

Enhanced Computational Infrastructure for Data Analysis at the DIII-D National Fusion Facility

D.P. Schissel, Q. Peng, J. Schachter, T.B. Tepstra,¹ T.A. Casper,² J. Freeman,
R. Jong,² K.M. Keith, B.B. McHarg, Jr., W.H. Meyer,² C.T. Parker,
and T. Warner

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

¹*Princeton Plasma Physics Laboratory, Princeton, New Jersey.*

²*Lawrence Livermore National Laboratory, Livermore, California.*

1. Introduction

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 are from the national and international collaborating institutions.

As the number of on-site and remote collaborators continues to increase, the demands on the DIII-D National Program's computational infrastructure become more severe requiring enhancing the computational infrastructure to satisfy this increasing demand. Work has concentrated on both software and hardware improvements to increase the DIII-D data analysis throughput and data retrieval rate. The underlying philosophy (Fig. 1) behind these development efforts is uniformity, both in terms of the look and feel of graphical user interfaces (GUIs), in terms of access methods to analyzed datasets, and access to existing computer power. This paper presents the enhancement plan and progress to date.

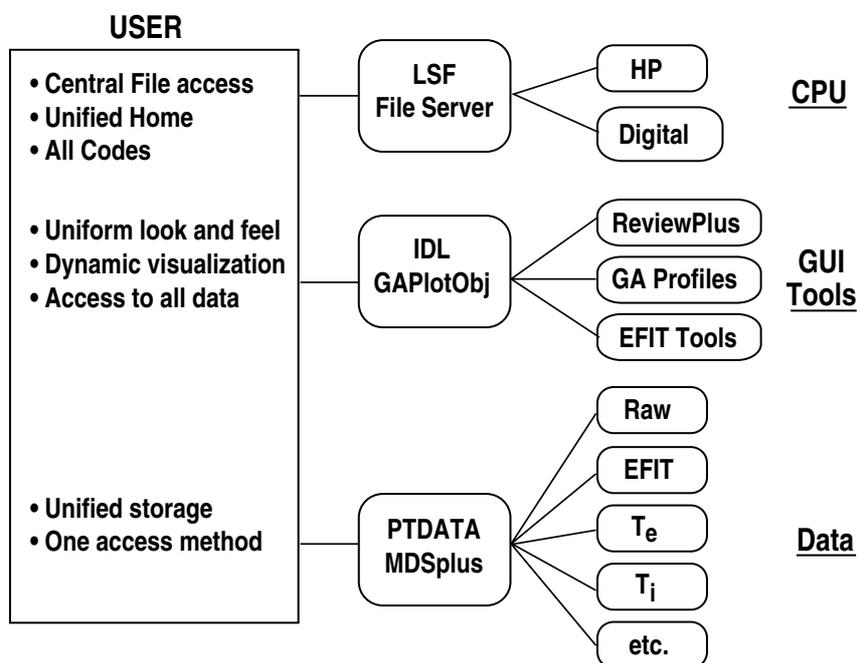


Fig. 1. Users benefit from the unification of a heterogeneous environment.

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. Presently a total of 1 TB of raw data, written into PTDATA format [1], has been acquired. A mass storage system with a total capacity of 3 TB has been installed for storage of DIII-D raw digitizer data. Along with 100 GB of hard disk capacity, this system consists of a 620 GB HP 600fx Magneto-optical jukebox and a 2.4 TB ATL 7000 DLT tape library. The mass storage system provides 7 day-24 hour data availability which is demanded by a facility with collaborators in time zones that are up to 12 hours apart. The magnetic disk and Magneto-optical components interactively provide 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 3 to 5 minute time scale

At the beginning of 1998 the MDSplus [2] data storage software was adopted for on-site storage of analyzed DIII-D data. Prior to 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 sixty instruments placing a severe burden on a new collaborator to learn a wide variety of analyzed data file formats. The unified format of the MDSplus system allows researchers 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 4730 archived shots representing 38 GB of analyzed data. Data serving by MDSplus is presently performed by a DEC AlphaServer 4000 5/300 running OpenVMS. The majority of MDSplus data is made available in between tokamak pulses giving the researcher critical information before setting up the next pulse. A multi-institutional team consisting of members from DIII-D, LLNL, MIT, and PPPL has recently completed porting the MDSplus software to the Unix environment. This Unix based MDSplus system has recently completed testing at DIII-D and will replace the existing OpenVMS version in the summer of 1999. The benefits of Unix MDSplus are faster serving of data and easier integration into the DIII-D Unix analysis environment. Future storage plans include the integration into MDSplus of data from advanced simulation and modeling codes allowing for easy comparison to experimental data with existing visualization tools.

The increase in data acquired per year has been accompanied by an increased demand for data retrieval as the size of the DIII-D National Team has grown. This in turn has created a substantial increase in network bandwidth. To satisfy this demand the network within the DIII-D control room and to the offices was increased to switched 100 Mbits/s at the start of this calendar year.

3. Data Viewing and Analysis Tools

Work on data viewing and manipulation tools has focused on efficient and uniform GUI design with object oriented programming for maximum code flexibility and access to both PTDATA and MDSplus data. The uniform GUI design combined with a thorough documentation system (described below) decreases the non-productive time a new researcher must spend learning a new system. Alternatively, existing users do not need to switch gears every time they switch analysis tools. GUI tools are being written in Interactive Data language (IDL) [3], a commercial product 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 providing a uniform GUI for data manipulation. The GaPlotObj graphics library allows multiple graphs with cursors for data readout, zooming, panning, and data selection for manipulation.

Two main viewing and analysis tools, EFITTools [4] and ReviewPlus [5], have been created that use the GaPlotObj graphics library. EFITTools combines the ability to perform

an interactive EFIT, a kinetic EFIT, a time dependent EFIT, and to view any EFIT results under one GUI umbrella. Visualizing EFIT results (EfitViewer) allows a researcher to examine a DIII–D magnetic equilibrium, the plasma profiles calculated during the equilibrium reconstruction, and the quality of the fit including the magnetics and MSE data in a more efficient method than was previously possible. The interactive EFIT capability dramatically reduces the time required to fine tune an MHD equilibrium reconstruction from hours to minutes by eliminating the unnecessary repeated initialization of the code. A kinetic EFIT typically took half a day but with streamlining provided in the kinetic EFIT GUI this time has been reduced to less than one hour.

The ReviewPlus tool is a general purpose data visualization program that provides interactive 2D and 3D graphs of data stored in either PTDATA or MDSplus. Researchers typically plot time histories, plasma profiles, or the temporal evolution of plasma profiles with this tool. As more analyzed data is stored in MDSplus it automatically becomes available for visualization in ReviewPlus. Besides adding more experimental data to the MDSplus archive, output results from theory/modeling analysis codes (e.g. ONETWO, GLF23, UEDGE) will also be added. Such storage will make it easier to compare experimental data with computer simulations.

The GAProfiles analysis tool has been in existence for several years at DIII–D and has not been ported over to use the new GAPlotObj graphics library. This tool performs interactive bi-cubic spline fitting of experimentally measured plasma profiles including T_e , T_i , n_e , V_r , and P_{rad} . These fitted profiles are used as inputs into energy transport and MHD stability codes.

Accompanying these new analysis tools is a Web based documentation system (<http://fusion.gat.com>) that brings critical knowledge to the on–site and remote collaborator. The documentation focuses on computer codes and tokamak data including PTDATA and MDSplus. This web based data documentation system is directly available from all analysis tools.

4. Computer Power for Data Analysis

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 three HP and five Digital Unix workstations. Interactive computer load balancing is accomplished with LSF Suite 3.2 [6]. 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. Such an implementation has been possible because of the fast network connecting the workstations, the central file server that is available from all workstations, and the unified data access methodology. This Unix cluster has more than tripled the Unix computer power available to the researcher.

Equilibrium reconstructions (EFIT) are performed in between tokamak pulses in a distributed computing environment [7] 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 ms.

The recent port of the MDSplus IP library to the MacOS allows all MDSplus and PTDATA stored data to be retrieved directly to the Macintosh for viewing and manipulation. Presently, both EfitViewer and ReviewPlus run on the MacOS with IDL. Usage of our existing G3 Macintosh computers for more intense IDL based data analysis and visualization is being investigated; this has the potential to greatly expand existing CPU power. Additionally, this capability 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 were in their office.

5. Support of Off-Site Data Analysis

Remote or temporarily on-site collaborators typically receive an account on the HP T-600 server with the total number of user accounts now exceeding 300. 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. Additionally, the new IDL based viewing and manipulation tools are being distributed to remote collaborators either in the form of a compiled binary executables 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 have presently been installed at C-Mod, NSTX, and JET.

Another aspect of remote data analysis is the ability to hold meetings to discuss on-going analysis. The DOE/OFES has facilitated this aspect by providing resources to aid the remote audio/video capability of our collaborators. This includes the use of video-conferencing, full-duplex audio-conferencing, and ShowStation Ip [8] to broadcast viewgraphs via a web browser.

6. Summary

A long term plan has been formulated and is being implemented to increase the DIII-D data analysis throughput and data retrieval rate. The basic component of this plan is to create a uniform interface to a very heterogenous environment. 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.

Acknowledgments

This is a report of work supported by the U.S. Department of Energy under Contract Nos. DE-AC03-99ER54463, DE-AC02-76CH03073, and W-7405-ENG-48. The authors acknowledge stimulating discussions T. Fredian, M. Greenwald, and J. Stillerman from Alcator C-Mod at MIT.

References

- [1] B.B. McHarg, Proc. of the 15th IEEE/NPSS Symposium on Fusion Engineering, Hyannis, Massachusetts, Vol. 1 (Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 1994) p. 123.
- [2] J. Stillerman, et al., Rev. Sci. Instrum. **68**, 939 (1997).
- [3] Interactive Data Language, Research Systems, Inc. (<http://www.risinc.com>).
- [4] Q. Peng, "IEFIT — An Interactive Approach to High-Temperature Fusion Plasma Magnetic Equilibrium Fitting," to be presented at IEEE Real Time Conf., Santa Fe, New Mexico.
- [5] J. Schachter, "Data Analysis Software Tools for Enhanced Collaboration at the DIII-D National Fusion Facility," to be presented at 2nd IAEA TCM on Control, Data Acquisition and Remote Participation on Fusion Research, Lisboa, Portugal.
- [6] Platform Computing Corporation (<http://www.platform.com>).
- [7] T.A. Casper, et al., Computers in Physics **12** (1998) 220.
- [8] Polycom, Inc. (<http://www.polycom.com>).