EXECUTIVE SUMMARY

This paper summarizes the architecture, including hardware and software, which is used at the DIII–D National Fusion Facility for analyzed data acquisition, analysis, and archival/ retrieval. A number of the new concepts developed for this system should be applicable to KSTAR. We would welcome the ability to discuss these ideas further and to explore potential collaborations between the U.S. Fusion Program and KSTAR.

The DIII–D system supports both near real-time on-line monitoring and direction of the operation of the DIII–D magnetic fusion (tokamak) experiment and longer-term detailed analysis of the resulting plasma physics data. This comprehensive data analysis capability facilitates efficient conduct of plasma science experiments using DIII–D, under the direction of an operation and science staff comprising more than 200 individuals drawn from both General Atomics and worldwide fusion laboratories, universities, and industries. 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. Since a large fraction of the DIII–D team is off-site, DIII–D data and applications are readily available at their various home institutions via high-speed network connections.

The DIII–D analyzed data acquisition, analysis and archiving system is fully modular and network-based, and is configured to provide a uniform data processing and display environment supporting both interactive and automatic analysis by local and remote users. Figure E–1 shows the overall configuration of this data analysis environment that handles storage for both raw (as digitized) and analyzed data from each DIII–D plasma operation pulse (commonly termed a "pulse" or "shot"). Data from each pulse comprises a time-sampled array of information from the various plasma diagnostic systems that view and record the plasma, plus corresponding records of the various DIII–D device (tokamak) operation settings that are used to produce the pulse. The overall duration of each DIII–D operation pulse is approximately 10 s, with the duration of the plasma portion typically being from 3 to 7 s. Each pulse typically produces 150 MB of compressed raw digitized data. Once the data is acquired, specialized analysis routines are automatically initiated on a variety of computers. As soon as any raw or analyzed data is acquired, it can be graphically displayed by the DIII–D operations and experiment staff. Typically, about 4 minutes after the pulse ends, a complete temporal history of the pulse is available including magnetic equilibrium and profiles of density, temperature, current, and plasma rotation. The DIII–D team, both on and off-site, uses this near real-time "control-room" data to monitor the results of the pulse and modify the parameters for the next. Figure E–2 shows a typical example of a control-room presentation.

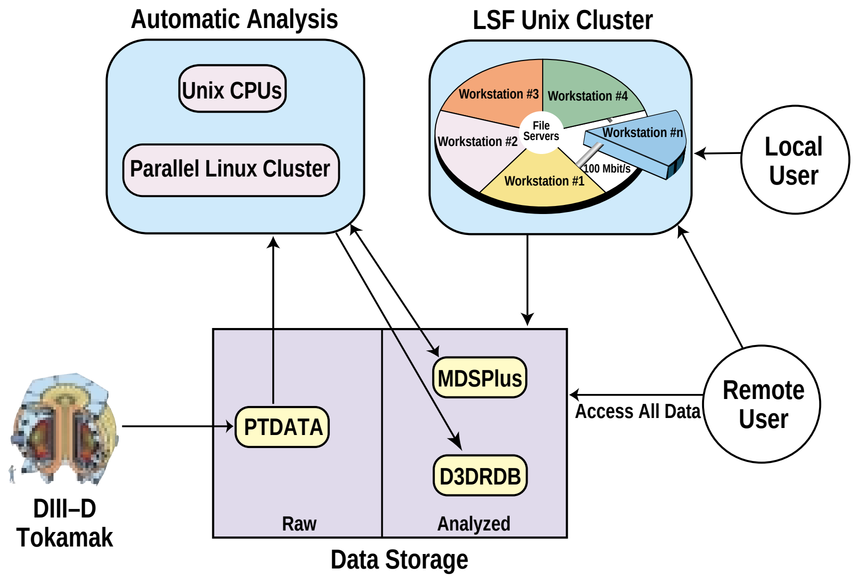

Fig. E–1. Data flow at the DIII–D National Fusion Facility.

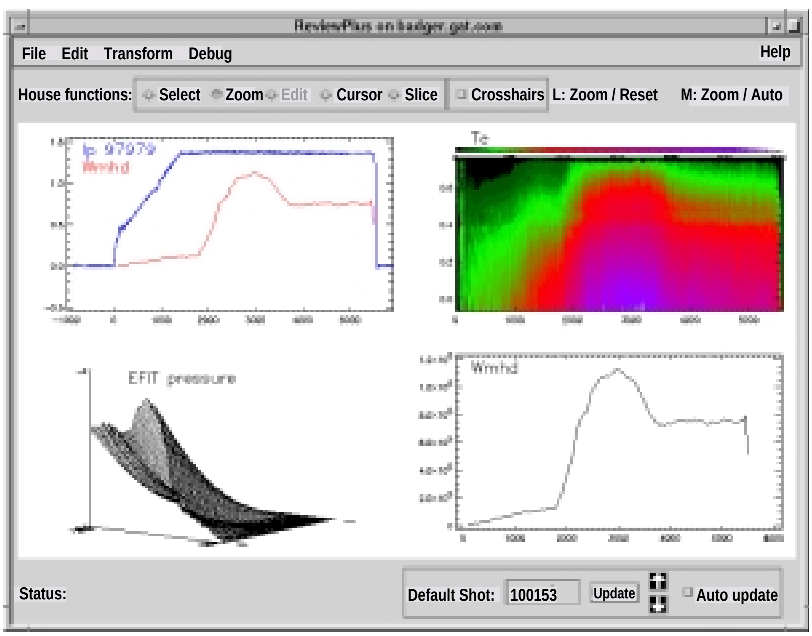

Fig. E–2. Interactive visualization is available near real-time during DIII–D operations .

Data is archived in three formats, with increasing degrees of analysis and reduction to higher level plasma parameters and attributes: raw (PTDATA, presently 1.4 TB), analyzed (MDSplus, presently 70 GB) and analysis highlights (DIII–D Relational Database or “D3DRDB”, presently 0.2 GB). The relationship among these three archives is illustrated schematically in Figure E–3. The D3DRDB contains the most refined and abstracted distillation of a pulse and can be used by the DIII–D team to investigate similarity and correlations among subsets of the data. The

MDSplus archive contains all analyzed data and allows the staff to conduct detailed studies of candidate pulses identified by inspection or D3DRDB searches. The PTDATA archive contains all raw or as digitized DIII–D data.

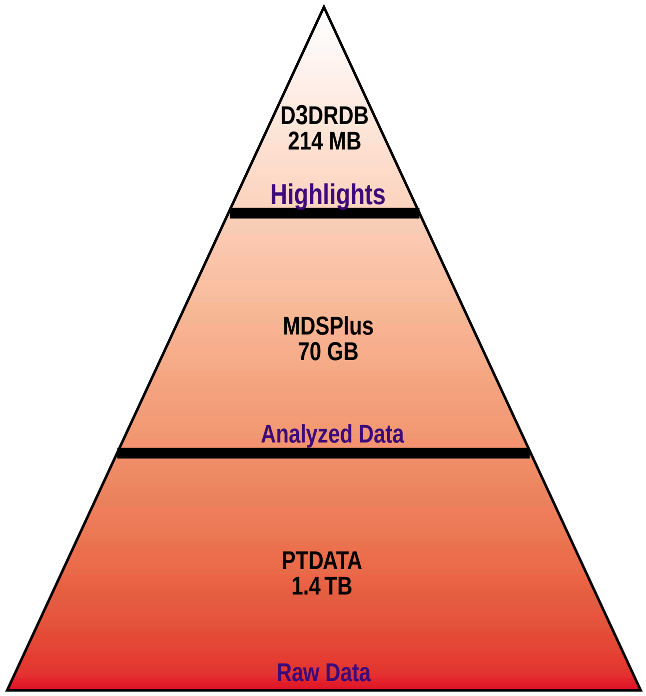

Fig. E–3. The DIII–D relational database, D3DRDB, works in concert with PTDATA and MDSplus by storing highlights of the data.

The raw data acquisition system (PTDATA) presently supports the acquisition of approximately 300 GB of compressed raw data per year (~2000 pulses) and also provides for cumulative storage and timely access to the 1.4 TB of data that has been acquired since DIII–D operation began in 1986. A 3 TB hierarchical data storage system comprising 200 GB of hard disk storage, 600 GB of magneto-optical storage and 2.2 TB of tape storage provides unattended access to all raw data with a delay that does not exceed 5 minutes. Data from the most recent or frequently accessed high interest pulses are available instantaneously from the hard disk. Data from less frequently requested pulses are available from the magneto-optical and tape archives on correspondingly slower but still interactive time-scales. Access to the data is completely transparent to the requester with only the time to retrieve the data varying. As the total amount of analyzed data is much less than raw data, both the analyzed data repository (MDSplus) and the DIII–D relational database (D3DRDB) are contained only on hard-disk.

As in most present and scientific and computer related endeavors, there is an overriding need to accommodate rapid and continued data growth. Figure E–4 shows the growth of the cumulative compressed raw data (PTDATA) since the start of DIII–D operations. This growth is presently scaling as $t^{2.7}$, where t is the time since the inception of DIII–D operation. If the present growth of raw data continues, this data archive will reach approximately 4 TB by 2004. It is anticipated that the increasing use of multi-dimensional (2D and 3D) experimental and simulation data will expand the analyzed data repository (MDSplus) to the 5 TB range by 2004.

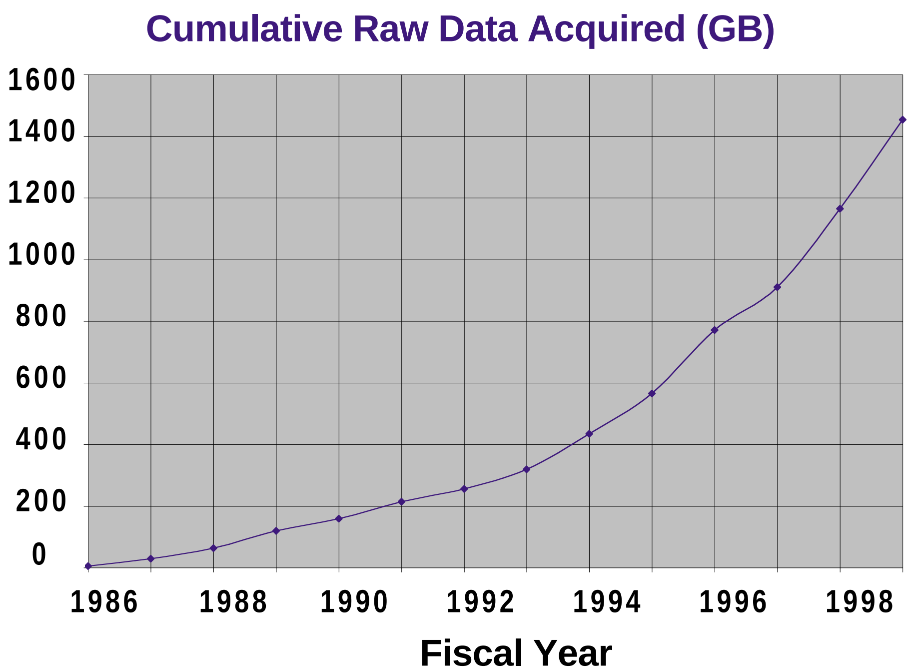

Fig. E–4. There has been a quadratic increase in total raw data (PTDATA) acquired over the lifetime of DIII–D operations.

The ability to update the system processing power to handle the increasingly sophisticated processing of 2D and 3D analysis, and to handle high time resolution analysis of longer duration pulses, is critical to the continued success of the DIII–D program. Support of general data analysis by both local and off-site staff is handled by a powerful and readily expandable load balanced Unix workstation cluster. Additionally, specialized workstations and a 12-processor Linux parallel processing cluster provide local users with near real-time automatic processing of key data used in the pulse-to-pulse conduct of experiments. Such recent increases in processing power have yielded near-real-time analyzed data feedback to the experimental team. Certain key pieces of data can now be compared between pulses without having to make allowances for approximations or interpolations. The combination of high time resolution and high spatial resolution with enhanced data visualization has resulted in more efficient utilization of DIII–D experimental time.

From the user’s point of view, all of the DIII–D data, regardless of where it resides, can be accessed and/or analyzed using a common and easily learned set of standard software tools and commands. This commonality of software and procedures, coupled with universal network access, facilitates rapid and efficient utilization of the DIII–D data by a national and international team of operations and research staff. Data visualization tools include specialized applications and one general purpose program that provides interactive 2D and 3D graphs. As the amount and complexity of data continues to increase, applications are being developed that use advanced graphics techniques that take advantage of specialized hardware for rapid visualization of 3D images.

The DIII–D analyzed data acquisition, analysis and archiving system had its origins in the now rudimentary data acquisition system that was initially set up in 1975 to support the operation of the Doublet III tokamak experiment, DIII–D's direct predecessor. Since then, hardware and software configurations of the DIII–D data system have evolved over the course of time into what is now a fully network-based and modular system predicated on standard computer industry data storage and processing protocols. There are many "lessons learned" behind the present configuration. Briefly put, the most important of these lessons learned are:

- The need for continuous expandability in data capacity and analysis computing power
- The need to be able to incorporate new hardware and software innovations without the need to modify the user interface and operation aspects of the system
- The importance of configuring the system in a modular fashion to be able efficiently use industry standard hardware and software
- The need to make both current and past/historic data, both raw and analyzed, available to present users in a timely and transparent fashion
- The need to provide a common and effective and user-friendly set of data retrieval, analysis and viewing/presentation tools
- The need to provide immediate and high-quality presentation of raw and initially analyzed data to the control room and remote-site operating staffs

As KSTAR progresses towards operation, it can benefit from our experience and our hardware and software solutions. This can save valuable time, allowing KSTAR staff to focus on machine operation and initial experiments. It will also benefit KSTAR scientists as they participate in the international community, since MDSplus and a lot of our software tools are used at a number of the international fusion laboratories.