MARK III INTERACTIVE DATA ANALYSIS SYSTEM

James W. Ryan and Chopo Ma

NASA/Goddard Space Flight Center

Bruce R. Schupler

Computer Sciences Corporation

ABSTRACT

The Mark III very long baseline interferometry (VLBI) software system is an integrated set of programs covering all aspects of VLBI computer activity from schedule generation through field data acquisition to the production of "publication ready" parameter values. The interactive data analysis system is a major subset of this system and consists of two major and a number of small programs. These programs provide for the scientific analysis of the observed values of delay and delay rate generated by the VLBI data reduction programs and produce the geophysical and astrometric parameters which are among the ultimate products of VLBI. The two major programs are CALC and SOLVE. CALC generates the theoretical values of VLBI delay and delay rate as well as partial derivatives based on a priori values of the geophysical and astrometric parameters. SOLVE is a least squares parameter estimation program which yields the geophysical and astrometric parameters using the observed values provided by the data processing system and theoretical values and partial derivatives provided by CALC. SOLVE is a highly interactive program in which the user selects the exact form of the recovered parameters and the data to be accepted into the solution.

INTRODUCTION

In the Mark III very long baseline interferometry (VLBI) system, complete processing of data from schedule generation through production of geophysical results involves a series of large and sophisticated programs. Figure 1 outlines this flow. The data processing can be thought of as consisting of three distinct elements: acquisition, reduction, and analysis. The programs SKED and FIELD constitute the acquisition element; these programs are used to schedule the experiments and to control the acquisition system while the raw data are being recorded. The programs DE-LOG, PREP, COREL, FRNGE, and EDIT constitute the reduction element; this includes all processing between the cross-correlation of the Mark III tapes and the production of the delay and rate observations. The programs CALIB, STRUC, ASTRO, CALC, and SOLV constitute the analysis element; included here is all data processing required to recover the geophysical and astrometric parameters from the delay and delay rate observations. This paper will be limited to a discussion of the analysis element.

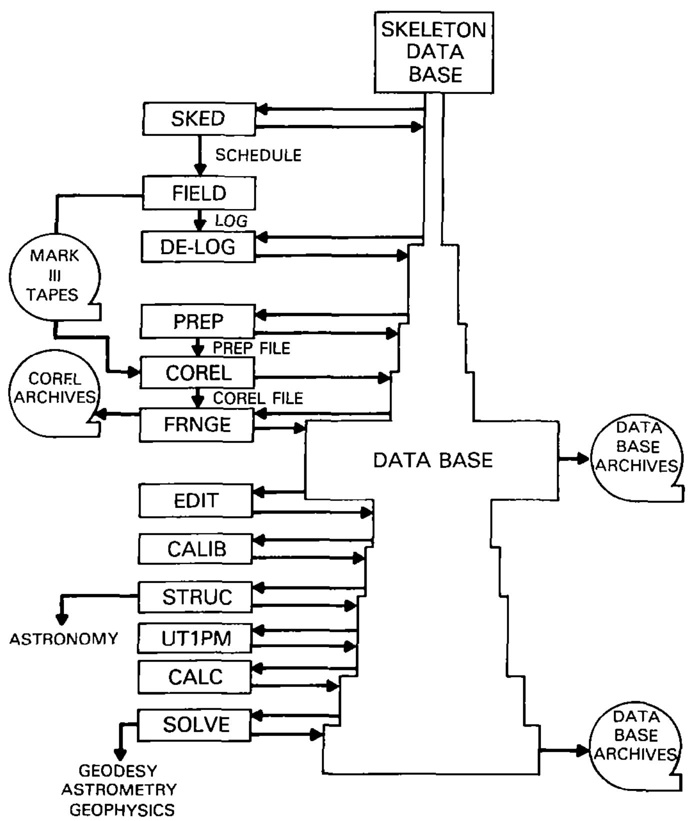

Figure 1.

INTERACTIVE ANALYSIS SYSTEM DESIGN CONSIDERATIONS

In 1975 when NASA elected to pursue the development of the Mark III VLBI system, the decision was made to develop an entirely new analysis system as well as hardware and data reduction systems. At that time, a program called VLBI3 was used to analyze VLBI data. It was a large batch mode program which could be run on the IBM 360 computers at Goddard and the Massachusetts Institute of Technology (MIT). Based on experience with VLBI3, a series of overall design considerations was set. Figure 2 highlights those considerations, and the following summarizes them.

1. HIGH LEVEL OF DATA INTEGRITY AND ACCESSIBILITY
2. HIGHLY RESPONSIVE SYSTEM
3. PROCESS AN UNLIMITED NUMBER OF POINTS AND RECOVER AN UNLIMITED NUMBER OF PARAMETERS
4. BUILT WITH STATE-OF-THE-ART MODELS
5. SIMPLE TO USE
6. WELL WRITTEN, DOCUMENTED, AND CONTROLLED
7. INEXPENSIVE TO USE

Figure 2. VLBI analysis system design considerations.

1. The system should provide for an extremely high level of data integrity. It should be easy to access any data set, and any data of permanent value should be saved automatically.

2. The system should be responsive. It should be possible to carry out simple analysis tasks in a matter of minutes and more complex tasks in minutes to hours.

3. Subject only to the physical limitations of the computer on which the system would operate, it should be possible to combine an unlimited number of data points and recover an unlimited number of parameters.

4. The system should contain state-of-the-art models for all phenomena which affect VLBI data.

5. The system should be simple to use. System operation should be straightforward and not require knowledge of the internal workings of the programs. With only a few hours of "hands on" experience it should be possible for a person who is familiar with VLBI analysis procedures to operate the system competently. The system should present the user with a default set of analysis specifications which in most cases would require only a small number of modifications to generate the desired result. The system should attempt to protect the user from blunders and should log with the output all analysis specifications set by the user.

6. The configuration of the system should be well controlled. It should be possible by reviewing the output of the system to know exactly what versions of the software generated the output. Moreover, the software should be completely documented and written with modern concepts of good code.

7. The system should be so inexpensive to operate that the cost of the computer time required to carry out data analysis would never be a consideration in deciding whether or not to perform the analysis.

RESPONSE TO THE DESIGN CONSIDERATIONS

In response to the first design consideration, the Mark III Data Base System was developed. The system consists of a set of files and a software system for accessing the files. In addition, there is a catalog system which allows the user to find any file quickly, which tracks the relationships among files, and which provides for nearly automatic archiving of files. The data system is covered in detail in the appendix to this paper.

Once the decision was made to develop a formal data system the second design consideration, system responsiveness, quickly surfaces as the driving consideration. The existing analysis program, VLBI3, was unresponsive because it was run as a batch job on a large computer system and because it calculated from first considerations all theoretical and partial derivatives for the data being processed even though that data may have been analyzed many times in the past. For complex analysis tasks, VLBI3 required amounts of computer time which were often available only at night or on weekends. Based on this experience, the decision was made that the new system should be an interactive system; that is, that the analyst using a terminal would set up the run via a dialog with the program. The program would carry out the estimation and return the result to the user's terminal or printer in at most a few minutes of elapsed time. This could not be accomplished using the large computer systems available at Goddard; only through the use of a dedicated minicomputer could an interactive system be achieved.

The decision to use a minicomputer, which has much slower execution speed than the large batch machines, forced the decision to divide the analysis into a number of smaller, separable tasks. In the main, the tasks were divided into two groups: those which are carried out only once or infrequently for a given data set, and those which must be carried out every time a new analysis is attempted. The programs CALIB, STRUC, and UT1PM were specified as small, "stand alone" minicomputer programs. Their functions are to add to a data base calibration data, source structure corrections, and UT1 and polar motion tables respectively. At least in principle, the information they store is to be entered only once for each data base. The function of computing the theoretical observations and partial derivatives and the function of doing the least squares estimation, both of which were done in VLBI3, were assigned to two programs: CALC and SOLVE, respectively. Computing the theoretical observations and partial derivatives is a typical example of a function which must be carried out only infrequently. Since in our application, the solutions need not be iterated, the only time the computation of the theoreticals and partials would be redone would be if the a priori constants were changed or if a model were improved. The least squares processing obviously depends on the detailed specifications of the analysis to be accomplished. Separating these functions proved to be the key decision in achieving adequate system responsiveness. With the system as it exists today, new data can be processed through the entire chain of analysis programs in less than 24 hours and old data can be reprocessed through SOLVE in a matter of minutes.

Achieving the second design consideration, the ability to process a large number of data points and to recover a large number of parameters, proved to be exclusively a function of SOLVE. It was implemented on a Hewlett-Packard (HP) 21MX minicomputer, which has rather severe constraints on program size (programs may not exceed 32K 16-bit words). Because of this limitation, it was not possible to maintain large matrices in core, and if a standard least squares algorithm had been implemented, fewer than 100 parameters could have been recovered. In order to circumvent this limitation, a algorithm called arc parameter elimination was implemented. The parameters to be recovered are divided into two classes: arc and global parameters. Arc parameters are those which affect only a limited, usually time-bound, subset of the data. These include clock polynomials, tropospheric refraction scale factors, and values of UT1 and polar motion. Global parameters are those which affect a broad cross-section of the data; these include the site and source coordinates, the precession constant, and earth tide parameters. Using the arc parameter elimination technique, data are added to the solution one experiment at a time, and at each step, the arc parameters are algebraically eliminated. Only a matrix containing the global parameters is carried forward. Once all the data have been added, the global matrix is inverted. The values of the parameters which result are identical to those which would have resulted from an inversion of a matrix containing the global parameters and all the arc parameters. With a modest amount of additional effort, the arc parameters can also be recovered. We have already produced a solution based on 8171 data points containing 535 recovered parameters of which 485 were arc dependent and 50 were global, and this solution scarcely tested the limits of the system.

The third design consideration, that state-of-the-art models should be used, involves nearly all of the analysis programs but is of special significance for CALC. Currently CALIB is little more than a prototype of the program which will ultimately calibrate the effects of cable delay and tropospheric refraction. The experiments which have been conducted to date have not produced on an operational basis the data needed to carry out these calibrations. Once the water vapor radiometers become operational, CALIB will be brought up to the state-of-the-art. STRUC is now no more than a program specification, and even the source structure models which would drive this program have not been tested to see if they would improve the analysis of delay and delay rate data. The quality of the data acquired to date has been such that source structure corrections have not been needed, but it is expected that the Mark III system will produce data which will require source structure corrections if the full potential of the data is to be achieved. CALC, the program which computes theoreticals and partials, has been fully operational for almost 2 years. CALC may well be the most complete program in existence for modeling VLBI delay and delay rate, and with the exceptions of ocean loading and ionospheric refraction, contains models for every phenomenon which significantly affects VLBI observations. The models in CALC were the subject of Chopo Ma's paper "Geophysical and Astronomical Models Applied in the Analysis of VLBI Data" already presented at this conference. SOLVE does not contain models for VLBI observations.

The fourth consideration, that the programs should be simple to use, is one that we have worked with great vigor to satisfy. UT1PM is an interactive program which operates on the HP 21MX and stores UT1 and polar motion information in data bases. It prompts the user in plain English at all decision points and is so foolproof that it could be run with no more instruction than how to set the program running. CALIB and STRUC are not yet operational, but they will be written with the same standards for simple use as UT1PM. CALC is maintained by a single individual in our group, but the operation of CALC is so simple that anyone of our group who desired a CALC run may generate independently. Anyone modestly familiar with VLBI processing could be taught to run CALC with no more than 30 minutes of instruction. Simplicity of use has had its greatest impact on the design of SOLVE. As stated above, SOLVE takes the actual and theoretical observations and partial derivatives placed in the data base by other programs and performs a least squares estimation to recover geophysical, astrometric, and other parameters. In order to carry out this function, data bases must be selected, individual data points must be accepted or rejected, and the configurations of clock polynomials, tropospheric refractions scale factors, and numerous other parameters must be specified. To accomplish this, SOLVE presents the user with a series of plain text menus which contain the various options which may be selected. After consulting the selected data bases, the program sets up a default configuration for all parameters which may be selected. To make a specific run, the user usually selects only a small number of options in the menus. The program contains numerous cross-checking procedures to help insure that the user does not blunder. For example, if more than one data base is accessed, the program checks to insure that the a priori constants used to generate the theoretical observations are identical. If the user attempts to solve for a parameter for which the data has no sensitivity, SOLVE detects that condition, informs the user, and generates the solution with that parameter eliminated. Since SOLVE presents the analyst with so many options for combining the data and selecting parameters, and since this power could generate many improperly specified solutions, considerable further effort will be spent placing more sophisticated cross-checking procedures into SOLVE.

The fifth design consideration is that of software configuration control and software coding and documentation standards. UT1PM and CALC have been the most successful in meeting this design consideration. Both programs were completed early in the project and have been stable for almost 2 years. All operational CALC processing has been accomplished with four thoroughly documented benchmark systems. Moreover, the old systems have been maintained so that any CALC run can be recreated. Both programs have been written with explicit coding and documentation standards. The two guiding principles have been "keep it simple stupid" and "machines are cheap and people are expensive." In any conflict between coding efficiency and coding clarity, clarity has always been chosen. We have not been as successful with configuration control for SOLVE, but that reflects more than anything else the nature of the SOLVE's job. It has been and still is the program which limits the complexity of the data analysis our system can achieve. There is relentless pressure to bring the latest improvements in SOLVE on line quickly. In the fall of 1979, a new version of SOLVE with significantly upgraded capabilities will become operational, and SOLVE should become more stable.

The final design consideration is that the cost of computer time to carry out the analysis should be negligible. This has been by far the easiest design consideration to achieve. When compared with the cost of reducing data such as satellite orbit measurements or planetary ranging data, the cost of analyzing VLBI data has always been small. Even so, at MIT, analysis with VLBI3 was often hampered by the availability of computer funds. In our current system, most of the work is carried out on the dedicated minicomputer, and no incremental cost is incurred in making a run. The total computer cost for the operation and maintenance of CALC on the IBM 360-91 at Goddard is less than $3000 per year, an insignificant cost in the overall project budget.

SUMMARY

Since 1975, our group has expended almost 15 man-years in the development of the Mark III VLBI analysis system. We have by no means completed our task, but we can already carry out analysis tasks which only a few years ago could only be dreamed of. It is our hope and expectation that the analysis system will always measure up to the quality of the data produced by the Mark III data acquisition system.