NATIONAL RADIO ASTRONOMY OBSERVATORY
SOCORRO, NEW MEXICO
VERY LARGE ARRAY PROGRAM

VLA COMPUTER MEMORANDUM NO. 151

DATA COMPRESSION BY AVERAGING EDITED VLA DATA

R. M. Hjellming

October 1979

The unfortunate implications of the large volume of VLA data have been known for a long time. Even by 1979 the large volume of data produced during astronomical observing was causing serious problems with computer through-put and disk data storage. In this memo I would like to discuss a specific suggestion for reducing the visibility data storage problem. This suggestion is based upon baseline-dependent averaging for visibility data within a "scan". A scan is defined by all visibility data records associated with a single index record.

Let f be a fraction ($f \lesssim 1$) averaging factor for a scan, and further let

$$f = f_{ARB}\, f_{uv}$$

where f_{ARB} is an arbitrary compression factor (1/n) chosen by the observer and f_{uv} is another (1/m) compression factor dependent upon $(u^2 + v^2)^{1/2}$ for each baseline. If we provided users with a program that averaged visibilities for each scan by both a factor f_{ARB} and a factor f_{uv}, with a scheme like that listed in Table I, considerable reduction in the size of visibility data bases could be achieved with no significant deterioration in data quality.

TABLE I

f_{uv}	$r_{uv} = (u^2+v^2)^{\frac{1}{2}}$ [msec]
1	>80
1/2	$60 < r_{uv} \leq 80$
1/3	$40 < r_{uv} \leq 60$
1/4	$30 < r_{uv} \leq 40$
1/5	$20 < r_{uv} \leq 30$
1/6	$10 < r_{uv} \leq 20$
1/12	$r_{uv} \leq 10$

In order to illustrate the gain of f_{uv} data compression I have used the scheme in Table I to evaluate the compression achieved for a scan of A-array VLA data for a snapshot observation at zero hour angle for a source at 89° declination. This represents a "worst case" since for lower declinations and other array configurations even more data compression would be achieved. From this observing situation simulated with the FAKVTB program, the following were computed:

$$N[(u^2 + v^2)^{\frac{1}{2}}] = \text{number of data points at radii } (u^2 + v^2)^{\frac{1}{2}}$$

and

$$N[>(u^2 + v^2)^{\frac{1}{2}}] = 351 - \int_0^{(u^2+v^2)^{\frac{1}{2}}} N[(u^2 + v^2)^{\frac{1}{2}}]\, d(u^2 + v^2)^{\frac{1}{2}}$$

(approximately, of course, by summation and use of discrete intervals). Note that the sum of all $N[(u^2 + v^2)^{\frac{1}{2}}]$ is 351. Plots of these functions normalized with division by 351 are shown in Figure 1.

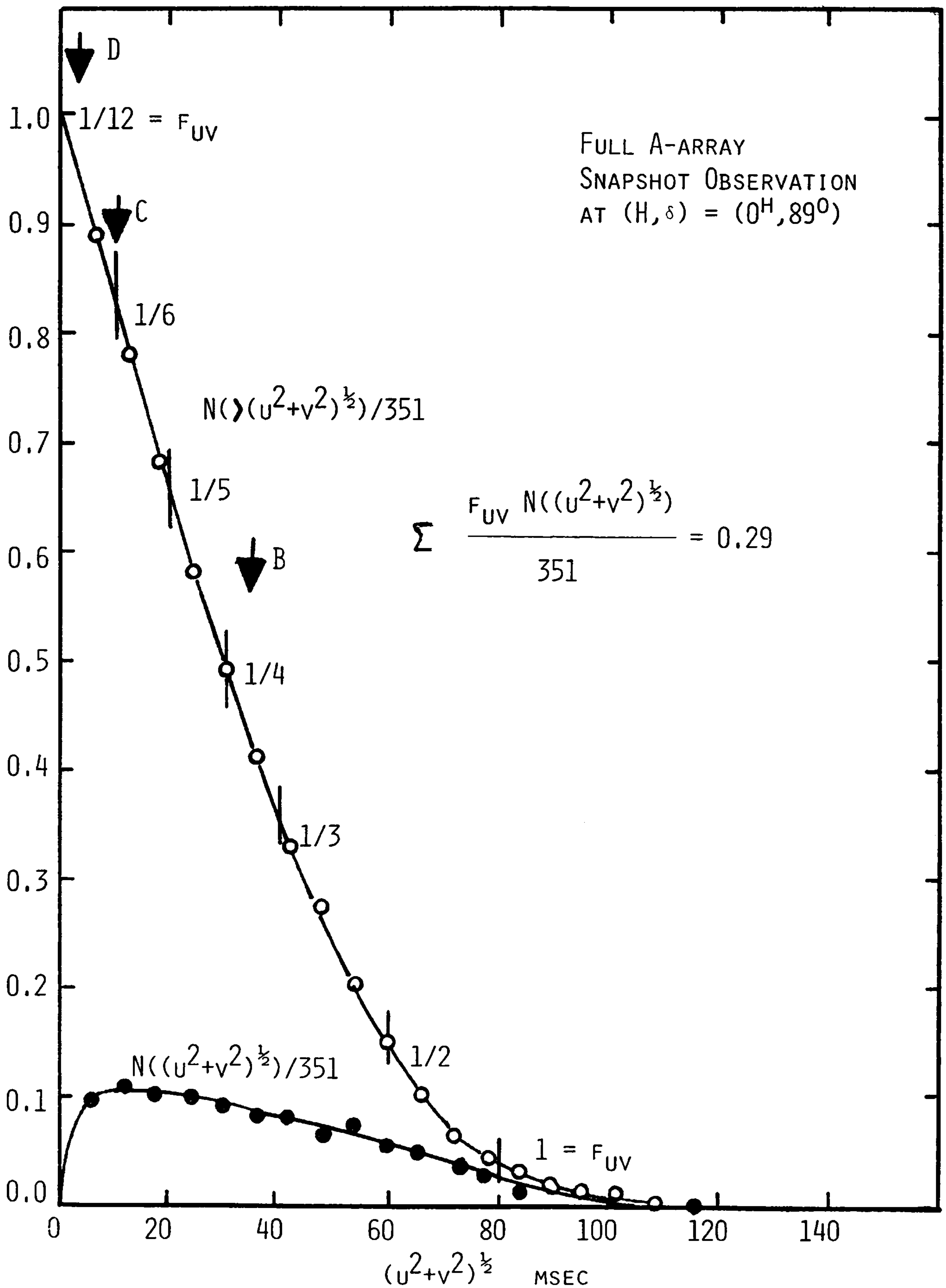
D
C
B
1/12 = F_UV
1/6
1/5
1/4
1/3
1/2
1 = F_UV
FULL A-ARRAY
SNAPSHOT OBSERVATION
AT (H,δ) = (0^H,89^O)
N(>(u^2+v^2)^½)/351
N((u^2+v^2)^½)/351
Σ F_UV N((u^2+v^2)^½) / 351 = 0.29
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
120
140
(u^2+v^2)^½ MSEC

FIGURE 1

Using the values of f_{uv} given in Table I, which are indicated in the appropriate zones of Figure 1, one can compute

$$\sum_{\substack{\text{all} \\ \text{data}}} f_{uv} \frac{N[(u^2+v^2)^{\frac{1}{2}}]}{351} \cong 0.29$$

which is the approximate fractional data compression in stored visibility data points under the averaging scheme of Table I.

For more compact configurations the observer can choose specific f_{ARB} as listed in Table II, without significant deterioration of source information in the data.

TABLE II

DATA COMPRESSION FOR DIFFERENT CONFIGURATIONS

Configuration	f_{ARB}	$0.29\ f_{ARB}$
A	1	0.29
B	1	0.29
C	1/3	0.097
D	1/9	0.032

In Table II we have taken the conservative approach of treating 10/3 sec as a more desirable basic A-array integration time to avoid too much visibility function averaging at the largest $(u^2 + v^2)^{\frac{1}{2}}$.

For sources that are known to be significantly more compact than the antenna beam size the observer can choose even smaller values of f_{ARB}.

It is practicable to write a program that will perform u,v dependent averaging that will compress visibility data bases preserving the standard format. Such averaging would be meaningful at any point after editing (reflagging). Only data with a specified PASSFLAG would be averaged and/or passed on to the new compressed data base. Averaging by f_{uv} factors would be built into the program and an additional averaging by f_{ARB} can be imposed by choice of the observer. Let us call such a program COMPRS solely for purposes of reference.

COMPRS would involve two basic steps. First, for each scan the visibility records would be sorted to have data for individual baselines in sequential order, perhaps with a baseline-to-baseline order based on

$$m_{ij} = (i - 1) * 28 + j - i*(i + 1)/2$$

for the i-j baseline. The second step, perhaps during the last sort/merge step, would be averaging based upon the largest $(u^2 + v^2)^{\frac{1}{2}}$ for a scan and baseline. The resulting DBNAME.VIS would differ from the original only in that averaged records would have different weights (original times $f_{ARB}^{-1} \cdot f_{uv}^{-1}$). According to their averaging times.

The details of the above-mentioned data compression scheme can probably be improved upon. However, I hope some such scheme could be adopted in a COMPRS program to give observers the option of working with data bases more compact than possible with the current AVGVIS program.