UCRL--93878-Rev.1

DE86 009894

DATABASE TOOLS FOR ENHANCED ANALYSIS OF TMX-U DATA

M. E. Stewart, M. R. Carter, T. A. Casper,
W. H. Meyer, D. E. Perkins, and D. M. Whitney

This paper was prepared for submittal to the Proceedings of the 6th Topical Conference on High Temperature Plasma Diagnostics, American Physical Society, Hilton Head Island, South Carolina, March 9-13, 1986.

March 6, 1986

Lawrence Livermore National Laboratory

DATABASE TOOLS FOR ENHANCED ANALYSIS OF TMX-U DATA

M. E. Stewart, M. R. Carter, T. A. Casper, W. H. Meyer,
D. E. Perkins, and D. M. Whitney
Lawrence Livermore National Laboratory, University of California
Livermore, CA 94550

ABSTRACT

A commercial database software package has been used to create several databases and tools that assist and enhance the ability of experimental physicists to analyze data from the Tandem Mirror Experiment-Upgrade (TMX-U) experiment. This software runs on a DEC-20 computer in M-Division's User Service Center at Lawrence Livermore National Laboratory (LLNL), where data can be analyzed offline from the main TMX-U acquisition computers. When combined with interactive data analysis programs, these tools provide the capability to do batch-style processing or interactive data analysis on the computers in the USC or the supercomputers of the National Magnetic Fusion Energy Computer Center (NMFECC) in addition to the normal processing done by the TMX-U acquisition system. One database tool provides highly reduced data for searching and correlation analysis of several diagnostic signals within a single shot or over many shots. A second database tool provides retrieval and storage of unreduced data for use in detailed analysis of one or more diagnostic signals. We will show how these database tools form the core of an evolving offline data analysis environment on the USC computers.

I. INTRODUCTION

Data acquired from the TMX-U experiment at LLNL is processed and analyzed on six HP-1000 computers that comprise the main data acquisition and analysis system for the experiment.[1] These computers form a stand-alone system for acquiring, processing, and analyzing data from numerous diagnostics and sensors used to measure plasma properties and other conditions within this experimental fusion device. Physicists analyze data using a combination of one or more custom programs and visual inspection of the plotted data. Because of the volume of data and the desire to maximize the experiment repetition rate, this system accommodates only a limited amount of interactive processing.

During the past 18 months, we have been working to create an interactive data analysis environment for TMX-U which complements the normal intershot batch style of data analysis. This environment is a mixture of databases and interactive software tools (programs) on a DEC-20 computer and provides the experimental physicists of M-Division with a highly interactive, offline environment for data analysis. Other benefits include the capability to perform batch style data analysis, access to experimental data for M-Division's theoretical physicists, and access to the NMFECC, where the compute power of several Cray supercomputers can be used to more closely model TMX-U processes.

In this paper, we give an overview of the databases and tools that have been assembled to form this environment. We will briefly discuss the features and uses of the databases and then we will present the major features of the data analysis tools. Finally, we will outline some future directions for the project.

II. DATABASES

The most important feature of this environment is the capability to store a reasonable amount of experimental data with the limited disk resources of the User Service Center's DEC-20 computer system. This is accomplished using two separate databases with very different approaches. One database, known as the full resolution database, is used to retrieve and store time series data from a given diagnostic at its full temporal resolution. The second database, known as the journal database, holds a highly reduced set of averaged data which generally represents most interesting features of a given shot. Figure 1 illustrates the role of these databases in giving users access to TMX-U data and making that data accessible to various data analysis tools. Both databases are managed using a commercial database management package from National Information Systems, Inc. known as Accent R.

Accent R (Ref. 2) is actually a database management system combined with a high level programming language (commonly know as a fourth generation language) with options for both compiled and interpreted modes. Each database consists of several data sets (or files), a utility program (written in the Accent R language) that provides the user interface to the database, and one or more Pascal/Fortran programs used by the database manager for moving data into and out of the database. In addition, there are high-level Fortran libraries that allow user Fortran programs to access any or all of the stored data.

The full resolution database gives users the capability to retrieve specific diagnostic data and store it online for a finite amount of time. Using the database utility, a user may request data retrieval. When a request is made, a batch job is executed that reads the complete shot database from an archive tape generated on the TMX-U acquisition computers. The requested data is located and retrieved from this database, which is subsequently deleted.

The database utility allows the user to examine the contents of the database to any level of detail. At the highest level, the user can determine which diagnostics are available for a given shot or, conversely, given a diagnostic, the user can see which shots are currently available. At the lowest level, the user can see data values at specific (nonsequential) times. The user can also plot, extract or delete the data for any diagnostic in the database. Data can be extracted into an ASCII file or moved directly into the SIG signal analysis code discussed below. Limited disk space makes an automatic removal facility necessary. Currently, a diagnostic is automatically removed from the database after 45 days if the space is needed for newly requested data.

The journal database gives users the ability to compare data and study correlations either between many diagnostics within a shot, or between diagnostics over many different shots. Using highly reduced data, shots can be stored as a collection of specific diagnostic signals and other information that represent the most important features of the shot. Currently, the journal database holds in the neighborhood of 600 shots in this form. With this many shots stored online at one time, correlations between shots become much easier to find and analyze.

Each shot in the journal database is represented by various kinds of data. Keywords describe various binary (i.e., on/off) characteristics of a shot. A typical keyword is "ECRH". If this keyword has been set for a given shot, it implies that ECRH heating was present for this shot. Parameters describe various scalar characteristics of a shot. An example is the frequency of a particular ICRH antenna. Vectors are provided to describe miscellaneous characteristics that can be grouped together such as steady-state magnet currents or the frequencies of an array of ICRH antennas.

The largest amount of data is made up of time series signals that describe various diagnostic results. To provide the data reduction and allow correlations between various diagnostics to be detected, all time signals are resampled to the same fixed frequency and range through a process of averaging and interpolation. The current time base is defined to be 1 kHz from 0 to 80 ms and results in a reduction of 25:1 for most current diagnostics. At this time, approximately 65 different diagnostics are used in the journal database to describe a shot. Each diagnostic is typically made up of several channels resulting in an average of about 325 channels per shot.

Finally, the database also provides space for user comments. Comments can be entered by a user at any time and can be attached to any data element except keywords. Thus a diagnostic's faulty signal can be flagged as part of the signal data or a particularly noteworthy shot can be indicated by a comment attached to the entire shot.

The journal database utility is much more flexible and complex than the full resolution database as is necessary for a larger database, but also because its intended function is much different. In addition to the features available in the full resolution database utility, the journal database utility is characterized by a special search mode. The simplest types of searches involve dates, shot numbers, or a specific keyword. For instance, searching for the keyword "ICRH" would immediately select all shots for which ICRH heating was used. Normally, searches act on the results of the previous search, resulting in an ANDing of all of the search criteria to form a subset of shots.

More complex searches can be made as well. Parameters, vectors, and time signals can be used singly or in combination as long as the criteria can be made into a relational expression. As an example, the relational expression

3.5 * A / X < Y[a](t) ,

where A is a parameter, X and Y are diagnostic names, a is a channel name, and t is a specific time, is a valid search criteria. If diagnostic X satisfies this relation for any channel and any time (looking at only one time and one channel for diagnostic Y), the shot is flagged as satisfying the search criteria. Shots for which this relation cannot be computed (e.g., shots for which X was not acquired), are put into a special holding set. Shots in this set can be moved into the selected subset singly or as a set (only as long as a subsequent search has not been made). More complex expressions can be built using standard Fortran functions such as SIN, EXP, LOG, and ABS. Additional functions include SLOPE, INT (integral) and DIFF (difference).

At any point, the current subset can be saved, including annotations on the search sequence that produced that subset. The user can also back up to the previous subset (by "UNDO"ing the previous search) or restore a previously saved subset. The number of shots in the subset is always included in the prompt. When in search mode, many commands act only on the currently defined subset. An added feature is an "ensemble extract" command which extracts a single value from each shot based on the same kind of relational expression as the basic search command. As an example, a user could extract the central cell density at the time when the ICRH antenna was turned off even though the ICRH was turned off at different times for different selected shots.

The journal database and its utility program give the user the capability to search large bodies of data in a rapid yet precise way. Although complex searches on large sets of shots may take a few minutes, many times the same search using the manual methods previously available would be unthinkable.

III. TOOLS

The two most notable tools used in our environment are packages developed outside the MFE program for general, interactive data analysis. Both combine sophisticated data analysis techniques with online help to get users up the learning curve quickly. Both also offer graphics capabilities to help the user quickly understand the results.

The first package is a signal analysis program known as SIG. SIG was developed in the Engineering Research Division at LLNL and provides facilities for analyzing, filtering, and transforming time signal and frequency data. Also available are commands to simulate various time signals, do point by point signal arithmetic, and edit time signals. SIG is customizable through the use of a parameter file and can be menu or command driven. Menu mode acts as a tutor by walking through the commands necessary to perform a menu item. In addition, SIG offers its own internal database so that users can retain processed data over many separate sessions. It is specifically designed to handle large amounts of data and this makes it a very attractive tool for analyzing the experimental data at its full resolution.

The second package used extensively to analyze TMX-U data is DATAPLOT,[3] a graphical and statistical analysis program developed at the National Bureau of Standards. DATAPLOT's major feature is its ability to fit data to nearly any algebraic model. The residuals can then be studied to determine whether an adequate model has been chosen. Techniques for studying the residuals are both statistical and graphical. The capabilities of DATAPLOT also include smoothing, spline fitting, and analysis of variance. Another useful feature of DATAPLOT is the ability to do algebraic manipulation: virtually any legal Fortran expression can be used to manipulate full or partial data sets. DATAPLOT's powerful techniques are most useful in analyzing correlations between diagnostics within a given shot and differences in diagnostic signals over several different shots. This makes DATAPLOT most useful in analyzing data from the journal database.

IV. FUTURE DIRECTIONS

The goal of this project is to create a fully integrated environment using databases and various data analysis tools as shown in Fig. 2. This will involve work in three basic areas: First, the journal database with its searching capabilities has been an experimental effort to put better analysis capabilities into the hands of TMX-U physicists. We will continue to investigate ways to improve and augment the capabilities of this database tool including other appropriate data analysis programs. Second, the data in the journal database is most useful when it is new. We are working on a local area network, MNET, which will be able to move TMX-U data to the User Service Center. Coupled with a more automated database manager, data will be available in the journal database sooner than with the current method which is largely manual. Finally, we will be working to better integrate these tools into a unified environment with a single user interface and the ability to move data between the database and data analysis programs with greater ease.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

REFERENCES

1. T. A. Casper, H. Bell, M. Brown, M. Gorvad, S. Jenkins, W. Meyer, J. Moller, and D. Perkins, "The TMX-U Computer System in Evolution," proceedings this conference.

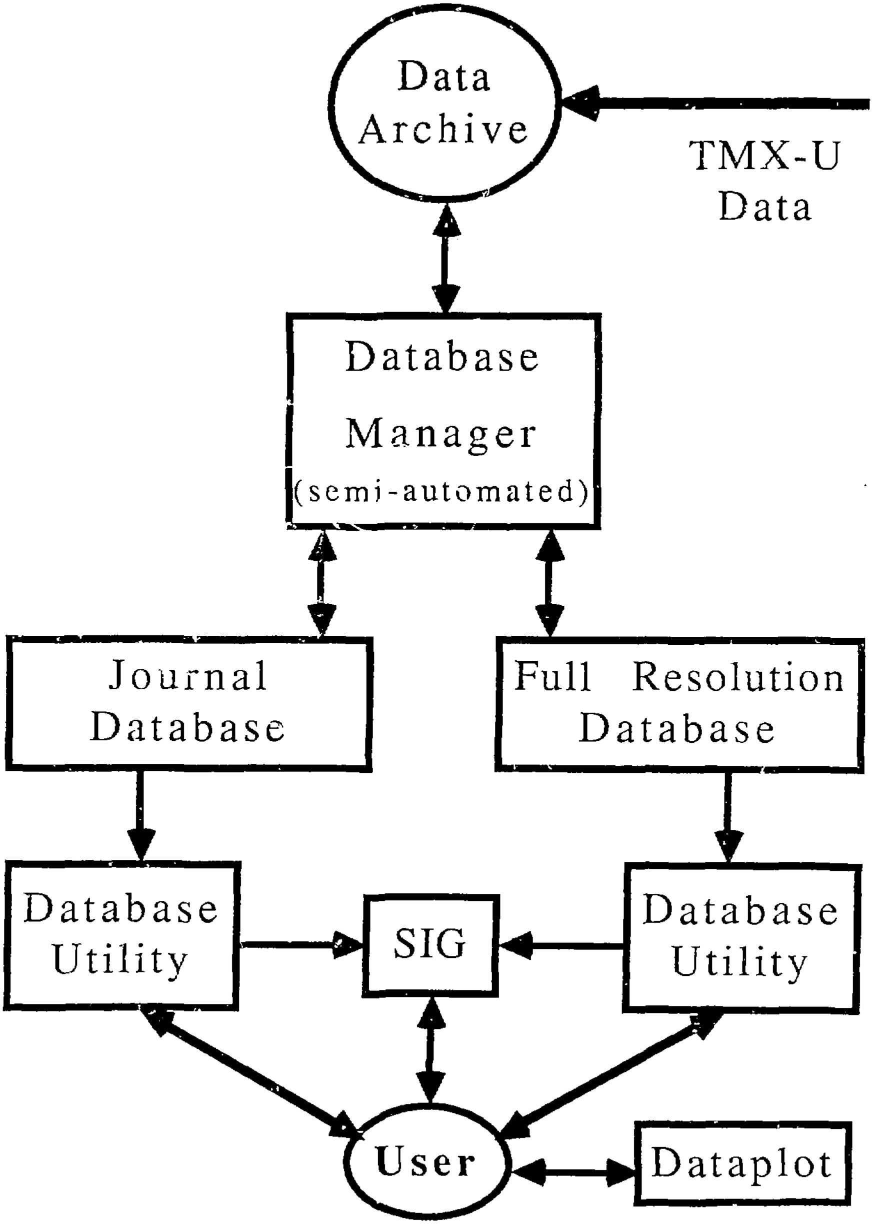
2. "ACCENT R User's Manual," National Information Systems, Inc., Cupertino, CA (1984).

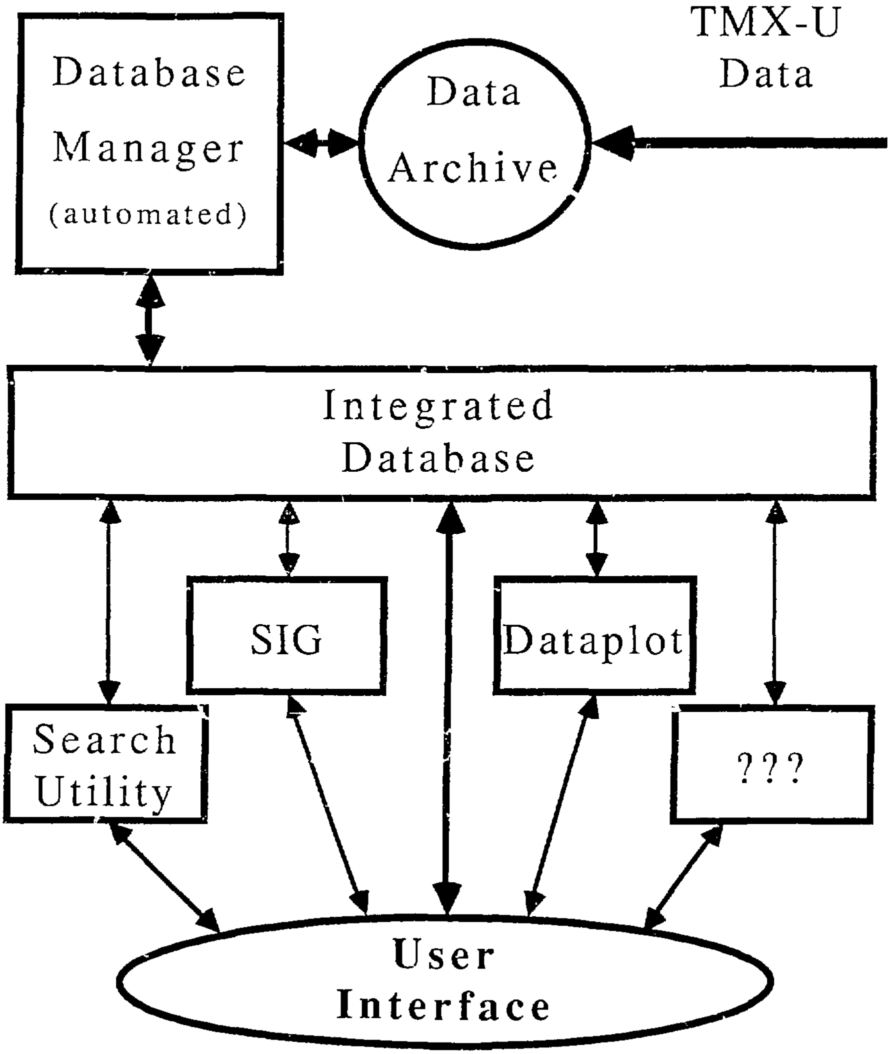
3. J. J. Filliben, "DATAPLOT--Introduction and Overview," National Bureau of Standards, Washington, D.C. (1984).

FIGURE CAPTIONS

Figure 1. Data flow between databases, tools, and users in the current environment.

Figure 2. Data flow in a future environment that closely integrates tools and databases.


Data
Archive
TMX-U
Data
Database
Manager
(semi-automated)
Journal
Database
Full Resolution
Database
Database
Utility
SIG
Database
Utility
User
Dataplot


Database
Manager
(automated)
Data
Archive
TMX-U
Data
Integrated
Database
SIG
Dataplot
Search
Utility
???
User
Interface