GA–A23471

RECENT ENHANCEMENTS TO ANALYZED DATA ACQUISITION AND REMOTE PARTICIPATION AT THE DIII–D NATIONAL FUSION FACILITY

by
D.P. SCHISSEL, J. BURTRUSS, Q. PENG, J. SCHACHTER,
T. TERPSTRA, K.K. KEITH, B.B. McHARG, JR., and J.C. PHILLIPS

OCTOBER 2000

DISCLAIMER

GA–A23471

RECENT ENHANCEMENTS TO ANALYZED DATA ACQUISITION AND REMOTE PARTICIPATION AT THE DIII–D NATIONAL FUSION FACILITY

by

D.P. SCHISSEL, J. BURTRUSS, Q. PENG, J. SCHACHTER, T. TERPSTRA,† K.K. KEITH, B.B. McHARG, JR., and J.C. PHILLIPS

†Princeton Plasma Physics Laboratory, Princeton, New Jersey

This is a preprint of a paper presented at the 21st Symposium on Fusion Technology, September 11–15, 2000 in Madrid, Spain and to be published in *Fusion Design and Engineering.*

Work supported by
the U.S. Department of Energy under
Contract Nos. DE-AC03-99ER54463 and DE-AC02-76CH03073

GA PROJECT 30033
OCTOBER 2000

ABSTRACT

A long–term plan is being implemented to enhance the computational infrastructure of the DIII–D National Fusion Facility. One of the goals of this plan is more efficient utilization of DIII–D experimental run–time by decreasing the time to analyze, store and distribute analyzed data during tokamak operations. A multi–processor Linux cluster is reducing data analysis time and a Unix based MDSplus data management system is providing rapid access to analyzed data. A second goal of the long–term plan is to reduce the time required for more detailed physics analysis after experimental operations. This goal is being accomplished with the underlying philosophy of uniformity, both in the look and feel of our own GUI–based tools, in terms of access methods to analyzed datasets, and access to existing computer power via a load balanced UNIX cluster. Additionally, we have enhanced our remote meeting capability resulting in improved communication within the geographically diverse DIII–D Research Team.

1. ANALYZED TOKAMAK DATA ACQUISITION, STORAGE, AND RETRIEVAL

The DIII–D tokamak operates in a pulsed mode producing plasmas of 5–10 s duration every 10–15 min, with 25–35 pulses per operating day. One way to more efficiently utilize experimental run time is to place more analyzed data between pulses at the fingertips of the research team. This can be accomplished by performing between pulse data analysis more rapidly and by faster serving of analyzed data.

At the beginning of 1998 the MDSplus [1] data storage software was adopted for on–site storage of analyzed DIII–D data [2]. Prior to MDSplus adoption, analyzed diagnostic data was written out in different formats and stored on a variety of computers (Fig. 1). The DIII–D plasma diagnostic set is made up of more than sixty instruments placing a severe burden on the research team to learn a wide variety of analyzed data file formats. Additionally, this lead to a variety of data retrieval rates due to different CPU speeds and network connections among the different computers often resulting in data only being available just prior to the next pulse. The adoption of MDSplus has numerous advantages: 1) a unified format for analyzed data (Fig. 1) requiring only a few commands to read a vast amount of data, 2) rapid data retrieval from a central server, 3) compression to save disk space, 4) only a handful of files are necessary to store the results from an entire pulse, and 5) storage of multiple versions of analysis.

The DIII–D analyzed diagnostic data stored in the MDSplus system continues to increase with presently 8900 archived pulses representing 70 GB of data. Unix MDSplus data service was chosen at DIII–D with the benefits of faster data service and easier integration into the DIII–D Unix analysis environment. MDSplus data is served from a Compaq DEC AlphaServer 800 5/500 with 1 GB of RAM running Compaq Tru64 Unix, with an nSTOR GigaRAID AA system containing 284 GB. This system provides a flexible upgrade path for future expansion in both CPU power and magnetic storage space.

Presently, up to 21 separate analyzed datasets are present for each pulse, including EFIT magnetic equilibrium results, kinetic T_e, T_i, n_e, V_r, and n_i profiles from the Thomson scattering and CER diagnostics, and spectroscopy and divertor imaging results. There are usually 21 MB of analyzed data stored per pulse. As more diagnostic data are added, this number will continue to increase to an anticipated 100 MB per pulse. During experiment operation, most of the data are loaded automatically between plasma pulses, giving the researcher critical information before setting up the next pulse. Between pulse analysis jobs and data storage are managed by the MDSplus dispatching system based on events declared when the requisite raw data are available

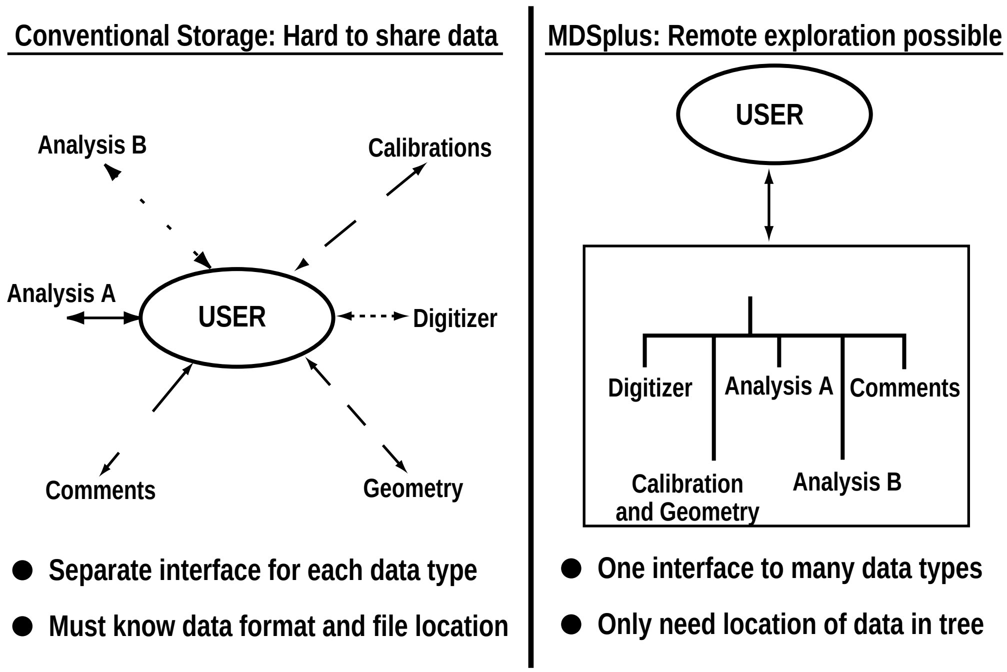

Fig. 1. MDSplus simplifies the interface to analyzed data making it easier for the DIII–D National Team to explore.

after acquisition. MDSplus loading can occur at any time after a plasma pulse so that the results of more detailed analyses can be stored in the days following an experiment.

A new DIII–D relational database, D3DRDB that uses Microsoft SQL Server 7 software, works in concert with MDSplus by storing highlights of these data repositories. Users can rapidly search through the data highlights stored in the relational database and find the subset of pulses that have special interest. Additionally, experimental summary information has been added to D3DRDB including daily text summaries, key experimental personnel, and electronic logbook comments. The electronic logbook, adopted from C–Mod, enhances DIII–D experimental operation through its many advantages over a traditional paper based system. The two primary advantages are the ability to see comments from others in real time and the ability to rapidly query past entries. The electronic logbook allows the DIII–D National Team to monitor, from anywhere, the pulse–to–pulse progress of the experiment.

2. CALCULATION OF ANALYZED TOKAMAK DATA

Data analysis between pulses is accomplished on either a load balanced Unix cluster or on a multi–processor Linux cluster. These analysis jobs can either be started by the tokamak data acquisition cycle or by an individual researcher. Results of these analyses are stored in MDSPlus and made available to the entire research team.

General computer power for data analysis at DIII–D is provided by an HP 9000 Model T–600 3 processor server and a load balanced cluster of six HP and five Digital Unix workstations. Interactive computer load balancing is accomplished with LSF Suite 3.2 [3]. This software operates in a heterogeneous computational environment thereby combining all of the newer Unix based computers into one CPU cluster. The benefit of such a cluster is that all computers are easily and transparently available to all researchers, that CPU upgrades are as simple as removing one workstation and adding another, and that a new on–site collaborator can easily add their own computer to the CPU cluster. This cluster has increased the Unix computer power available to the research team by over a factor of ten and has greatly increased the amount of interactive between pulse data analysis. Additionally, LSF Suite has reduced operating costs since it manages cluster–wide commercial software licenses thereby requiring fewer total licenses.

EFIT magnetic equilibrium reconstructions are performed in between tokamak pulses on a new 12 processor Pentium III 256 MB PC Linux cluster [4]. This work required modifying EFIT to use the parallel message passing interface (MPI) library so that multiple independent equilibria can be rapidly generated from experimental data. This new system calculates equilibria eight times faster than the previous system yielding a complete equilibrium time–history on a 25 ms time–scale two minutes after a pulse ends (Fig. 2). Additionally, the PC based hardware for this increased analysis capability costs significantly less than a comparable HP or Compaq Unix workstation solution. A new interface has also been created so that the physics operators and session leaders can directly control the EFIT setup including the time resolution and the input parameter file being used. Users may run their own EFITs on the Linux cluster when the tokamak is not operating. Some users are performing statistical analysis requiring on the order of 100,000 EFIT equilibria. Problems of this magnitude could not be solved without the Linux cluster. Future uses of the cluster include an Interactive Data Language (IDL) [5] based application in beta release that calculates the time history of a complete T_i profile within four minutes of the pulse termination. Presently, this analysis is only able to be completed overnight using the LSF cluster.

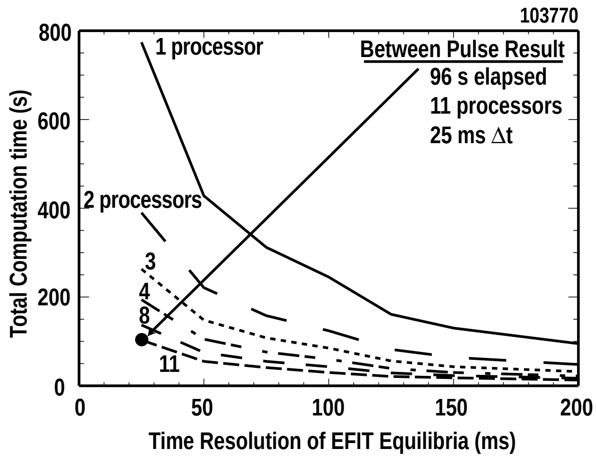

Fig. 2. Total computation time versus the time resolution of EFIT equilibria calculation for different number of processors. All analysis covers the 6 s length of pulse 103770 which for the 25 ms resolution case requires 240 equilibria. The data point represents the actual result for between pulse calculation during experimental operation. The EFIT equilibria are calculated for a 65×65 grid size at double precision.

These between–pulse EFITs have been integrated into the Fault Identification and Communication System (FICS) for rapid error detection during experimental operations [6]. This system has been used routinely during DIII–D operations and has led to an increase in tokamak productivity. In the near term over a dozen between pulse data analysis codes will be added to FICS allowing diagnostic consistency checks during operations resulting in rapid problem identification and repair. This is to be contrasted with the present situation where some inconsistencies are found only after experimental operations.

3. SUPPORT FOR REMOTE PARTICIPATION

The DIII–D National Team consists of about 120 operating staff and 100 research scientists from 9 U.S. National Laboratories, 19 foreign laboratories, 16 universities, and 5 industrial partnerships. Presently, about two–thirds of the research physics staff are from the national and international collaborating insituitions. The DIII–D National Team is therefore geographically diverse which has placed an increased demand on improving communication and data access to off–site researchers.

To alleviate the ever increasing load placed on our CPU resources by off–site collaborators our analysis environment encourages usage of off–site computers. Such analysis is simplified by the availability of raw and analyzed DIII–D data via the MDSplus client/server interface. Utilization of off–site hosts benefits both the remote collaborator and the DIII–D computing infrastructure. Collaborators would experience faster response from their operations on the data, as all interaction with the window management system is local, and the DIII–D network and computers would be less burdened. IDL based viewing and manipulation tools are being distributed to remote collaborators. These tools have presently been installed at C–Mod, NSTX, SSPX, and JET, as well as DIII–D.

ReviewPlus is one of the tools currently in use in the U.S. fusion community that is already capable of viewing data from multiple sites simultaneously. For each signal of interest, a user can specify the IP address of the MDSplus server from which to retrieve data. Since nothing in the code is specific to DIII–D, it can be easily setup and used at any remote site with MDSplus and an IDL license. Users familiar with ReviewPlus therefore do not need to learn how to use other tools to view data from other sites.

Another aspect of remote data analysis is the ability to hold meetings to discuss on–going analysis. The DIII–D National Team has recently improved off–site communication by upgrading the remote audio/video hardware. At the DIII–D facility, the current capability includes two conference rooms near the staff offices that have been equipped to share a ShowStation IP from Polycom [7]. This device acts as a viewgraph machine for the researcher in the conference room and a Web server for those not in the conference room. The off–site collaborator may see the viewgraphs simply by using a Web browser or by means of a remote ShowStation if their conference room is so equipped. The off–site collaborator may also become a presenter in such a conference, showing their viewgraphs to the larger audience.

Video conferencing is presently handled with a ViewStation MP from Polycom [7] in both conference rooms. Broadcast is accomplished over ISDN at 384 kbits/s or over 10 baseT

Ethernet (10 Mbits/s). Mutli–point ISDN video conferences are handled by an ESNET bridge allowing an unlimited number of remote sites to conduct a meeting. Therefore, a complete working meeting can involve multiple sites participating on the ViewStation, additionally sharing viewgraphs with the ShowStation.

The DIII–D tokamak is physically located approximately 1.5 miles from the National Team offices. A conference room, located next to the DIII–D control room, is similarly equipped as the conference rooms described above. However, DIII–D's conference room and control room have the ability to broadcast audio and video over the web continuously [8]. The morning meeting, held prior to operations every experimental day, is viewable using a web browser. At the conclusion of the meeting the web broadcast is switched to the control room so that the days experimental operation can be monitored. To view, a client need only have a Netscape browser with Java enabled (to receive the audio stream). A diverse client base is supportable because the broadcast does not require any plug–in software like RealPlayer or Shockwave. The web broadcast is available off–site without restriction, allowing the worldwide DIII-D National Team active participation in any DIII–D experimental campaign.

REFERENCES

[1] J. Stillerman, *et al*., Rev. Sci. Instrum. **68**, 939 (1997).

[2] J. Schachter, "Data Analysis Software Tools for Enhanced Collaboration at the DIII–D National Fusion Facility," Proc. of the 2nd IAEA Tech. Com. Mtg. On Control, Data Acquisition and Remote Participation on Fusion Research, to be published in Fusion Eng. And Design (1999).

[3] Platform Computing Corporation (http://www.platform.com)

[4] Q. Peng, et al., "A Linux PC Cluster for Between-Pulse EFIT and Other CPU Bound Analyses at DIII–D," General Atomics Report GA-A23473 (2000).

[5] Research Systems, Inc. (http://www.rsinc.com)

[6] M.L. Walker, *et al.*, "Automated Fault Detection for DIII–D Tokamak Experiments," to be published in the proc. of the 18th IEEE/NPSS Symposium on Fusion Engineering (1999).

[7] Polycom, Inc. (http://www.polycom.com).

[8] http://d3dpc1.gat.com/computing/NFF/meet.html.

ACKNOWLEDGMENT

This is a report of work supported by the U.S. Department of Energy under Contracts No. DE-AC03-99ER54463 and DE-AC02-76CH03073.