GA–A22922

REMOTE COLLABORATION AND DATA ACCESS AT THE DIII–D NATIONAL FUSION FACILITY

by
D.P. SCHISSEL for the DIII–D NATIONAL TEAM

SEPTEMBER 1998

DISCLAIMER

GA–A22922

REMOTE COLLABORATION AND DATA ACCESS AT THE DIII–D NATIONAL FUSION FACILITY

by
D.P. SCHISSEL for the DIII–D NATIONAL TEAM

This is a preprint of a paper to be presented at the Workshop on Remote Participation in Fusion Experiments, June 28–29, 1998, Prague, Czech Republic, and to be published in the *Proceedings.*

Work supported by
the U.S. Department of Energy
under Contract DE-AC03-89ER51114

GA PROJECT 3466
SEPTEMBER 1998

REMOTE COLLABORATION AND DATA ACCESS AT THE DIII–D NATIONAL FUSION FACILITY

D. P. Schissel for the DIII–D National Team

General Atomics, P.O. Box 85608, San Diego, CA 92186-9784

1. Introduction

When Doublet III was converted into the DIII–D tokamak in 1986, an expansion of collaborations was sought as a goal of the new DIII–D National Program. Today, the DIII–D National Team consists of about 120 operating staff and 100 research scientists drawn from 9 U.S. National Laboratories, 19 foreign laboratories, 16 universities, and 5 industrial partnerships. This multi–institution collaboration carries out the integrated DIII–D program mission which is to establish the scientific basis for the optimization of the tokamak approach to fusion energy production. Presently, about two–thirds of the research physics staff is from the national and international collaborating institutions.

As the number of on–site and remote collaborators has increased, the demands on the DIII–D National Program's computational infrastructure has become more severe. The Director of the DIII–D Program recognized the increased importance of computers in carrying out the DIII–D mission and in late 1997 formed the Data Analysis Programming Group. Utilizing both software and hardware improvements, this new group has been charged with increasing the DIII–D data analysis throughput and data retrieval rate. Understanding the importance of our remote collaborators, this group has developed a long term plan that will allow for fast 24 hour data access (7×24) with complete documentation and a set of data viewing and analysis tools that can be run either on the collaborators' or DIII–D's computer systems. This paper presents the group's long term plan and progress to date. A discussion on remote DIII–D tokamak operation is discussed in a separate paper at this workshop [1].

2. Data Storage and Retrieval

The amount of raw digitizer data acquired per tokamak pulse has been increasing steadily since the start of the DIII–D program. In FY96, 215 GB of raw compressed shot data were acquired with the largest single shot being acquired this year at 130 MB. Presently a total of 1 TB of raw data have been acquired. The increase in data acquired per year has been accompanied by an increased demand for data retrieval as the number of collaborators has grown. This in turn has created a substantial increase in network bandwidth.

A mass storage system has been purchased and installed this fiscal year to handle the large raw data requirements of the DIII–D program. This system consists of a 620 GB HP 600fx Magneto–optical jukebox with 238 platters and 6 drives and a 2.4 TB ATL 7000 DLT jukebox with 68 bays and 4 drives. These systems feed into a Sun Ultra–1 clone with 100 GB of Raid 5 magnetic hard drive storage providing over 3 TB of storage for DIII–D raw digitizer data. The mass storage system interactively provides 720 GB of raw data which is the equivalent of approximately 9000 compressed tokamak discharges. The remaining 2.4 TB from the DLT system is available on a 2 to 5 minute time scale. The hierarchical storage manager commercial software purchased from Veritas Software Corporation to operate the mass storage system provides 7×24 data availability which is demanded by a facility with collaborators in time zones that are up to 12 hours apart.

Network bandwidth into the DIII–D mass storage system is large enough to support remote collaborators. Off–site researchers connect into the DIII–D National Fusion Facility via the Energy Sciences Network (ESnet) utilizing a T3 ATM data line at 45 Mbits/s. The mass storage system is on an FDDI ring (100 Mbits/s) that connects directly to the ESnet.

Raw tokamak data is written into PTDATA format [2] and moved to the mass storage system at the end of day's operation. Data acquisition and data serving during tokamak operations is currently performed by an older MODCOMP computer system which is connected to the main DIII–D network via Ethernet (10 Mbits/s). Starting with January 1999 operations this system will be substantially upgraded. The MODCOMP computer will be replaced by a new DEC Personal Workstation 433au system running Digital Unix. This new computer will be connected to the main DIII–D network via FastEthernet (100 Mbits/s) utilizing an Alantec Power Hub series 8000 network switch. This enhanced data serving capability during tokamak operations will be approximately 100 times faster than current capabilities which will benefit the on–site researcher as well as the remote collaborator.

At the beginning of 1998 the Alcator C–Mod MDSplus [3] data storage software was adopted for on–site storage of analyzed DIII–D data. Prior to the MDSplus adoption, analyzed diagnostic data was written out in different formats and stored on a variety of computers. The DIII–D plasma diagnostic set is made up of more than fifty instruments placing a severe burden on a new collaborator to learn a wide variety of diagnostic file formats. The unified format of the MDSplus system allows a new researcher to learn a few computer commands and read a vast amount of DIII–D data. The diagnostic data stored in the MDSplus system continues to increase with presently 1700 archived shots representing 20 GB of analyzed data. MDSplus data serving is performed by a DEC AlphaServer 4000 5/300 running OpenVMS and connected to the DIII–D ESnet network via an FDDI ring. The majority of MDSplus data is made available in between tokamak pulses giving the researcher critical information before setting up the next pulse. Access to the MDSplus data by a remote collaborator is made easy by using a TCP/IP sockets interface. A multi–institutional team consisting of members from General Atomics, LLNL, MIT, and PPPL are porting the MDSplus server software to the Unix environment.

The DIII–D relational database of analyzed physics results was designed over ten years ago and operates under the commercial software package S1032 on a DEC Vax 4000 Model 600A computer system. A replacement commercial database is targeted for fiscal year 1999 and is presently being investigated with input from our MIT and PPPL collaborators. This new system will operate in the Unix environment with access available from either a Web interface or customized GUI data mining tools. The planned interface and documentation will allow easy access for remote collaborators.

3. Data Viewing and Analysis Tools

Work on data viewing and manipulation tools has focused on efficient graphical user interface (GUI) design with object oriented programming for maximum code flexibility and access to both PTDATA and MDSplus data. The GUI design combined with a thorough documentation system (described below) decreases the non–productive time a new collaborator must spend learning a new system. Tools are being written in IDL, a commercial software package for scientific data manipulation and visualization.

A new object oriented IDL based direct graphics library GaPlotObj has been created in collaboration with Fanning Software Consulting. This graphics library is a fundamental component of the new DIII–D viewing tools and allows multiple graphs with cursors for data readout, zooming, panning, and data selection for manipulation. The first package to use this graphics library is the new EfitViewer which allows a researcher to examine a DIII–D magnetic equilibrium, plasma profiles calculated during the equilibrium reconstruction, and the quality of the fit including the magnetics and MSE data. The second tool to use GaPlotObj is REVIEWplus which is presently in beta testing. This package allows general 2D and 3D data plotting with arithmetic data manipulation. Researchers typically plot time histories and plasma profiles with this tool. These two new codes are gradually replacing the functionality of the 4D analysis tool that was the DIII–D program’s first attempt at creating a GUI analysis package. The profile analysis part of 4D will be retained but with an enhanced GUI for ease of use.

Existing data analysis capabilities have been streamlined for greater efficiency and ease of use by a new collaborator. A kinetic EFIT typically took half a day but with recent streamlining in the 4D profile analysis section this time has been reduced to less than one hour. Furthermore, a researcher does not have to know a multitude of codes and Unix shell scripts but instead only needs to know the one 4D GUI. A new interactive EFIT GUI is being created for rapid MHD equilibrium reconstruction. This type of interactive fit is only possible because of the dramatic increase in computational power that is available in the workstation environment. Future plans include storing the output results of theory/model analysis codes (e.g. ONETWO, GLF23, UEDGE) in MDSplus. Such storage will make it easier to compare experimental data with computer simulation, allow for easy plotting of simulation data in existing viewing tools, and provide a long term centralized storage mechanism. Where appropriate, a GUI will be constructed which will make an analysis tool easier to use.

Accompanying these new analysis tools is a Web based (http://fusion.gat.com) documentation system that brings critical knowledge to the on–site and remote collaborator. The documentation focuses on computer codes and tokamak data including PTDATA and MDSplus. Access to the data documentation system directly from an analysis tool has been demonstrated and is presently being implemented.

The new IDL based viewing and manipulation tools are being distributed to remote collaborators either in the form of a compiled binary executable or from a source code management system (CVS). Computer code management allows the interested researchers to modify existing tools and merge their changes back into the main repository. Creating tools in IDL has the added benefit of being able to move among different operating systems with minor modifications. These tools are presently being run on HP–UX, Digital Unix and the MacOS platforms. The recent port of the MDSplus IP library to the MacOS allows all MDSplus stored data to be retrieved directly to the Macintosh for viewing and manipulation. This allows a researcher who is visiting another laboratory or attending a workshop to bring a PowerBook G3 computer, establish an IP connection, and perform data analysis as if they never left their home laboratory. A drawback to utilizing IDL is that remote collaborators are forced to purchase a license in order to operate these DIII–D software packages. The Data Analysis Group has investigated writing tools in Java and Python but at the present time the IDL language is a much richer environment for scientific applications. As these languages evolve they will continue to be evaluated as a potential replacement for IDL.

4. Computer Power for Data Analysis

The backbone of data analysis at DIII–D is performed on an HP 9000 Model T–600 server wth 3 processors running HP–UX 10.2. Remote collaborators typically receive an account on this server with the total number of user accounts now exceeding 300. Complementing this system are numerous HP and DEC Unix workstations scattered throughout the DIII–D site with two HP workstations dedicated to data analysis during tokamak operations. Equilibrium reconstructions (EFIT) are performed in between tokamak pulses in a distributed computing environment consisting of workstations at General Atomics and LLNL. This system allows for two types of EFIT reconstructions to be run in between every tokamak pulse with a time resolution of 25 msec.

In order to make the most efficient use of computational resources the LSF distributed load sharing and job scheduling commercial software by Platform Computing Corporation is currently under review. The concept under review is that this software will operate in a heterogeneous computational environment potentially combining all Unix analysis workstations into one large CPU pie. This arrangement will allow the DIII–D program to keep pace with the rapid advancements of the computer industry since older pieces of the pie can be replaced by newer computers. A researcher will receive one account, as opposed to an account per workstation under the current system, greatly simplifying the appearance of the computational environment. Additionally, a remote or on–site collaborator can add their own computer to the CPU pie without the need for the computer to reside at the DIII–D facility.

5. Summary

A long term plan has been formulated and is being implemented that will allow remote collaborators easy and fast access to both raw and analyzed data from the DIII–D tokamak. GUI designed viewing tools with ample documentation will guide the researcher through their data mining experience. Future work will include storing the results of complex modeling codes, often run off–site, alongside DIII–D experimental data for easy comparison. Running integrated modeling and simulation codes in a massively parallel computing environment coupled with new visualization tools that will be required to handle the order of magnitude increase in data is presently being investigated.

Acknowledgements

The Data Analysis Programming Group (DAPG) led by D. P. Schissel consists of C. Campo, Q. Peng, and J.S. Schachter from General Atomics, and B. Davis and T. Terpstra from PPPL. The work of K. Keith, B. McHarg, and C. Parker from General Atomics has been critical in developing and maintaining the computer hardware and PTDATA infastructure at the DIII–D National Fusion Facility. The DAPG acknowledges stimulating discussions with T. Casper, B. Meyer, and J. Moller from LLNL, and T. Fredian, M. Greenwald, and J. Stillerman from Alcator C–Mod. This is a report of work supported by the U.S. Department of Energy under Contract No. DE-AC03-89ER51114.

References

[1] T.A. Casper, this workshop; B. B. McHarg, J. of Fus. Eng. and Design, to be published.
[2] B.B. McHarg, Proceedings of the 15th IEEE Symposium on Fusion Engineering (1994).
[3] J. Stillerman, et al., Rev. Sci. Instrum. 68, 939 (1997).