus: .5 to 1 hour as set up;

thereafter 8 to 16 minutes/map. (1k x 1k)

Ie. 45 to 90 maps/12 hour. (with no time for analysis.)

This analysis bypasses the I/O problem: with large data cubes, sorting the data may be a serious problem - if a video disk is not available.

Alternative timing estimates:

a. AAO vax: 256x256 maps took an average of 7.5 mins. to clean. (1000 components, 3 major cycles). Scaling up by a factor of 16x1.1 to our canonical task gives 2.2 hours. Thus we would seem to imply a factor of 8 improvement over the AAO vax by assuming an AP.

b. NRAO data, (from the AIPS HELP files) :

T = 25 minutes/map.

c. Simply summing the operations required:

N about 600 Mflop.

Thus at 5 Mflops, need 2 minutes;
1 Mflops, need 10 minutes.

All this implies that multi-channel wide field mapping will be difficult if a large number of channels is required.

Caveats:

1. Current figures are based on current hardware. making better use of modern APs could conceivably bring us closer to the 2 mins/map mark. (350 maps/12 hr.)

2. An observing mix of continuum/line in the proportions 1:1 would release a further 8 (?) hours cpu - ie. up to 600 maps/ "12 hr".

3. More modest map sizes would help.

Conclusions:

On present form, we would not be able to process more than about 100 widefield/spectral maps in 12 hours.