REMOTE COLLABORATION AT THE DIII–D NATIONAL FUSION FACILITY

Presented by

J. SCHACHTER

for the DIII–D National Team

Presented to

DOE Remote Collaboration Workshop

December 16–17, 1998
Germantown, Maryland

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

ACKNOWLEDGMENTS

- The Data Analysis Group
  - Qian Peng, Jeff Schachter, Dave Schissel from GA
  - Ted Terpstra from PPPL
- DIII–D Computer Staff
  - Kristi Keith, Bill McHarg
- 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-AC03-99ER54463 and DE-AC02-76CH03073

DIII–D PROGRAM COLLABORATORS

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

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

COLLABORATOR SUPPORT IS FUNDAMENTAL TO COMPUTING AT DIII–D

- Two-thirds of onsite scientists are collaborators
- 150 of 300 users off-site collaborators
- Increasing number of collaborators places more demand on computing infrastructure
- Data Analysis Group formed in part to support our collaborators
- In this presentation:
  - The needs of our collaborators
  - What we are doing to meet them
  - Future plans

THE COMPUTATIONAL NEEDS OF A REMOTE COLLABORATOR ARE NO DIFFERENT THAN A RESIDENT SCIENTIST

- Computer
- Access to data
- Tools to view and analyze data
- Communication with other scientists
- Control of equipment

The difference is sometimes in how these needs are met

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

COMPUTATIONAL SUPPORT OF DIII–D REMOTE COLLABORATORS

- **Computer**
  - Leverage use of remote CPUs
- **Access to data**
  - 7x24 remote access via MDSplus, PTDATA
  - And a relational database
- **Tools to view and analyze data**
  - Easy-to-use GUIs (IDL) with ample documentation
  - Tools run from remote CPU
- **Communication with other scientists**
  - Both in real-time and off-line

Much of this plan is in place now. Future work will build on current activity

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

COMPUTATIONAL SUPPORT OF DIII–D REMOTE COLLABORATORS

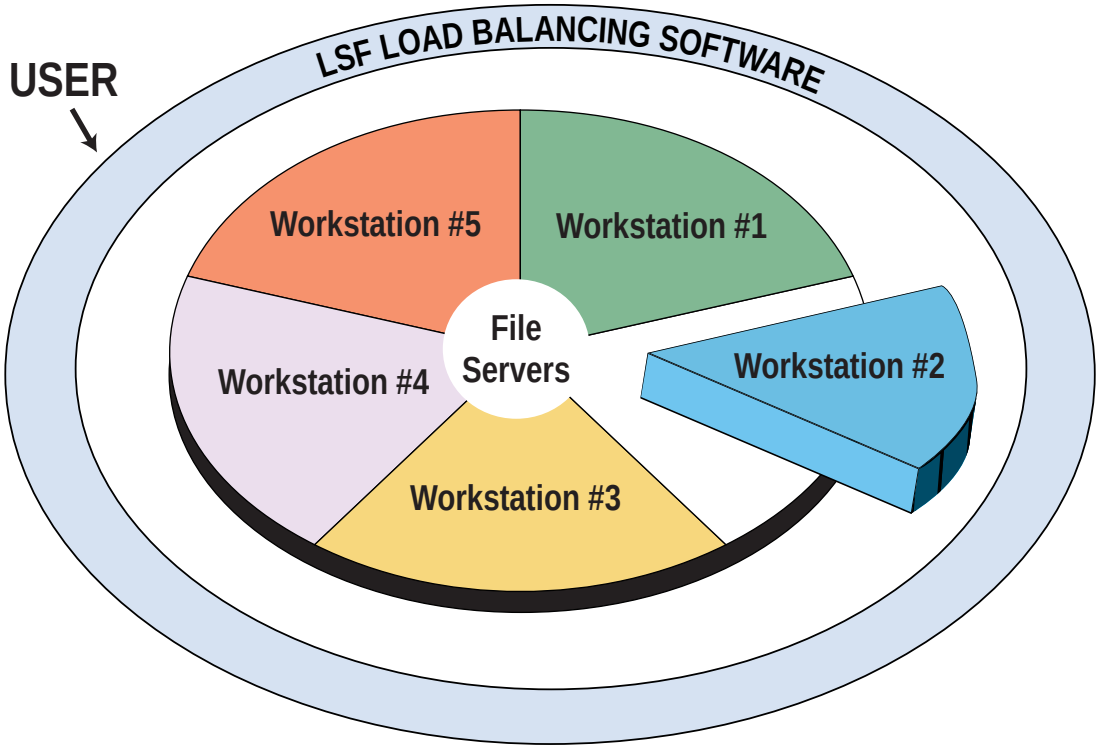
- Computer
  - Leverage use of remote CPUs
- Access to data
- Tools to view and analyze data
- Communication with other scientists

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

LOAD BALANCING SOFTWARE PROVIDES SUBSTANTIAL COMPUTATIONAL POWER TO ONSITE USERS

- Heterogeneous 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 file servers for data and user files



DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

REMOTE COMPUTING COULD BENEFIT COLLABORATOR AND DIII–D

Remote Site

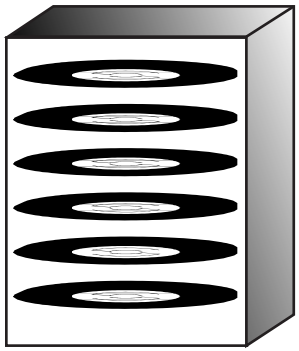
- Computation
- Window management

Data (MDSplus)

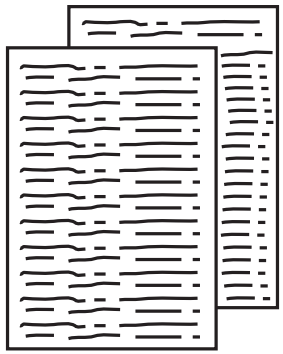
TCP/IP

Codes (CVS)

DIII–D



- Data repository



- Code repository

Faster response from tools

Less load on infrastructure

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

FUTURE REMOTE COMPUTING ISSUES

- Do we integrate remote CPUs into load-balanced environment?
  - Technically possible
  - Performance?
  - Management? Security?
- Remote file sharing via DFS?
  - Easier to share files, codes
  - LLNL uses — Tom Casper will say more
  - Data already served via TCP/IP sockets

COMPUTATIONAL SUPPORT OF DIII–D REMOTE COLLABORATORS

- Computer
- Access to data
  - 7x24 remote access via MDSplus, PTDATA
  - And a relational database
- Tools to view and analyze data
- Communication with other scientists

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

ALL DIII–D DATA AVAILABLE AROUND THE CLOCK

- All raw shot data (PTDATA) 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 (MDSplus) 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

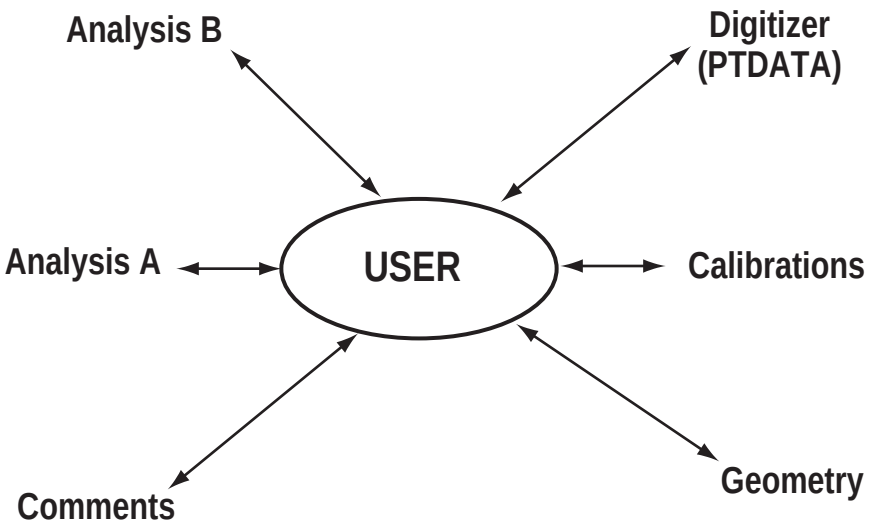
ATL DLT 7000 drives
2.4 TB from 68 bays
35 GB/tape, 4 drives

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

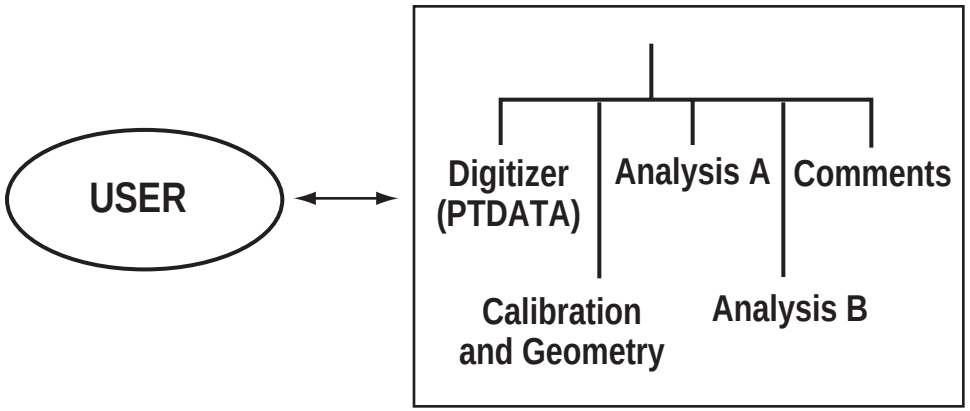
MDSplus SIMPLIFIES DATA ACCESS

Conventional Storage



- Separate interface for each data type
- Must know data format and file location
- Data and context stored separately
- Hard to share results

MDSplus



- One interface to many data types
- Only need location of data in tree
- Store all relevant information
- Remote exploration of data productive

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

A NEW 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 tools connected to 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

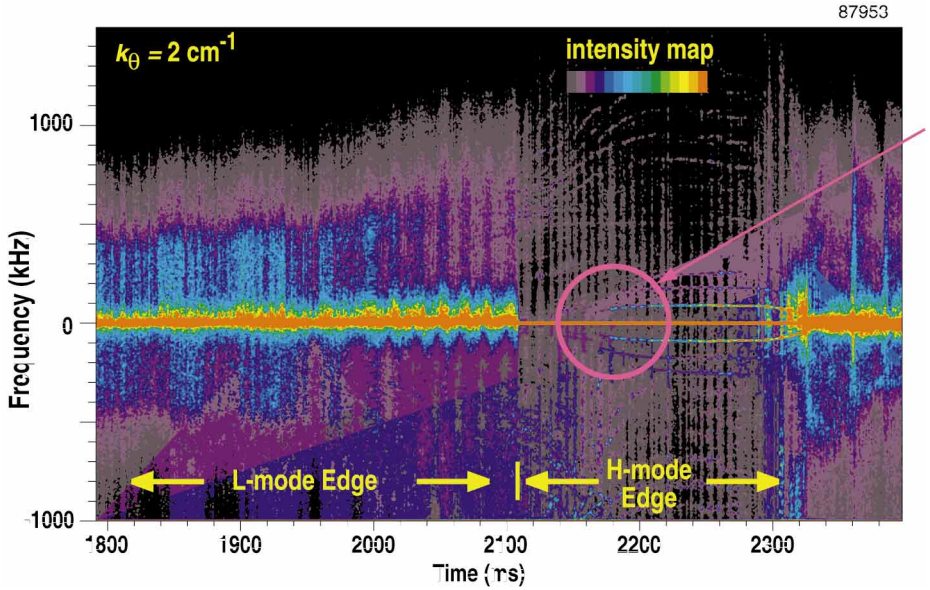
IDENTIFY THE ADVANTAGES OF OBJECT RELATIONAL DATABASES FOR THE FUSION COMMUNITY

- 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

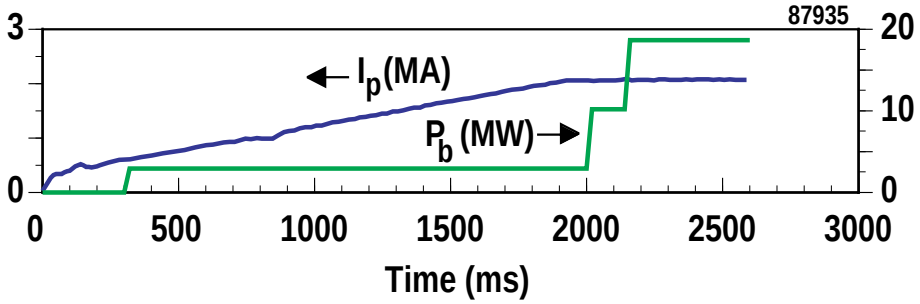
DIII–D NATIONAL FUSION FACILITY SAN DIEGO

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

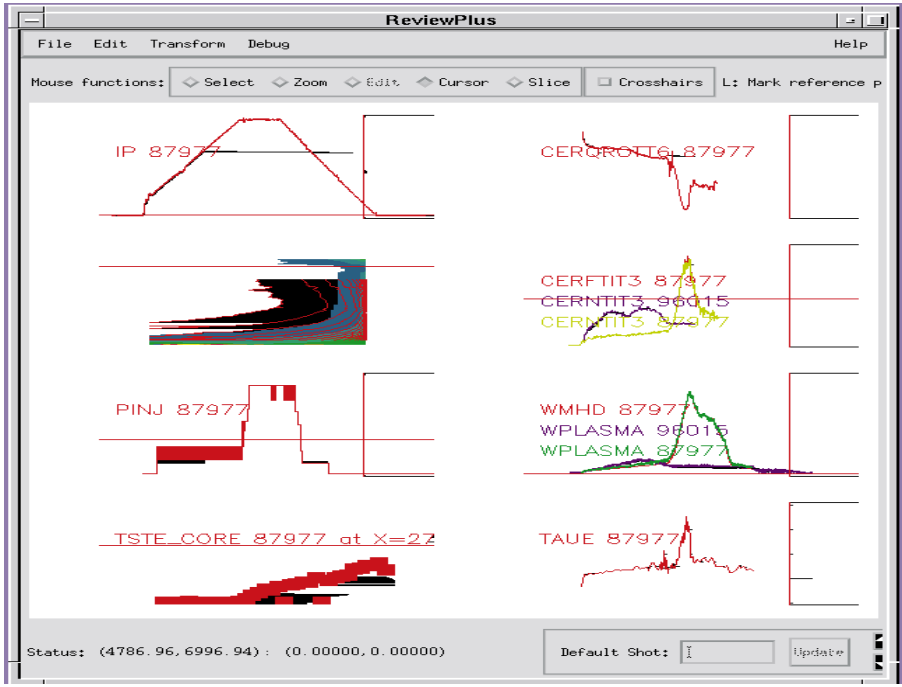
GENERAL ATOMICS

COMPUTATIONAL SUPPORT OF DIII–D REMOTE COLLABORATORS

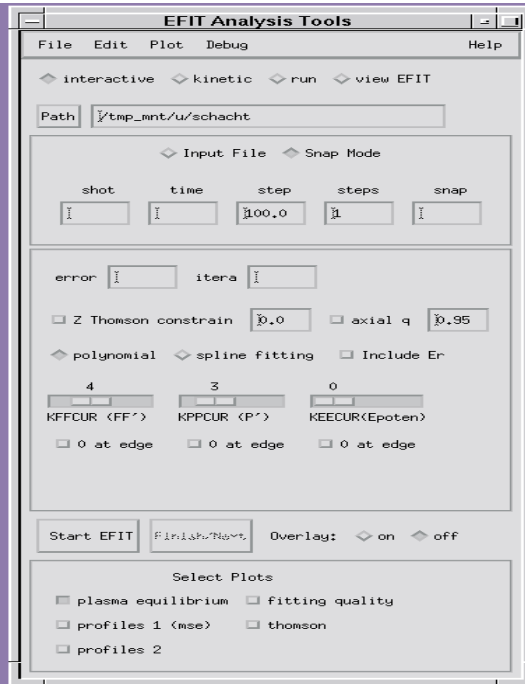
- Computer
- Access to data
- Tools to view and analyze data
  - Easy-to-use GUIs (IDL) with ample documentation
  - Tools run from remote CPU
- Communication with other scientists

GRAPHICAL USER INTERFACES SIMPLIFY DATA VIEWING AND ANALYSIS

ReviewPlus



EFITtools



- Same "look and feel" in all GUIs
  - Use GA-developed object oriented IDL based graphics library
- GUI data viewing/analysis tools: ReviewPlus, EFITtools, GAprofiles
- GUIs simplify use of FORTRAN analysis codes
  - EFIT, ONETWO
- Easier for collaborators to learn tools

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

WEB BASED DOCUMENTATION BRINGS CRITICAL INFORMATION TO THE ONSITE AND REMOTE COLLABORATOR

- **Computer code documentation**
  - By the scientific research staff
  - By the computer scientist staff
- **Data documentation**
  - Includes both raw digitizer data and analyzed data repositories
- **One source, multiple paths to reach the source**
  - Direct from Web browser
  - From within a tool

TOOLS RUN ON REMOTE CPUs

- **Can now run ReviewPlus, EFITtools, GAprofiles at LLNL, MIT**
  - Retrieve DIII–D data via MDSplus
  - Any site with IDL and MDSplus client can use
  - Some tools will run on MacOS, Win95/NT
- **Distributed EFITs (LLNL — Tom Casper)**
  - EFIT analysis on one shot shared by computers at DIII–D and LLNL

IDL IS LANGUAGE OF CHOICE FOR GUI TOOLS REQUIRING GRAPHICS

- **Benefits of IDL**
  - Extensive, powerful interactive graphics capabilities
  - Rich, easy to use language
  - Platform independent
- **Drawback: remote collaborators must purchase a license**
- **Public domain software (Java, Python) does not have required graphics**
  - Cost of implementing interactive features outweighs cost of IDL license
- **Adapt as technology grows**

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

FUTURE DATA VIEWING AND ANALYSIS TOOL ISSUES

- **Running the same code on different experiments**
  - How to make code tokamak-independent? (e.g. EFIT)
- **Simultaneous access to data from different experiments**
  - Possible with MDSplus and most relational databases
  - ReviewPlus does it now
  - Problem — what are the data signal names?
- **Sharing code development**
  - How to manage concurrent development?
  - Dialog begun between PPPL, MIT, LLNL, and GA

COMPUTATIONAL SUPPORT OF DIII–D COLLABORATORS

- Computer
- Access to data
- Tools to view and analyze data
- Communication with other scientists
  - Both in real-time and off-line

LLNL HAS TAKEN LEAD IN SUPPORTING REAL-TIME REMOTE COMMUNICATION AT DIII–D

- **Real-time = live (especially during experiment operation)**
  - IRC
  - Internet audio/video
  - Need more efficient tools
- **Off-line communication**
  - Web-based documentation
  - Documentation stored in MDSplus
  - Future: electronic logbook (based on MIT's)

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS

SUMMARY: DIII–D NATIONAL FUSION FACILITY IS AN ACTIVE, FRUITFUL COLLABORATION

- **Well-established plan to support collaborator computing**
  - Leveraging remote CPUs
  - 7x24 remote data access via MDSplus, PTDATA
  - Easy-to-use GUIs for data viewing and analysis
  - Remote real-time communication with DIII–D
- **Future will be built on current success**
  - Remotely accessible relational database and tools
  - Explore the benefits of object relational technology
  - Cooperative use and development of tools
  - More effective real-time interaction (LLNL)

DIII–D NATIONAL FUSION FACILITY SAN DIEGO

GENERAL ATOMICS