ornl

OAK RIDGE NATIONAL LABORATORY

UNION CARBIDE

ORNL TM-7464

MASTER

EBT Data Acquisition and Analysis System

R D Burris
D E Greenwood
J S Stanton
K A. Geoffroy

OPERATED BY
UNION CARBIDE CORPORATION
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

ORNL/TM-7464
Dist. Category UC-20 f,g

Contract No. W-7405-eng-26

FUSION ENERGY DIVISION

EBT DATA ACQUISITION AND ANALYSIS SYSTEM

R. D. Burris
D. E. Greenwood
J. S. Stanton

Computer Sciences Division

and

K. A. Geoffroy

Fusion Energy Division

Date Published - October 1980

NOTICE This document contains information of a preliminary nature. It is subject to revision or correction and therefore does not represent a final report.

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
DEPARTMENT OF ENERGY

CONTENTS

ACKNOWLEDGMENTS

Our thanks to G. R. Haste, F. W. Baity, and R. J. Colchin for guidance and suggestions in the structuring of this system. We are indebted to C. R. Stewart, Jr., for the PDP-11 Tektronix plotting package, C. O. Kemper for the PDP-11-to-PDP-11 communications software, and D. R. Overbey (FED), John O'Donnell, and Karen Hill (both from the Princeton Plasma Physics Laboratory) for the CAMAC support software. Finally, we appreciate the efforts of Teresa Craig in the preparation of this manuscript.

ABSTRACT

This document describes the design and implementation of a data acquisition and analysis system for the EBT fusion experiment. The system includes data acquisition on five computers, automatic transmission of that data to a large, central data base, and a powerful data retrieval system. The system is flexible and easy to use, and it provides a fully documented record of the experiments.

1. INTRODUCTION

1.1 GENERAL DESCRIPTION

The ELMO Bumpy Torus (EBT) is an experimental device for studying controlled fusion reactions and their application to power generation. It includes several diagnostic devices which acquire some set of static or dynamic parameters of the plasma. Since the purpose of EBT is research, the acquisition of diagnostic data is the primary objective of the device.

The implementation of the computer system described herein has enhanced the quantity and quality of the acquired data. In addition, the development of this system allows the design of the experiments which were extremely difficult or impossible because they required a high rate of data acquisition, a massive amount of data to be collected, or real-time adjustments in parameter settings. The computer will also be used to provide control of the experiment and to perform ordinary repetitive tasks.

Several computers are involved in the EBT experiment, including the Fusion Energy Division (FED) DECsystem 10, two PDP-11/34's, two PDP-8's, and a Nuclear Data 6600 (ND6600). We will describe their applications, the relationships between them, the interface between the computers and the users, and the measures taken to ensure reliable, useful operation of the entire system.

1.2 OVERVIEW OF THIS DOCUMENT

This document contains a description of the computer hardware configuration and discussions of each of the following design goals of the EBT data acquisition and analysis system:

(1) usefulness,
(2) reliability,
(3) flexibility, and
(4) maintainability.

2. HARDWARE/SOFTWARE CONFIGURATION

2.1 GENERAL STRUCTURE

The EBT system has three levels of computer hardware — the FED DECsystem 10, PDP-11/34's, and data acquisition computers (PDP-8's and a ND6600 analyzer). They are connected as shown in Fig. 1.

The interconnections shown in Fig. 1 are more complicated than a single line can show. There are two physical paths between each of the EBT computers, one for each mode of data transfer. In the more traditional mode, files of data are passed from one computer to another, with control of destination being provided by the FED network command string.[1] The other mode is a request/response facility that permits the exchange of limited quantities of information between computers.[2,3]

2.2 DECsystem 10

The DECsystem 10 is a time-sharing system which a relatively inexperienced person can use effectively. It has several software packages which make its utilization even more attractive, such as the DISSPLA plotting package,[4] the Tektronix plotting packages,[5] the System 1022 data base system,[6] and several numerical analysis libraries including NAG and IMSL.[7,8] It also has a considerable amount of mass storage available (disks and magnetic tapes), so it is useful for the accumulation of large quantities of data.

2.3 SPECIAL PURPOSE COMPUTERS

There are two PDP-11/34's in the EBT system. One is known as the PDP-11/34-C to denote its chief function: communications. This is the central processor in the EBT system and the computer to which all others are connected. All processing related to communications is performed on this processor, including the following:

(1) providing system date and time to all computers via request/ response facilities (known as function calls);

ORNL-DWG 80-3025 FED

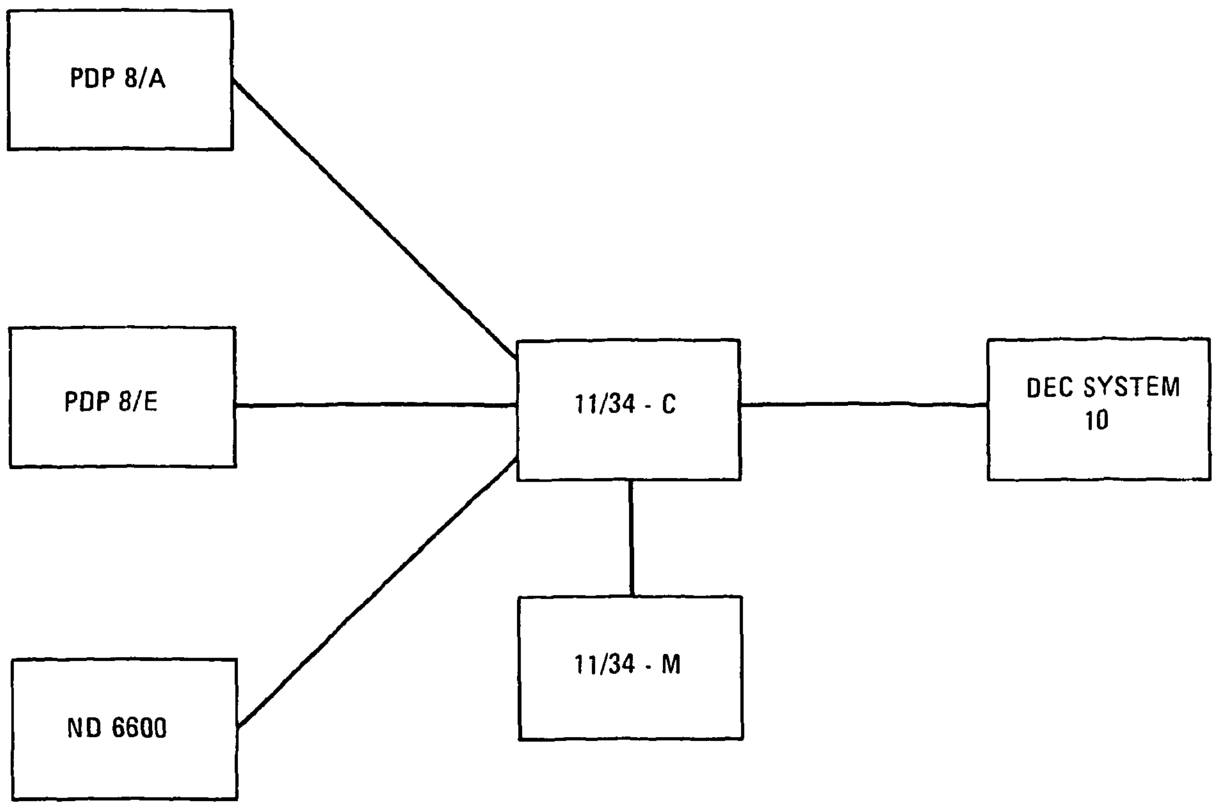

Fig. 1. EBT computer connections.

(2) providing sequencing information (described later) via function calls;

(3) receiving, translating, storing, and retransmitting data files (described later).

The PDP-11/34-C also has a CAMAC[9] computer data acquisition interface, but data acquisition programs are not permitted to degrade the performance of communications tasks. The charge exchange and RPI heavy-ion beam probe diagnostics (The term "diagnostic" will be used throughout to refer to a diagnostic device attached to EBT and the data acquisition program associated with it.) satisfy this constraint without loss of data integrity and, so, are currently implemented on the PDP-11/34-C.

The second PDP-11/34 is designated PDP-11/34-M, denoting machine, since it acquires EBT machine condition data such as microwave power, pressure, etc. One terminal attached to this computer displays machine condition data, and another is used by the experiment coordinator to control aspects of the experiment and data acquisition. Additional diagnostics will use this machine for data acquisition; plans include computer support for Thomson scattering, error field measurement, and microwave interferometry experiments.

2.4 DATA ACQUISITION COMPUTERS

The two PDP-8's, the PDP-8/E and the PDP-8/A, are used entirely for data acquisition and associated control of a single diagnostic at a time (since the operating system will permit only one program to run at a time). The PDP-8/A will be used for spectroscopic data acquisition, both for the monitoring of H-alpha signals and for the monitoring of impurity levels. The PDP-8/E is currently being used for the acquisition of soft x-ray data.

Finally, the ND6600 provides acquisition and some preliminary reduction of x-ray data. The ND6600 was purchased with considerable packaged acquisition software, and analysis programs for preliminary reduction have been added. The preliminary results can be used as a basis for decisions regarding immediate changes in machine conditions. It is anticipated that final detailed analysis will be performed on the DECsystem 10.

3. USEFULNESS

One of the goals of the development of this system was to produce a system that would be easy for experimentalists to use regardless of their computer expertise. We called this attribute "usefulness" and during the design made several decisions based on usefulness which had great impact upon the system. These decisions designated a primary programming language, a standard format for data storage, the structure of the stored data, and a procedure for data retrieval. Each will be discussed in this section. Some usefulness considerations which also received emphasis in the design and implementation of the system were automatic operation (where possible), clear instructions, and effective documentation. The value of these qualities is obvious, so they will not be discussed further.

3.1 PROGRAMMING

Programming in this system is done in FORTRAN whenever feasible for several reasons:

(1) FORTRAN is available on all of our computers.

(2) The language is simple and easily understood by programmers and experimentalists. Many experimentalists can develop or maintain their own programs.

(3) FORTRAN is portable, so the programs can be run on other machines.

(4) FORTRAN programs are easier to maintain than programs in lower-level languages.

All diagnostic and machine monitoring programs are written in FORTRAN, as is the PDP-11 plotting software.[10]

3.2 DATA FORMAT

EBT has several types of computers, and each type has its own architecture and, therefore, its own format for the internal representation of data. The PDP-8's store integer data as 12-bit words and floating point data as three 12-bit words. The PDP-11's and the ND6600 store integer data as 16-bit words and floating point data as two 16-bit words. The DECsystem 10 stores integer and floating point data as 36-bit words.

3.2.1 Format Translation

Since the architecture of the EBT data system requires that data acquired at any computer be usable on other computers, some conversion of the data must take place. It could be done at the source computer, at an intermediate computer, or at the DECsystem 10.

The alternative of translating the data at the source was considered but rejected because the processing power necessary to perform translations was needed for other purposes. We rejected the third option because all possible implementations imposed unacceptable overhead on the DECsystem 10. Accordingly, we decided to convert the data at the PDP-11/34-C, since it is connected to all the other computers and since its primary purpose is to do communications processing. The translations to be performed always produce an output file in the format of the computer to which the data are to be sent. (See Ref. 11 for a description of the translation routines.)

3.2.2 User Interface

Translation has been made as simple as possible. A PDP-8 program saves the data on disk on the source computers using a special subroutine. When the file is complete and the program exits, the user invokes the transmission program. (See Fig. 2 for the flow of data as described here and in the next few subsections.)

The following actions take place automatically:

(1) a file name is created;
(2) the file is sent to the PDP-11/34-C;
(3) the data are translated to PDP-10 format;
(4) the file is sent to the PDP-10;
(5) the file is entered into the data base.

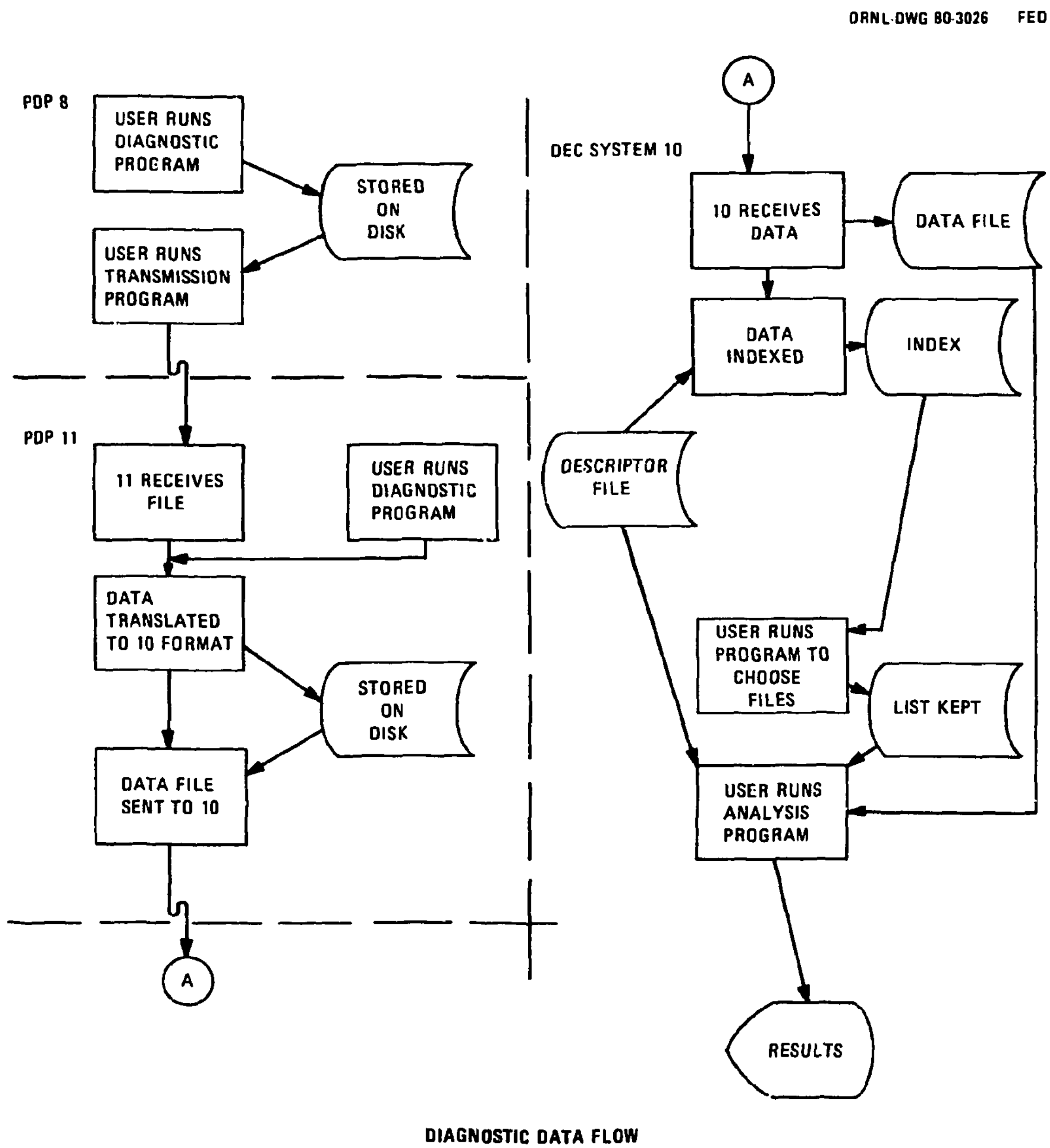

Fig. 2. Diagnostic data flow.

A PDP-11 program differs somewhat in that it calls subroutines to initialize the translation facility and then calls a translation subroutine (as often as necessary) to store the data on disk. Finally, the program calls a termination routine which completes the actions listed above.

3.3 DATA STORAGE

Data are stored in the EBT data base as conglomerates (files) of several basic types:

(1) information that defines the state of EBT (machine condition data);

(2) data that are acquired from some individual diagnostic device (diagnostic data);

(3) data which are so important that they provide a means of choice between groups of data (indexed data, either from machine condition or diagnostic data);

(4) data that describe the format of the types above.

These files are organized into two types of structures.[12] The index structure is implemented under the System 1022 data base management system and is used to determine which data files satisfy some set of criteria. The data structure enables the retrieval of the data kept within a particular file.

3.3.1 Index Structure

The index provides a hierarchical structure to the important data in the data base. It has three levels:

(1) important machine condition data;

(2) diagnostic status data, reflecting the presence of data from a diagnostic;

(3) important data from diagnostic files.

The first level points to the file containing the original machine condition data, and the third level points to the file containing the original diagnostic data. The second level points to both files. This structure permits the user to devise retrieval requests such as:

(1) identify all files containing data acquired between April 10 and May 13, 1980;
(2) list all files acquired when 28-GHz power was applied;
(3) identify the files containing machine conditions prevalent when the temperature was over 250 eV;
(4) list the diagnostics active at 10 AM on May 27, 1980.

The responses to such requests comprise a list of file names that the user can then retrieve and analyze.

Additional features of System 1022 permit the retrieval, analysis, and reporting of data contained within the index structure without reference to the diagnostic or machine condition files.

3.3.2 Data Structure

There are two elements in the EBT data structure — the data file itself and a file (the descriptor file) that defines the format of the data file.

The data file

Both machine condition and diagnostic data files have the same format — a header area followed by a data area. The header area contains identifying data and, in the case of diagnostic data files, provides a reference to the file containing the machine conditions prevalent when the diagnostic was run. The data area contains the data acquired during one machine state (machine condition files) or during one run of a diagnostic.

Header. The header contains:

(1) header format identifier, which designates the format of the header itself;
(2) diagnostic identifier, which identifies the diagnostic which acquired the data or which states that this is a machine condition file;

(3) sequence number, which provides a unique identifier to each run of a diagnostic or to a machine condition;
(4) descriptor key, which identifies the descriptor file;
(5) startup parameters, which include the identifier of the file containing the machine conditions when this acquisition began and the start time and date;
(6) end parameters, which include the machine condition identifier, the time, and the date at the end of the diagnostic run.

These fields uniquely identify the data and provide information enabling correlation with all other data taken during the same machine conditions.

Data. The data portion of a file may include integer, ASCII, or single- and double-precision fields arranged to suit the convenience of the user. The fields may be arranged singly or in arrays with fixed, variable, or undefined bounds (delimited by end-of-file) and with any number of dimensions.

Descriptor file

When an experimenter is acquiring data, the information about the signal being acquired and the conditioning applied to that signal, etc., must be rigorously recorded, for if they are not, the meaning of the data will eventually be lost. When the data are analyzed, the descriptive information must be included to properly utilize the data.

When data are acquired and analyzed using computer programs, the above requirements are not relaxed. In fact, further requirements must be imposed, for with a computer program data are kept in a file, and the nature of the data (format, number of elements in arrays, etc.) must be recorded. If there are changes made to the experiment, such as changes to signal conditioning or number of data recorded, the analysis program must be able to process data files in both the old and new formats.

In the EBT system these requirements are met by the use of a descriptor file. This file contains the attributes of each datum or group of data in the data file it describes, including name, data type, number of points recorded, precision, scale factor, and descriptive comments. Each change in the data recorded by a diagnostic is accompanied by the creation of a new descriptor file which is kept in a special part of the data base with all other descriptor files ever produced. A printed record of the descriptor file is also produced to aid in documenting the system.

3.3.3 Maintenance

Given a data base into which data are constantly being placed, several functions must be performed:

(1) Descriptor files must be created and modified.
(2) Data must be added to the data base index.
(3) When the disk space becomes insufficient, some files must be moved to secondary storage.
(4) The data base must be backed up (a second copy maintained in case of catastrophe).

The creation and modification of file descriptors are performed by the BLDDFN program.[12]

The addition of data files to the data base proceeds automatically by means of a program which runs periodically. That program (named INDEX) can identify and incorporate files that have been received at the DECsystem 10 but have not yet been included in the data base.

The function of freeing disk space is performed by an archiving program, which identifies candidates for transfer to secondary storage and builds a command file to do so.

The backup of new data in the data base is carried out by the normal FED DECsystem 10 daily backup routine.

3.4 DATA RETRIEVAL

The first step in the retrieval of a particular datum from the data base is to identify the file(s) containing the datum. The user may already know the files of interest, or the index structure can be used to identify such files, as described earlier. Once the file or files have been identified, an analysis program can be run to acquire and manipulate the datum.

Two subroutines must be called by the analysis program. One, NEWFIL, finds and opens a file from which the datum is to be extracted. The file name provided to NEWFIL as an argument may be entered manually or acquired from a file produced by the index processing.

The second subroutine is GETDAT, which retrieves a datum (or a set of data) using its name. The descriptor file associated with the current data file provides the mapping between the name of the datum and its position in the data file. Note especially that neither the analysis program nor the GETDAT subroutine knows the structure of the data file — the descriptor file provides the structure. Accordingly, the analysis program can retrieve the datum from any file which contains a datum of that name. This feature makes the analysis program independent of changes to the experiment or to the data file and makes data acquired at any time or under any conditions equally accessible.

4. RELIABILITY

When EBT is down for any length of time, considerable money is wasted, and milestones may not be met. The computerized data acquisition and analysis system is so central to the operation of EBT that if that system is inoperable, then EBT is, effectively, also inoperable. As a result, a great deal of attention has been paid to the system's reliability.

4.1 RECOVERY FROM HARDWARE MALFUNCTION

Reliability is designed into this system at several levels. First, the system is easy to reconfigure if a computer should fail. Since there are two PDP-8 computers, both utilizing CAMAC for data acquisition, the functions performed by one can also be performed by the other by reconnecting the CAMAC system from the broken computer to the operational one.

The same degree of backup is present between the two PDP-11/34's. Furthermore, communications hardware is switchable between the two computers and can, therefore, be removed from a disabled computer and made available to the operational one. All software on either computer will run on the other.

Finally, reliability is achieved by the use of CAMAC modules which can be easily replaced.

4.2 RECOVERY FROM COMMUNICATIONS FAILURE

Since communications is so vital to this system, considerable care has been taken that data are not inadvertently lost due to malfunctions in the communications system. Protection against this eventuality is a feature of the FED network. In the event that some computer in the network becomes unavailable, data are transmitted to the most distant available intermediate computer and kept there for automatic transmission when the program is repaired. Enough disk storage capacity is available to permit several days' operation of EBT with no communication path to the DECsystem 10.

4.3 DATA BACKUP

If some catastrophe were to strike and disk files were to be destroyed, they should be recoverable. A convenient method of assuring recovery is to make a copy of all data on magnetic tape. The entire index is duplicated on magnetic tape every day. New data stored in the data base are copied to magnetic tape daily. A complete copy of a disk pack is made when it becomes full enough to warrant the use of a new pack. Later additions to the older pack are copied to tape monthly.

5. FLEXIBILITY

The system incorporates flexibility at several levels. First, the work in redesigning diagnostic software is reduced by the wide range of functions available in CAMAC modules. The revision of acquisition software is facilitated by the standard CAMAC subroutines, FORTRAN programming, the translation subroutines, and a standard question/answer/help routine named QUEST.[13] Changes in data analysis are readily accommodated by FORTRAN programming, by the numerical analysis and graphics libraries, and by the internal structure and intrinsic power of the data base. Finally, changes on one level (hardware, data acquisition software, or data analysis software) can frequently be made without changes on any of the other levels.

6. MAINTAINABILITY

Maintainability is an extremely important design goal of the EBT data acquisition and analysis system. We will discuss the maintainability of both hardware and software.

6.1 HARDWARE MAINTAINABILITY

The hardware involved in the system includes computers and data acquisition devices. Maintenance of the computers is eased by extensive use of a single manufacturer, which eases the requirements for training and which reduces the inventory of spare parts. Maintenance of the data acquisition hardware is made easier and more reliable by the use of CAMAC modules. If such a module should fail, we have a sufficient spares inventory to simply unplug the module and replace it.

6.2 SOFTWARE MAINTAINABILITY

A high level of software maintainability has been achieved through the following features:

(1) standard operating systems on all computers, thereby avoiding pitfalls resulting from local changes to the operating system;

(2) the use of standard software for critical functions, including data base management (System 1022), CAMAC data acquisition, plotting, and some communications (routines for these functions were already installed in FED and in other computer centers around the nation);

(3) strong emphasis upon documentation, both within programs and with external documents;

(4) use of FORTRAN wherever possible, greatly easing the burden upon the programmer who must modify or maintain a program.

7. SUMMARY

The major goal of EBT research is to evaluate the potential of EBT as a steady-state fusion reactor. This research requires accurate and rapid acquisition, storage, retrieval, and analysis of machine condition and diagnostic data. The EBT computer system was designed to provide these requirements in a flexible and straightforward manner.

Use of standard capabilities — CAMAC hardware and protocol, FORTRAN software, the System 1022 data base management system, and FED communications — has been employed throughout. The implemented system has fulfilled the design parameters of usefulness, reliability, flexibility, and maintainability. As such, it provides an exceptionally powerful, flexible, and useful tool for the EBT experimental organization.

REFERENCES

1. R. D. Burris, *Net — An Inter-Computer File Transfer Command*, ORNL/CSD/TM-56, Oak Ridge, Tennessee (May 1978).
2. R. D. Burris, *EBT Data Acquisition System Control Routines*, ORNL/CSD/TM-131, Oak Ridge, Tennessee (in preparation).
3. D. E. Greenwood, *CAMAC-Based Inter-Computer Communications System*, ORNL/TM-5678, Oak Ridge, Tennessee (in preparation).
4. The DISSPLA system is a proprietary graphics product of Integrated Software Systems Corporation (ISSCO), P.O. Box 9906, San Diego, California 92109.
5. The PLOT10 and Advanced Graphing-II packages are the property of Tektronic, Inc., Beaverton, Oregon.
6. System 1022 is a proprietary data base management product of Software House, Cambridge, Massachusetts.
7. The NAG library of numerical analysis subroutines is a product of NAG (USA), Inc., 1250 Grace Court, Downer's Grove, Illinois 60515.
8. The IMSL library of mathematical and statistical subroutines is a product of IMSL, Inc., 7500 Bellaire Boulevard, Houston, Texas 77036.
9. "IEEE Standard Modular Instrumentation and Digital Interface System (CAMAC)" (IEEE Std. 583-1975), the Institute of Electrical and Electronics Engineers, Inc., 1975.
10. C. R. Stewart, Jr., W. D. Joubert, D. R. Overbey, and K. A. Stewart, *TEK 11 Graphics User's Guide*, ORNL/TM-6546, Oak Ridge, Tennessee (October 1978).
11. R. D. Burris, *Data Format Translation Routines*, ORNL/TM-7462, Oak Ridge, Tennessee (in preparation).
12. J. S. Stanton, *EBT Data Base*, ORNL/TM-7463, Oak Ridge, Tennessee (in preparation).
13. D. E. Greenwood, *A Standard Operator Interaction Facility*, ORNL/CSD/TM-132, Oak Ridge, Tennessee (in preparation).