NATIONAL RADIO ASTRONOMY OBSERVATORY
VERY LARGE ARRAY PROJECT

VLA COMPUTER MEMORANDUM NO. 143

REPORT OF THE
VLA OFF-SITE DATA PROCESSING AD HOC STUDY GROUP

October 21, 1977

M. S. Roberts

I. PROLOGUE

A study group was formed on March 21, 1977 to consider the general problems related to VLA off-site data processing. This group met seven times to review and analyze the various factors involved in such post-processing. Interested parties were contacted and a variety of written and oral presentations were given the committee. A preliminary version of the report served as basis for discussion for a meeting of VLA personnel and committee members held at the site. Comments were solicited and, where appropriate, included in the final report.

The most remarkable aspect of this study, one that is not readily visible in this report, is the similarity in individual views as to what is necessary and how this might best be accomplished; there is no minority report. However, it is unlikely that such unanimity will be encountered when the specific details of post-processing are considered.

II. THE ON-SITE PRODUCT

As an underlying philosophy we accept the view that, because of its magnitude, detailed re-editing or recalibration of the raw data is neither feasible nor likely in most circumstances. However, the ability for gross editing, e.g., omitting baselines or time intervals, will be requested by some. No distinction is made in the following between line and continuum data since we envision continuum data to be taken on the A array while line data are obtained on the more compact C and D arrays: the amount of data to be handled in either case is comparable.

At the end of an observing run there should be available to the observer:

1) u-v data which has been edited, calibrated, and perhaps sorted. These ungridded data are tagged with baseline, time information. Some display of these data, most likely graphical and perhaps in an averaged form, is highly desirable.

2) A summary of diagnostics which resulted in data rejection or flagging.

3) Since on-line maps have been produced for decision making during the run, hard copy of these maps should be available. For line work, this may consist of maps of selected and/or averaged frequency channels.

and perhaps

4) To speed the throughput it is desirable (but not necessary) that at the end of the run, or in a reasonably short time thereafter, a map of fairly high quality (e.g., corrected, cleaned) be supplied the observer. (Timing here may be determined by map size and number of either line or continuum observations.) This should be available in both hard copy and Fortran readable magnetic tape. Compact (i.e., averaged, gridded) forms of the visibility data should also be available on magnetic tape.

Note to above: This last item regarding maps and magnetic tapes is considered only as "desirable" as part of the on-site product. It involves decisions by the observer (i.e., taper, size, number) that can be made after he is off the telescope without affecting the observations. When requested these maps should be supplied as soon as possible.

A minimum amount of map manipulation and extraction of numerical information may be sufficient for some projects and should therefore be obtainable on-site immediately after the observing run. This point requires some elaboration.

Map manipulation must be possible on-site for staff use and experimentation. If excess capability exists (e.g., nighttime), then usage by the most recent observers should be supported. In a similar vein excess capacity in the computer system is likely during many experiments. Transform operations (on a low priority but still at a useful rate), in addition to those made for the quick-look maps, should also be supported. Such usage must be carefully restricted, but the possibility of limited on-site post-processing for recent observers should be made available if possible.

The u-v data are the fundamental product. As questions arise during examination of the maps, different tapers, fields of view, and source subtraction (for example) will often make retransformation of the u-v data necessary. Gross editing of these data may eventually be little used, but it surely will be in demand initially.

A viable storage, transportation, and archiving system of the u-v data is an important problem for the VLA. It is not addressed in this report.

III. POST-PROCESSING

The output described above will be adequate for some experiments: e.g., detection, variability, position, approximate size. But for many problems further, i.e., post-processing is a necessity. A two-page list of such processing needs is given in the memo "VLA Post-Processing: Phase I", by W. R. Burns and E. W. Greisen (March 25, 1977), see Appendix A.* In brief, the astronomer wants (i) to be able to produce results of publishable quality, (ii) to be able to experiment with the data, and (iii) to re-examine the data, especially as new questions arise. And the very nature of the VLA with its movable elements requires a time scale for data reduction (for some observations) which cannot be done in real time. Post-processing is a necessity. It may be performed at various locales:

*Appendix A is classified separately as VLA Computer Memorandum No. 140.

1) At the user's home institution;
2) At an NRAO computing center;
3) Via time and space rented at a large computing facility;
4) Via a number of small (mini) computers dedicated to VLA processing and located in various high user-density areas; and
5) Some combination of the above.

Two additional possibilities may be included for completeness: a general astronomical computing center, and on-site post-processing. The former appears too far in the future and would require significant staffing with astronomers familiar with VLA techniques. The latter is not within the purview of this committee.

Each of these possibilities have been considered in detail. The recommendation of this committee is (5); specifically, a combination of (2) and (4). It is described in some detail below. The reasons for rejecting the other alternatives will be given first.

For many observers, the ideal reduction and post-processing procedure would have the user return to his home base to work on his data while involved in other commitments, such as teaching. If he returns with u-v data for just one typical 12-hour observation, he would likely be carrying one or several high-density (6250 bpi) tapes. The handling of these u-v data, even for the minimum situation, would be a major burden for a university computing center. This first step can be bypassed by going directly to a map-tape (as described above).

While some users will require only limited analysis of their map data, we envision that the bulk of the post-processing will require interactive manipulation of these data. To accomplish this a disk drive dedicated to that one observer must be available as well as a display terminal. The latter will require a coaxial cable link to the computer. An additional constraint is the potentially large amount of cpu and "real" time requirements for typical work, See Appndix B* for time estimates. Finally, software

*Appendix B is classified separately as VLA Computer Memorandum No. 142.

development assistance from NRAO will be sought; this is a reasonable involvement for NRAO but one that may be difficult to satisfy because of the many different machines and systems that are in operation.

All of these hurdles can be overcome, but it seems to be a grossly inefficient approach. The concept of interactive map manipulation at many university computer centers is not a likely approach to post-processing.

The time-rental concept, unless at some already operating government facility, means that we (the astronomical community) are willing to pay a premium for operating management. The flexibility that may be necessary for VLA post-processing will likely be absent. Going to Los Alamos or Livermore (the former was casually looked at) may, even more, remove the flexibility and will always pose, at some level, a security problem inherent in such installations. At present, 12^h/day shifts on a CDC 6600 are available at Los Alamos. This may increase when their new Cray becomes operative. Renting time is feasible but it cannot be described as desirable in so far as we have been able to ascertain.

Remaining possibilities include an NRAO center, many minis, or a combination of these two. Many minicomputers dispersed about the country without an NRAO center is not viable. Many questions (which we make no attempt to answer) are raised in this extreme situation, e.g.: who develops and maintains the software for this battery or minis? How are special projects and needs handled? Is this a cost-effective (and science-effective) approach? These questions and concerns can be answered satisfactorily by replacing the many minis with a large central computer facility. But other disadvantages can be envisioned: The potential for bureaucratic inefficiency; the necessary lack of flexibility in approach and in programming; and perhaps most important, a computing center would keep users away from their home institutions.

Must we have one extreme or the other? We hope not. A central computer facility for development work and for some visitor (and staff) usage is necessary for a repository of VLA knowledge and expertise; it is a center

for both normal and special jobs. In addition to this central (presumably modest sized) facility, several minicomputers would be available for regional use. These computers should have common, NRAO supplied, programs written in Fortran, where practical, and which can be modified by local users. Price estimates of $150,000 to $300,000 for a minicomputer, array processor, and interactive display system have been quoted. These will allow interactive map manipulation. Longer jobs (e.g., retransforming, extensive cleaning) may be run overnight if necessary. Less expensive, more modest systems also appear feasible.

Problems remain. Who gets such systems? Are they regional in the sense that those outside the institution housing such a system will have access to the system? And if so, who will maintain the machine and aid the visitor? We think such problems and the inconveniences they imply will be more than outweighed by having a reduction system in one's own backyard. An added responsibility will go with this added convenience. The user distant from such a local computing system or one with special needs would go to the central NRAO facility which would also be equipped with such minis (or a system which to the user should look like a mini). This NRAO center, in the situation envisaged here, is of modest size, perhaps not much larger (physically) than our present computer center. In addition to a computing facility, such an NRAO center should have auxiliary equipment for the most productive use of the VLA-generated data, e.g., a photographic plate measuring machine, overlay and related photographic facilities. This center must also have a scientific staff to aid the visitor in use of the computer programs and in the problems of data manipulation that will arise.

The development and growth of such a center must proceed in a natural and orderly manner. We already have some experience with post-processing on large computers. Experimentation as to the large computer's role and capabilities should continue. A minicomputer system with necessary peripherals should also be developed and tested. Once this latter system is established, others can copy both the hardware and software.

IV. IN SUMMARY WE PROPOSE:

An NRAO computing center and a number of regional, university-based, minicomputer systems dedicated to VLA post-processing.

The NRAO computing center would contain a modest sized general purpose computer and necessary peripherals for the maintenance of active NRAO programs. It would have facilities to service several VLA post-processing users simultaneously. Re-editing, if requested, could be done on-site, or with the above general-purpose computer, whichever is more feasible. Scientific and technical staff would be located at this center to aid the user. Close communication between this center and on-site personnel is of great importance.

Several (3 to ?) minicomputer systems with adequate memory and interactive displays would be located at supportive universities across the country. An identical minicomputer system would remain at the NRAO computing center for development work. These systems are fast enough to handle many post-processing tasks in a reasonable time.

V. RECOMMENDATIONS

The above is the background and justification for the following recommendations:

1. Identify and delegate responsibilities and authority to a post-processing group to begin detailed planning and implementation of the post-processing system.

2. Begin immediate development of post-processing by a minicomputer system, including software and hardware.

3. Develop preliminary cost projection of the post-processing system, including site location options and begin budgetary action now.

Study Group Members:
R. L. Brown
J. J. Condon
D. S. DeYoung
J. R. Fisher
W. Jaffe
M. S. Roberts (chairman)
L. Rudnick