

GA-A23187

**DATA ANALYSIS SOFTWARE TOOLS FOR
ENHANCED COLLABORATION AT THE
DIII-D NATIONAL FUSION FACILITY**

by
J. SCHACHTER, Q. PENG, and D.P. SCHISSEL

JULY 1999

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

**DATA ANALYSIS SOFTWARE TOOLS FOR
ENHANCED COLLABORATION AT THE
DIII-D NATIONAL FUSION FACILITY**

by
J. SCHACHTER, Q. PENG, and D.P. SCHISSEL

This is a preprint of a paper to be presented at the 2nd IAEA Technical Committee Meeting on Control, Data Acquisition and Remote Participation on Fusion Research, July 19–21, 1999, in Lisboa, Portugal, and to be published in *Fusion Engineering and Design*.

**Work supported by
the U.S. Department of Energy
under Contract No. DE-AC03-99ER54463**

**GA PROJECT 30033
JULY 1999**

ABSTRACT

Data analysis at the DIII-D National Fusion Facility is simplified by the use of two software packages in analysis codes. The first is “GAPlotObj,” an IDL-based object-oriented library used in visualization tools for dynamic plotting. GAPlotObj gives users the ability to manipulate graphs directly through mouse and keyboard-driven commands. The second software package is “MDSplus,” which is used at DIII-D as a central repository for analyzed data. GAPlotObj and MDSplus reduce the effort required for a collaborator to become familiar with the DIII-D analysis environment by providing uniform interfaces for data display and retrieval. Two visualization tools at DIII-D that benefit from them are “ReviewPlus” and “EFITviewer.” ReviewPlus is capable of displaying interactive 2D and 3D graphs of raw, analyzed, and simulation code data. EFITviewer is used to display results from the EFIT analysis code together with kinetic profiles and machine geometry. Both bring new possibilities for data exploration to the user, and are able to plot data from any fusion research site with an MDSplus data server.

1. INTRODUCTION

The DIII-D National Fusion Facility hosts 300 researchers, most from 54 collaborating institutions around the globe. Support for collaborating scientists, both on and off-site, is therefore fundamental to the development of data analysis software tools at DIII-D. All scientists at DIII-D have benefitted from ongoing improvements to these software tools, their underlying data management scheme, and the computing hardware infrastructure on which they run. For collaborators in particular, the improvements enhance their ability to perform data analysis either at DIII-D or remotely from their home institution, by making the process simpler, more efficient, and more productive, and by introducing new possibilities for data exploration. This paper focuses on the data analysis software tools in use at DIII-D, specifically the tools used for data visualization, and on the underlying analyzed data management system. For a broader picture of the computational environment at DIII-D [1,2].

Visualization tools at DIII-D are written in the Interactive Data Language (IDL), and employ an in-house developed object-oriented graphics library, GAPlotObj, as a uniform graphical user interface (GUI) for interactive data display. Access to DIII-D data, from visualization tools and other codes, has also been made uniform through the adoption of the MDSplus system [3] as a centralized repository for analysis results. The uniformity of data interaction and access methods simplifies the data analysis process. In addition, both the GAPlotObj library and MDSplus bring new capabilities to the DIII-D environment. These two software packages are discussed in Sections 2 and 3 respectively.

Two data visualization tools employing the GAPlotObj library and MDSplus are described in Section 4. The first, ReviewPlus, is a general purpose plotting tool that provides interactive 2D and 3D graphs of raw, analyzed and simulation data. The second, EFITviewer, is used for examining the results from the EFIT equilibrium fitting code. Part of the effort to support collaborators at DIII-D has been to allow scientists to run these tools from their own site, connecting via the internet to the DIII-D site only to retrieve data. Of greater benefit would be

the ability to use the same tool to access data from multiple institutions. Section 5 discusses remote operation of ReviewPlus and EFITviewer, and the issues involved in making them site-independent.

2. GAPlotObj

IDL, from Research Systems Incorporated [4], is the language of choice for data visualization applications at DIII-D, as it combines a rich, easy-to-use language with extensive and powerful interactive graphics, and can run on a variety of CPU architectures with minimal porting. One of the strengths of IDL is its ability to couple with the window management system to provide dynamic plotting; that is, the ability for the user to change how the data is displayed, and even the data themselves, with a mouse-driven, “point and click” interface.

Dynamic plotting must be included in an application by the developer; it is not automatically available in IDL. Rather than coding the same interface over again in each application, the Data Analysis Applications Group at DIII-D, together with Fanning Consulting, developed an object-oriented library, GAPlotObj, that can be used in all applications for dynamic plotting. From the user's point of view, this means that in every application the same action, either a keystroke or a mouse click, gives the same result. Users therefore need to learn the interface only once, allowing newly arrived collaborators to quickly become productive in their analysis of DIII-D data, and saving all users the need to switch gears each time they switch tools.

An object-oriented approach benefits code developers as well by encapsulating the properties and behaviors of the plotting functions so that they can be easily extended, reused, and maintained independently of the applications that employ them. While the GAPlotObj library is object-oriented, it should not be confused with IDL's built-in Object Graphics library. This latter library is a set of classes that use OpenGL for rendered graphics display, whereas the GAPlotObj library uses IDL's “direct graphics” for display.

In all applications using the GAPlotObj library, the user is presented with a window containing one or more plots, as shown in Fig. 1. Each plot contains one or more data traces. Dynamic plotting is controlled by actions of the mouse directly on the plotting window, and by configuration dialogs. The former allow the user to manipulate the plots directly, giving them rapid feedback on their actions and saving them from changing their focus to another window.

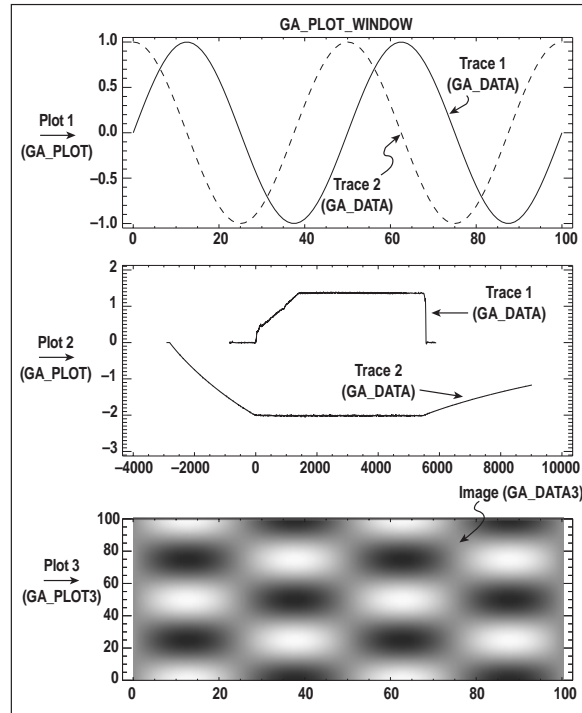


Fig. 1. Display and class organization of GAPlotObj Library.

Several types of actions are possible, more than the available number of mouse buttons (three on a typical system). Therefore, the actions are divided into “mouse modes” that group like functions together. The application can be in only one of the following five mouse modes at any given time: select, zoom, edit, cursor, or slice. Each is described in turn.

Select mode allows the user to choose a plot or data trace for manipulation, copying, or deletion by clicking near the item of interest. Once selected, the appearance of a data trace can be altered, or mathematical operations can be performed on the data. Clicking and dragging in zoom mode changes the ranges of the plot displayed to the region defined by the mouse. If many plots are in the window, the user can toggle to a single plot display by clicking near the plot of interest. Edit mode allows the user to add or remove data points within a trace; note that this is different than adding or removing entire traces from the plot. This feature is disabled in applications that must preserve data integrity. In cursor mode, the user can mark reference points on the plots. Slice mode refers to behaviors specific to 3D plotting described below. In all of these modes, the

user has the choice of having crosshairs follow the mouse cursor, displaying in all plots the coordinates of the cursor in the reference frame of the plot in which it resides.

The GAPlotObj library provides both standard 2D Y-versus-X plots and 3D displays. 3D data items can be shown as surfaces, images, or contour plots. They can also be “sliced,” meaning that they can be plotted as a function of one of their two independent dimensions. Furthermore, 2D and 3D plots can be linked so that actions performed on one plot affect others. For example, Fig. 2 shows three views of a 3D temperature profile from the core Thomson scattering system on DIII-D. The top view is a contour plot showing temperature versus time and position (elevation). The middle view is a time history slice plot of the temperature at the midplane, that is, at a fixed position. The bottom view shows a profile slice plot of temperature versus elevation at a fixed time. In slice mode, the plots can be linked as follows: if the user moves the crosshairs in the contour plot from $t = 1400$ ms to $t = 3000$ ms, the crosshairs also move in the time history slice plot. In the profile plot, the time at which the profile is displayed changes from 1400 ms to 3000 ms. Mixing and coupling 2D and 3D plots in one display is a capability not previously available in any general use tool at DIII-D, and provides a rapid means of examining 3D data.

The breakdown of the class structure of the GAPlotObj library, also shown in Fig. 1, reflects the organization of the display for the user. An application will typically have one object of the GA_PLOT_WINDOW class, which in turn will contain one or more objects of the GA_PLOT class. A GA_PLOT object contains one or more GA_DATA object together with GA_ANNOTATION objects. Each class encapsulates the methods and properties specific to it. The GA_PLOT_WINDOW class contains the methods for handling mouse modes and events, and display configuration. It also has properties determining layout. The GA_PLOT class has methods for drawing the axes of a plot, and for determining its ranges and the coordinates of the mouse cursor. Finally, the GA_DATA and GA_ANNOTATION classes have methods for drawing individual data traces and annotation like lines and text. The 3D capabilities of the GAPlotObj library are obtained by subclassing the GA_PLOT and GA_DATA classes to

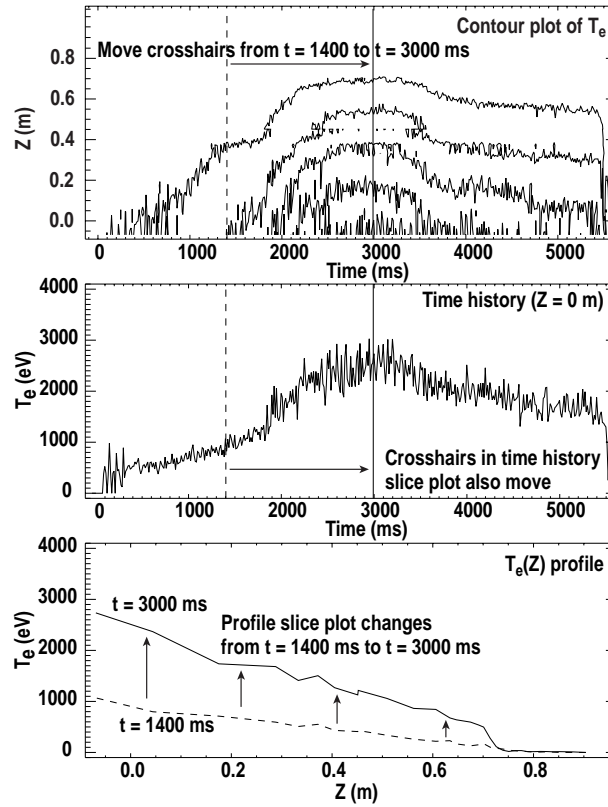


Fig. 2. Interaction between 3D contour plot and 2D slice plots of temperature profile.

GA_PLOT3 and GA_DATA3 respectively. The subclasses inherit common properties and methods, and add others specific to 3D plotting.

3. MDSplus

DIII-D uses the PTDATA system [5] for acquisition, storage, archival and retrieval of raw data. Prior to the adoption of MDSplus for DIII-D in September of 1997, no equivalent system existed for the results of analyses performed on the raw data. Each analysis code or diagnostic therefore developed its own system, writing their results in differently formatted flat files, sometimes hundreds per plasma pulse, scattered across the disks of multiple computers. Often these data were stored without any contextual information or comments, so that in addition to the confusion of having to use a different code to access each analyzed dataset, it was not possible for scientists to explore each others results without first meeting to determine the context of the data. Sharing data with remote collaborators was therefore difficult.

The MDSplus system is now being used as a centralized repository for analyzed data at DIII-D. All data from each diagnostic or code is accessible with the same set of commands, again benefitting collaborators and resident scientists alike by reducing the amount of requisite knowledge. Scientists need to know only the location of the data in the MDSplus tree, and do not need to worry about the format in which the data is stored or the location of the files. The browsable, hierarchical tree structure, together with comments, geometry, and documentation stored with the data, make it easier for a scientist to understand the results of the analyses of others, and thus remote exploration of DIII-D data can be productive.

MDSplus also simplifies the tasks of the Data Analysis Applications Group at DIII-D. Before MDSplus, each data access routine had to be maintained on each of the operating systems in the heterogeneous DIII-D computing environment. Now, only the MDSplus libraries need to be maintained. Data in MDSplus is also compressed, saving disk space, and only a handful of files are necessary to store the results from an entire plasma pulse. The flexibility of MDSplus allows addition of new datasets to the existing archive and storage of multiple versions of analysis.

A key benefit of MDSplus is its ability to interface with other data systems. For example, the interface to the PTDATA system allows users to transparently retrieve the raw data stored within through MDSplus. This is accomplished by writing MDSplus functions containing calls to the PTDATA shared library routines. In situations where analyzed data are simply calibrations applied to raw measurements, MDSplus will store the calibrations and an expression multiplying them by the raw data stored in PTDATA. This reduces the amount of disk space used, and allows calibrations to be changed independently of the raw data, while automatically providing the most up-to-date results to the users.

As of the beginning of June, 1999, analyzed data from over 4700 DIII-D shots are stored in MDSplus, using 38 GB of disk space. Up to 16 separate analyzed datasets are present for each shot, including EFIT results, kinetic profiles from the Thomson scattering and CER diagnostics, and spectroscopy and divertor imaging results. Currently the full set of analyzed data stored in MDSplus for one shot uses about 20 MB, where the entire raw data set for one shot uses about ten times as much space. More analyzed datasets are being added both to new shots and retrospectively to old ones. During experiment operation, most of the data are loaded automatically between plasma pulses, with analysis jobs and data storage triggered by events declared when the requisite raw data are available after acquisition. MDSplus loading can occur at any time following a plasma pulse, so that the results of more detailed analyses in the days following an experiment can be stored.

The initial installation of MDSplus at DIII-D supplied data to the heterogeneous Unix workstation cluster from an OpenVMS server. With the completion of the Unix MDSplus port, a joint effort of MIT, GA, PPPL and LLNL, DIII-D recently switched to Unix MDSplus data service. Data is now served from a Compaq DEC AlphaServer 800 5/500 running Compaq Tru64 Unix, with an Andataco GigaRAID AA system containing 100 GB for storage. This system integrates more closely with the DIII-D computing environment, and provides an upgradable path for future expansion in both CPU power and magnetic storage space. Given the projected growth rate of MDSplus storage needs and the falling cost of hard disk storage, the long term

plan is to store all MDSplus data on magnetic disk, rather than involving slower, more complex optical or tape mass storage systems.

4. TOOLS

ReviewPlus and EFITviewer are two IDL-based data visualization tools in use at DIII-D that employ the GAPlotObj library and MDSplus. ReviewPlus is the main data viewing tool for experimental operations, and is also used for more in-depth analysis afterwards. It takes full advantage of the 3D capabilities of the GAPlotObj library to display interactive time histories, profiles, and 3D plots. Figure 3 shows the ReviewPlus main window with four plots representative of the types that it can make. In the top left is a simple time history of the plasma current obtained from the PTDATA system. Tools in general use prior to ReviewPlus could make only these types of graphs. In the bottom left, a spatial profile of the electron temperature at a fixed time has been composed by overlaying measurements from the three Thomson scattering systems on DIII-D. Using the slice mode of the GAPlotObj library, users can adjust the time for which the profile is displayed by moving the mouse crosshairs in any time history plot. The top right shows a surface plot of the pressure profile output from the EFIT code, where pressure is plotted versus both time and normalized flux. The surface can be interactively rotated around two axes to help the user identify features of interest. The bottom right graph illustrates ReviewPlus' ability to plot "any Y versus any X;" in this case, the time history of the central electron temperature is plotted versus the time history of the central electron density. ReviewPlus interpolates both signals to a common timebase and then plots one versus the other.

ReviewPlus is also able to plot any mathematical combination of data signals from MDSplus or PTDATA, including allowing built-in or user-written IDL functions to operate on the data. Menu options are provided for the most common operations: smoothing, integration, differentiation, and fast Fourier transforms. A feature especially useful during experiment operation is the ability to track the most recent pulse, so that users can create a set of plots that automatically refresh every shot. For users unfamiliar with the signals available in the DIII-D repository, a searchable menu is provided, along with the ability to link to a web browser to

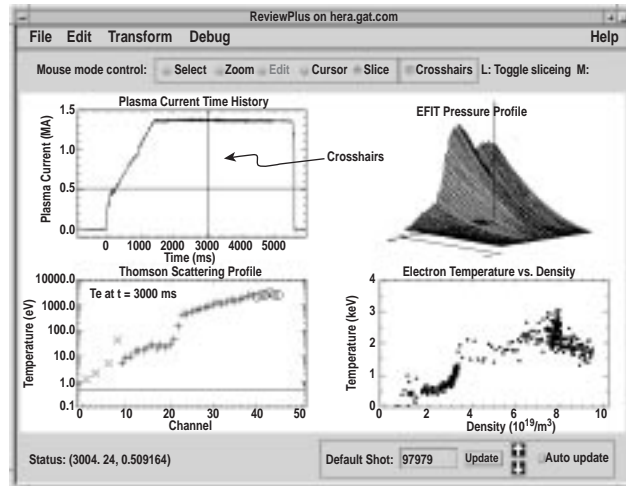


Fig. 3. ReviewPlus main window with representative plots.

display more detailed information on a signal of interest. All of these features combine to make data visualization more efficient and more productive.

The EFITviewer code is also used heavily during both experiment operation and for in-depth analysis. Its main purpose is to provide a rapid means of assessing the magnetic equilibrium fit from EFIT. To this end it displays a poloidal cross section of the magnetic flux contours overlaid upon tokamak and diagnostic geometry, and target shapes input to the plasma control system, as shown in Fig. 4. This display is dynamic; users can zoom in on features of interest, turn geometry overlays on and off, and overlay the cross sections of multiple shots for quick comparison. Other plots available in EFITviewer are equilibrium fit quality, profiles and time histories of various calculated parameters, and kinetic profiles for comparison. At the press of a button users can map the profiles to any of three possible coordinate systems: major radius, toroidal flux, or poloidal flux. The profile plots are organized into separate customizable pages grouping related information together so that users can focus on plots of interest to their analysis needs. EFITviewer is part of a larger suite of tools developed for enhancing use of the EFIT code; details can be found in [6].

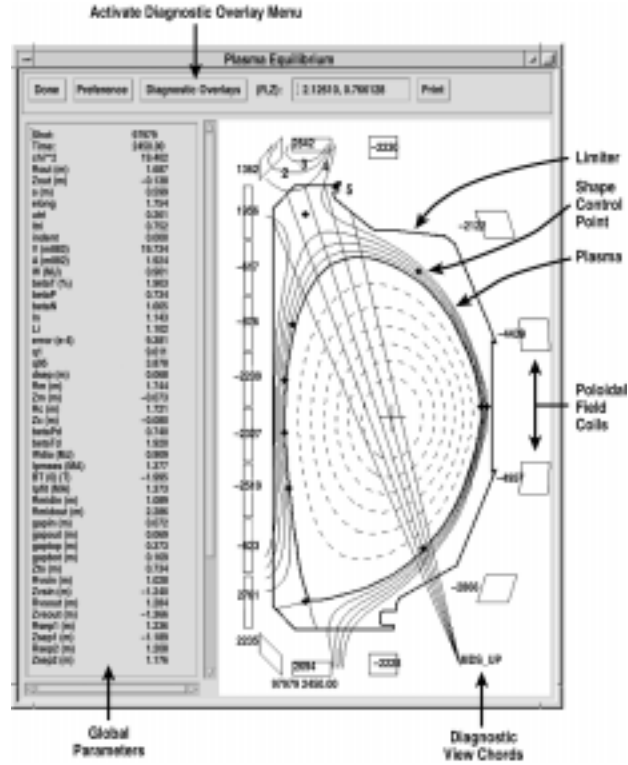


Fig. 4. EFITviewer plot of plasma equilibrium and machine-specific overlays.

5. MULTIPLE SITES

Both ReviewPlus and EFITviewer are designed to run on remote host systems licensed for IDL, connecting to DIII-D via MDSplus and PTDATA for data retrieval. These tools have already been installed and operated at the Lawrence Livermore, Oak Ridge, and Princeton Plasma Physics laboratories, at the MIT Plasma Science and Fusion Center, and even at JET. Utilization of off-site hosts benefits both the remote collaborator and the DIII-D computing infrastructure when the bandwidth required for interacting with the tool exceeds that required for transmitting the data. 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 CPU's would be less burdened. In situations where remote collaborators cannot obtain IDL licenses, or where the bandwidth required for data retrieval is greater than for interactive display, users are able to log into the DIII-D computing network to use the tools on DIII-D CPU's, setting their display back to their own, off-site host.

Based on the success of existing collaboratively developed software tools, including TRANSP from PPPL, MDSplus from MIT, and EFIT from GA and LLNL, the U.S. Magnetic Fusion Energy sites are exploring new areas for collaboration on software tools of common interest. Future projects will include tools to view results from and prepare data for existing power balance transport analysis codes, and database applications for analysis code run management and signal name translation. These discussions have the promise of creating a situation where collaborators could use their tools of choice to view data from any experiment or simulation code at any location around the world.

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 the 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.

On the other hand, while EFITviewer can display EFIT results from other tokamaks, its full capabilities are not available to them as it contains hardcoded information specific to DIII-D. Most of this hardcoded information is machine and diagnostic geometry. A plan is underway to take advantage of the flexibility of MDSplus to make EFITviewer fully site-independent. Machine specific information will be removed from the code and stored in the MDSplus EFIT tree together with the EFIT results. The structure of EFIT trees at different sites will be standardized so that they will all be readable by EFITviewer for display. EFITviewer would then be able to present machine-specific menus for geometry overlays, configured by the content of the EFIT trees it reads rather than by hardcoded site-specific information.

The idea of making tools like EFITviewer “data driven,” that is configurable by the structure of the data that they use, has the potential to open up many more codes to use at multiple sites. Long term plans at DIII-D are to make the EFIT code itself site-independent, and to work with the rest of the community on other codes of common interest.

6. SUMMARY

Collaborating scientists are an integral part of the DIII-D National Fusion Facility. The GAPlotObj visualization library and MDSplus data system are being used by the Data Analysis Applications group to enhance their ability to analyze DIII-D data. ReviewPlus and EFITviewer, two tools employing these systems, do so by making data visualization easier, faster, and more productive, by opening up new possibilities for analysis, and by their ability to run on remote hosts accessing data from multiple sites.

REFERENCES

- [1] D.P. Schissel, Q. Peng, J. Schachter, T. Terpstra, B. Meyer, T.A. Casper, and R.Jong, “Enhanced computational infrastructure for data analysis at the DIII–D National Fusion Facility,” these proceedings.
- [2] B.B. McHarg, Jr., “The DIII–D Computing Environment: Characteristics and Recent Changes,” these proceedings.
- [3] J.A. Stillerman, T.W. Fredian, K.A. Klare, and G. Manduchi, MDSplus data acquisition system, *Rev. Sci. Instrum.* **68** (1997) 939–942.
- [4] Research Systems, Incorporated, <http://www.rsinc.com>.
- [5] B.B. McHarg, Jr., “Access to DIII–D Data Located in Multiple Files and Multiple Locations,” *Proc. 15th Symp. on Fusion Engineering* (1993) 123.
- [6] Q. Peng, *et al.*, “IEFIT – An Interactive Approach to High Temperature Fusion Plasma Magnetic Equilibrium Fitting,” *Proc. 11th IEEE NPSS Real Time Conference*, Santa Fe, New Mexico, to be published in *IEEE Transactions on Nuclear Science* (1999).

ACKNOWLEDGEMENTS

The authors wish to thank K. Keith, B. McHarg, C. Parker, and T. Warner from General Atomics for their critical efforts in developing and supporting the computing infrastructure on which our tools depend. We would also like to thank T. Fredian, M. Greenwald, and J. Stillerman from Alcator C-Mod for their advice and support of MDSplus, and T. Casper, B. Meyer, and J. Moller from LLNL. Finally, we would like to thank the DIII-D scientific community for their valuable feedback during the development process. Work supported by U.S. Department of Energy under Contract No. DE-AC03-99ER54463.