ENHANCED DATA ANALYSIS AT THE DIII–D NATIONAL FUSION FACILITY

Presented by

D.P. SCHISSEL

for the DIII–D National Team*

Presented to

40th Annual Meeting, APS Division of Plasma Physics

November 16–20, 1998
New Orleans, LA

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

279–98 1998 APS/DPP Meeting

GENERAL ATOMICS

IN COLLABORATION WITH

- *The Data Analysis Group*
  - *Qian Peng, Jeff Schachter from GA*
  - *Bill Davis and Ted Terpstra from PPPL*
- *DIII–D Computer Staff*
  - *Kristi Keith, Bill McHarg, Carl Parker, and Tony Warner*
- *The DIII–D User Community*
- *From LLNL — Tom Casper, Bill Meyer, and Jeff Moller*
- *From MIT/C–Mod — Martin Greenwald, Tom Fredian, and Josh Stillerman*
- *Supported by U.S. DOE Contracts DE-AC03-89ER51114 and DE-AC02-76CH03073*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

MISSION STATEMENT IS TO INCREASE DIII–D DATA ANALYSIS THROUGHPUT AND DATA RETRIEVAL RATE

- **Both in–between shot analysis and post-experiment analysis**
- **CPU power**
  - In a load balanced multi-workstation environment
- **Networks**
  - Deploy switched 100 baseT to the office desktop
- **Unified file and data storage**
  - Unified interface for all raw shot and analyzed data
  - User files centralized with large NFS file server
  - All data and files available 24 hours/day, 7 days/week
- **Enhanced computational tools**
  - Streamlined analysis and automation where possible
  - New analysis and viewing capabilities
  - GUI tools to enhance the user base
  - Major analysis and viewing tools operate on collaborator's off–site computer
  - Detailed documentation with help
  - Interface and coding standards

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

COMPUTER HARDWARE FOR DATA ANALYSIS

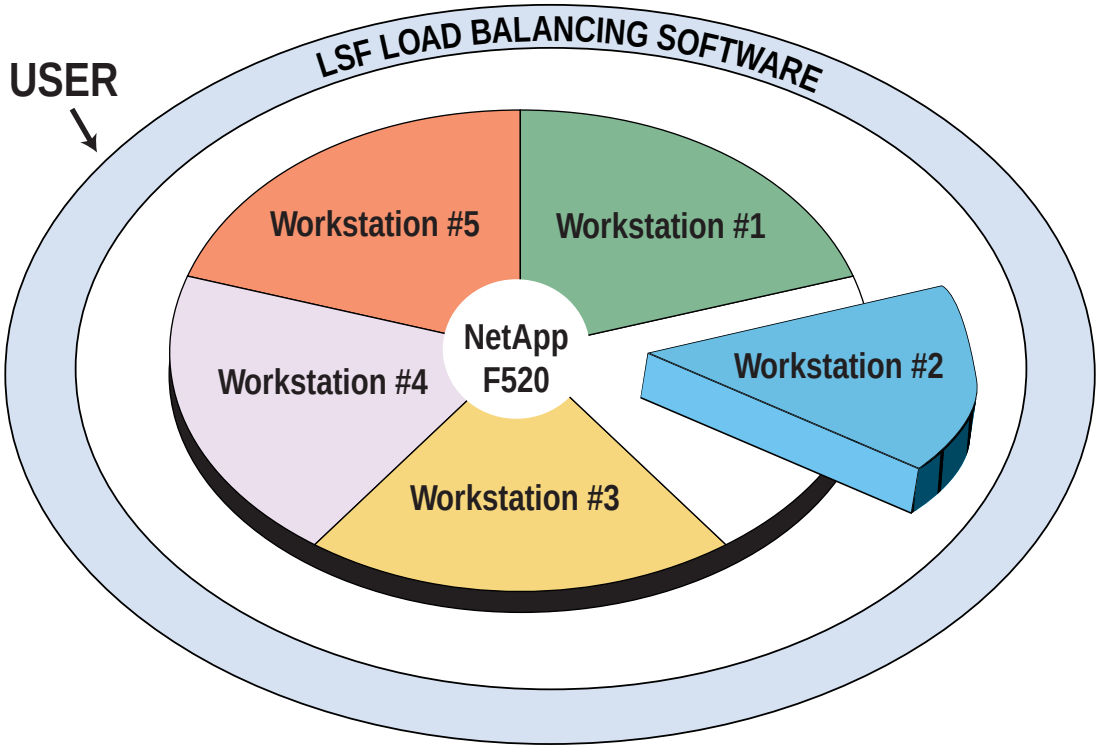
- *Large Unix Server – HP 9000 T600 Enterprise Server*
  - *3 CPUs with a total of 1GB memory to a total of 44.7 SPECfp95*
  - *Over 350 accounts including on–site and off–site collaborators*
- *HP 9000 based Unix workstations*
  - *C180 and a 735/125 dedicated to tokamak operations*
  - *Another C180 and numerous 735s available*
- *DEC 433au Unix Personal Workstations*
  - *5 machines placed on the desktop but in a shared environment*
  - *Each workstation has a speed of 18 SPECfp95*
- *DEC AlphaServer 4000 5/300 running OpenVMS*
  - *13 SPECfp95 available to all on–site and off–site collaborators*
  - *One OpenVMS Personal Workstation 433au running OpenVMS for legacy code*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

BETTER UTILIZATION OF COMPUTING POWER WITH LOAD BALANCING SOFTWARE

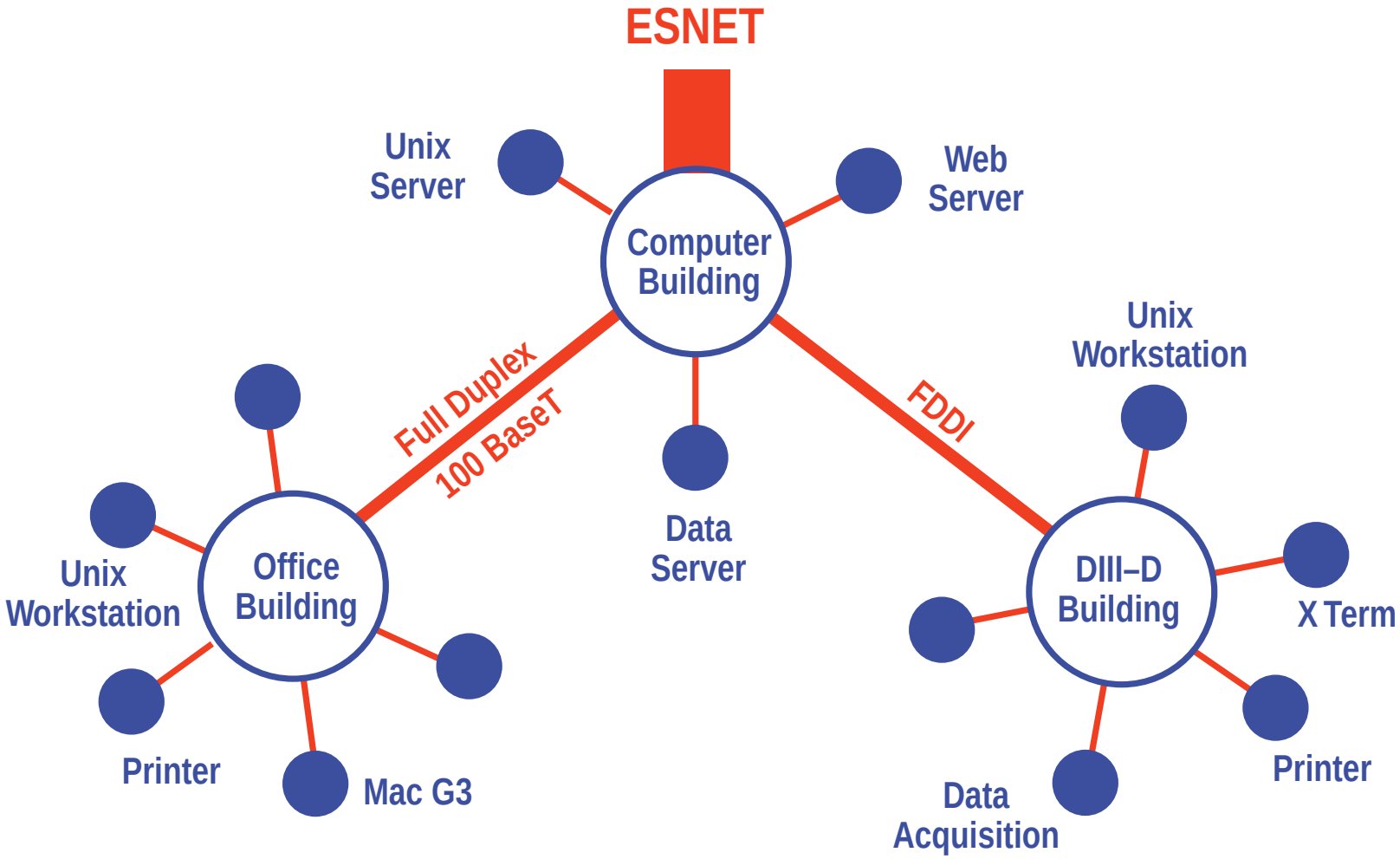
- Heterogenious Unix environment with a large server and numerous workstations
- Platform Computing's LSF Suite performs interactive load sharing
- Cost effective for sharing CPUs and commercial software
- Central user file repository via a 100GB NetApp F520 NFS file server



DIII-D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

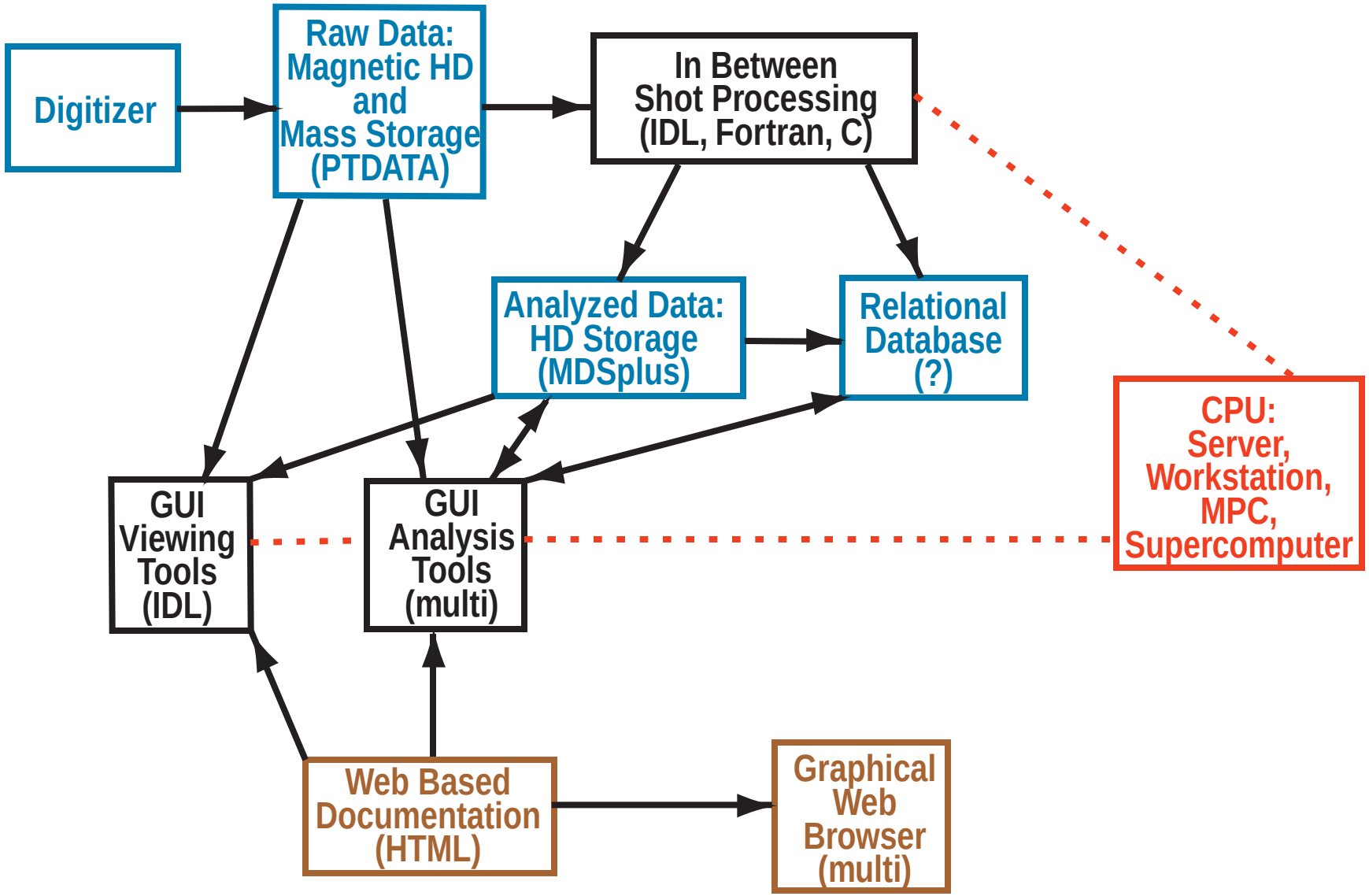
AN UPGRADED STAR NETWORK TOPOLOGY WILL UNIFY THE COMPUTING ENVIRONMENT



DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

DIII–D DATA FLOW CHART

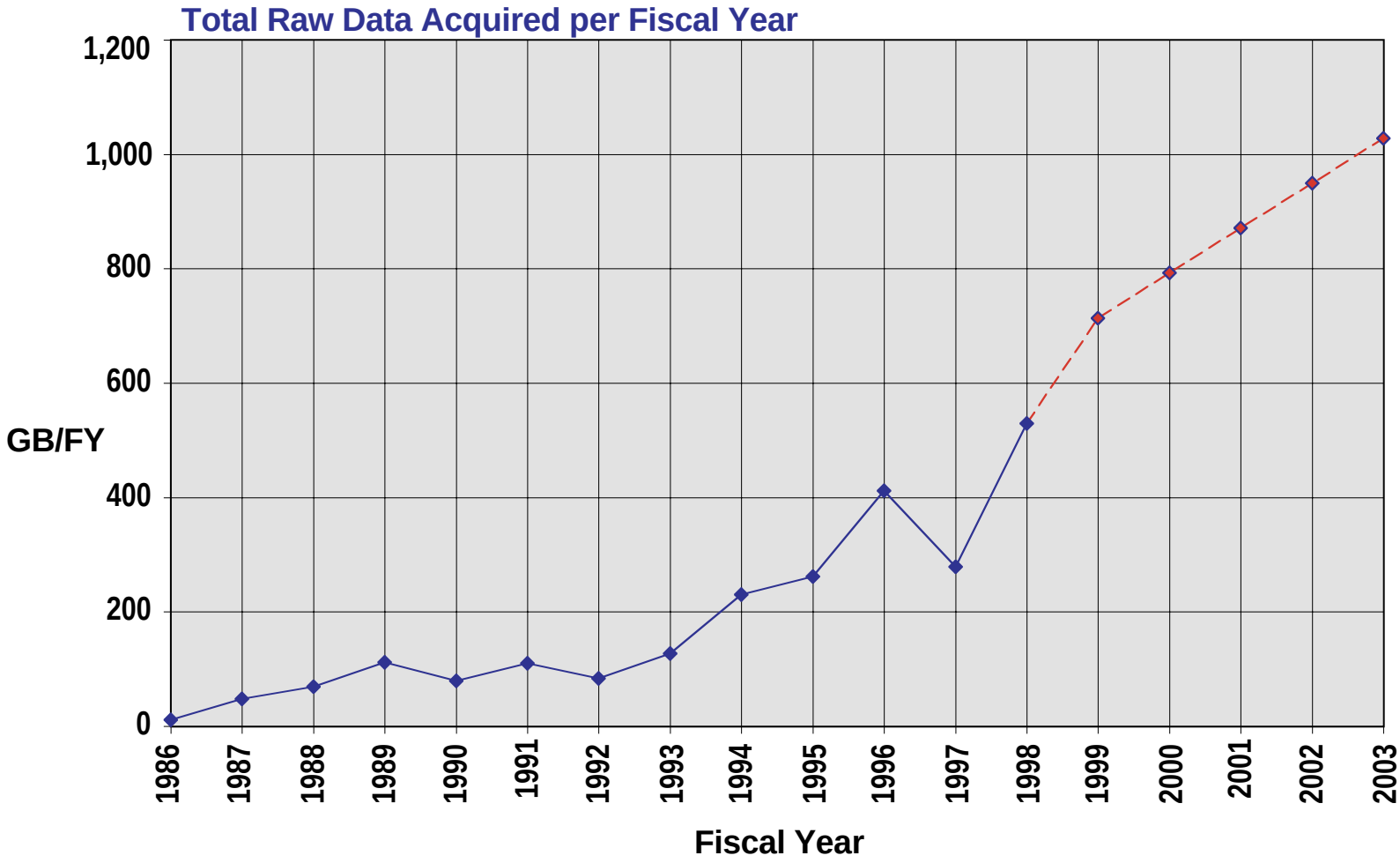


DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

RAW SHOT DATA FILE SIZE CONTINUES TO INCREASE

- More diagnostics, faster digitization rates, potentially longer pulse lengths
- Assume 15 weeks in FY98 and 18 weeks thereafter



DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

ALL DIII–D DATA AVAILABLE AROUND THE CLOCK

- All raw shot data automatically available (7x24) for data mining
  - 720 GB interactively (15–30 sec minimum), ≈9000 raw compressed shots
  - 2.4 TB on a 2–5 minute minimum time scale per request
  - Present needs are 1 TB with about 0.3 TB added per year
- All analyzed shot data on magnetic hard drive

Sun Ultra–1 Clone
100 GB Magnetic–Raid 5

HP Magneto–optical 600fx
620 GB from 238 platters
2.6 GB/platter, 6 drives

ATL DLT 7000 drives
2.4 TB from 68 bays
35 GB/tape, 4 drives

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

MDSplus HAS PROVIDED A UNIFIED DIII–D ENVIRONMENT FOR BOTH RAW AND ANALYZED DATA

- *Utilized to unify data storage and retrieval*
  - *All 1998 shots and selected older ones for a total of 3600*
  - *PTDATA can be accessed through MDSplus*
  - *MDSplus data retrieval integrated into existing and new tools*
- *A wide variety of data has been stored*
  - *Control room EFITs*
  - *CER and Thomson Scattering data only available through MDSplus*
  - *Confinement related quantities, impurity density, auxiliary heating*
- *Presently operational with an OpenVMS server*
  - *Approximately 25 GB of data is being served*
- *Unix port of MDSplus*
  - *Joint effort between MIT, GA, LLNL, and PPPL*
  - *Testing has begun with scheduled implementation in a few months*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

PHYSICS ANALYSIS TOOLS ARE BEING ENHANCED ALONG WITH DOCUMENTATION FOR EASE OF USE

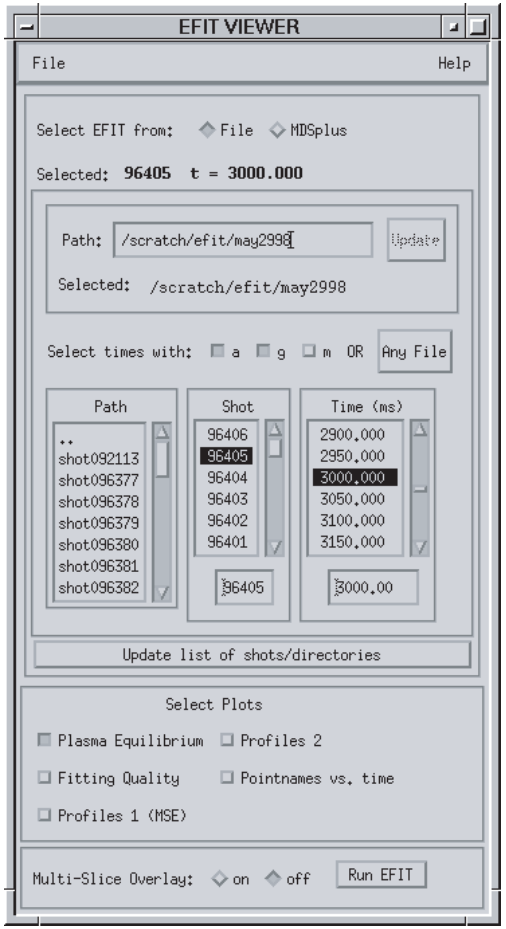
- *Streamlined analysis*
  - *Taking existing analysis packages and creating a more efficient interface*
  - *The time to create a kinetic EFIT has been reduced by a factor of 10*
- *Automatic between shot analysis*
  - *Using an event driven system analysis is automatically started after a shot*
  - *Data stored directly into MDSplus*
  - *Analysis includes EFIT, CER, confinement quantities, and Thomson Scattering*
- *GUI systems*
  - *Has helped to streamline some analysis*
  - *Enhanced the user base due to an easier interface*
  - *New capabilities including 3D graphics and interactive EFIT*
- *Web-based documentation also available from the analysis tool*
  - *PTDATA and MDSplus*
  - *Enhanced help on widely used codes such as ReviewPlus, 4D and EFIT*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

GUI-BASED VIEWING AND ANALYSIS TOOLS

EFIT Viewer GUI



- **Efficient GUI design**
  - Tools written in RSI's IDL, a commercial data package
  - Standardized interface
- **Created an object-oriented IDL-based graphics library (GaPlotObj)**
- **Viewing tools created to examine EFITs, time histories, and 2D&3D plasma profiles**
  - See Peng (R8P.23) and Schachter (R8P.24)
- **Analysis tools perform interactive EFIT, kinetic EFIT, and profile fitting**

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

WEB-BASED DOCUMENTATION BRINGS CRITICAL INFORMATION TO THE RESEARCHER

- *Computer code documentation*
  - *Both callable routines and compiled applications*
  - *Fortran and IDL based codes*
  - *Documentation has multiple levels for a user and a programmer*
- *PTDATA and MDSplus data documentation*
  - *Includes both raw digitizer data and analyzed data repositories*
  - *2/3 of the 3500 PTDATA pointnames have been documented*
  - *As data is added to MDSplus it is documented*
- *One source, multiple paths to reach the data documentation*
  - *Direct from Web browser*
  - *From within a tool*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

NEW ANALYSIS ENVIRONMENT EASES THE LEARNING REQUIREMENTS OF THE NEW COLLABORATOR

- *Unified data format with fast access*
  - *Raw data in PTDATA available 7/24*
  - *Analyzed data stored in MDSplus available 7/24*
  - *Remote PTDATA and MDSplus calls easy from a collaborator's code*
- *Viewing and analysis tools run on a collaborator's off-site computer*
  - *Requires an IDL license*
  - *Fast network access to data via a TCP/IP socket interface*
- *GUI systems*
  - *Efficient and simple interface design*
  - *Enhanced help capability via the Web based documentation*
- *Ability to examine multi-tokamak data simultaneously*
  - *Presently testing on C–Mod, DIII–D, and TFTR*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

A DRAFT DATA USAGE AND PUBLICATION POLICY HAS BEEN CREATED TO CLARIFY A COLLABORATOR'S RESPONSIBILITY

- *DIII–D raw and analysis data available without a local DIII–D computer account*
  - *For PTDATA access DIII–D can supply the software*
  - *For MDSplus access the MIT group can supply the software*
  - *MDSplus data access security is by IP number*
- *Collaborators' are offered full access to DIII–D data as it is collected and analyzed*
  - *No data will be supplied without signing the policy agreement*
- *Presentations of DIII–D data at conferences and workshops*
  - *Require approval of DIII–D Director*
  - *For major conferences a rehearsal presentation is expected*
- *The DIII–D program peer technical review is required for all papers*
- *Papers will be published on the Web either at DIII–D or on the collaborator's site*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

DIII–D PROGRAM COLLABORATORS

National Laboratories	Universities	International Laboratories
ANL	Cal Tech	Academia Sinica (China)
INEL	Columbia U.	Cadarache (France)
LANL	Hampton U.	CCFM (Canada)
LLNL	Johns Hopkins U.	Culham (England)
ORNL	Lehigh	FOM (Netherlands)
PNL	MIT	Frascati (Italy)
PPPL	Moscow State U.	Ioffe (Russia)
SNLA	RPI	IPP (Germany)
SNLL	U. Maryland	JAERI (Japan)
	U. Texas	JET (EC)
	U. Washington	KAIST (Korea)
Industry Collabs	U. Wisconsin	Keldysh Inst. (Russia)
CompX	UCB	KFA (Germany)
CPI (Varian)	UCI	Kurchatov (Russia)
GA	UCLA	Lausanne (Switzerland)
Gycom	UCSD	NIFS (Japan)
Orincon		Troitsk (Russia)
		Southwestern Inst. (China)
		Tsukuba U. (Japan)

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

A NEW UNIX-BASED RELATIONAL DATABASE WILL ALLOW THE SCIENTIFIC COMMUNITY TO MINE FUSION DATA

- *Desired user functionality*
  - *"Broad & shallow" containing summary data from all tokamak pulses*
  - *"Highly Analyzed" containing publication quality data*
  - *"Special Purpose" containing user specified information*

- *GUI interface layered on top of database engine*
  - *Relational queries from the Web, Fortran, C or IDL*
  - *Multi-platform clients*
  - *Independent of database engine*

- *Combined analysis of different tokamak databases will be straightforward*
  - *Similar to the sharing of our viewing and analysis tools*
  - *The development of these tools must be a coordinated effort*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

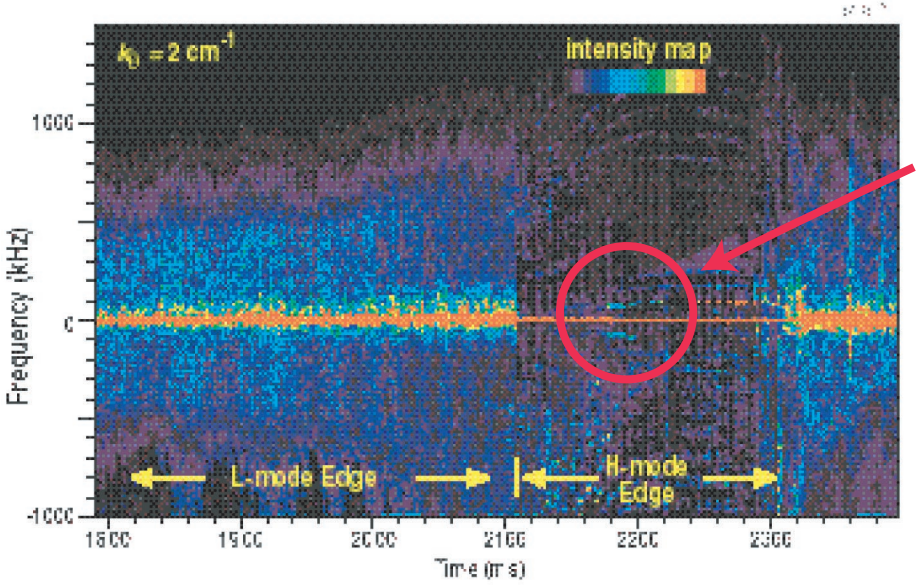
IDENTIFY THE ADVANTAGES OF OBJECT RELATIONAL DATABASES FOR THE FUSION COMMUNITY

- *Advanced multi-dimensional queries should revolutionize our data mining*
- *How will object relational technology compliment our usage of MDSplus?*
- *Collaborate with the High Energy Physics Community*
  - *Extensive use of object databases at SLAC*
- *Create a proof-of-principal test bed for community evaluation*
  - *Possible data choices are a TRANSP and ONETWO analysis or results from transport simulation codes*

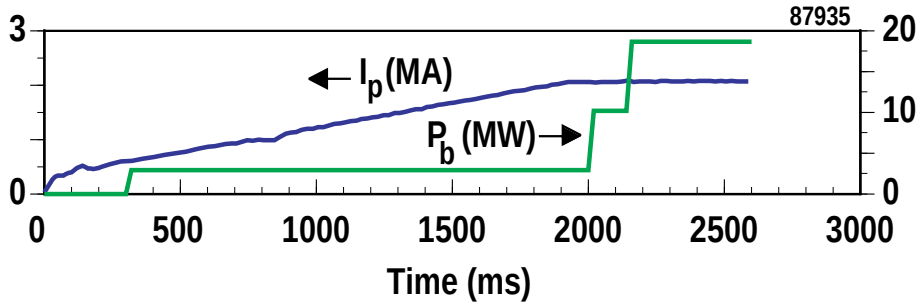
DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

OBJECT RELATIONAL QUERIES SHOULD REVOLUTIONIZE DATA MINING



Find other fluctuation data with similar mode behavior



Find discharges where Pb during the Ip ramp is 1/5th Pb during the Ip flattop phase

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

SUMMARY — A NEW ANALYSIS ENVIRONMENT

- *A load balanced multi-workstation environment supplying ample CPU power*
- *Fast 100 Mb/sec network for rapid data retrieval*
- *Unified data file storage for easy retrieval on a 7/24 basis*
- *Enhanced computational tools for efficient data analysis*
- *Analysis tools operational on remote CPUs to ease the local DIII–D burden*

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS