



ENHANCED COMPUTATIONAL INFRASTRUCTURE FOR DATA ANALYSIS AT THE DIII-D NATIONAL FUSION FACILITY

Presented by

D.P. Schissel

for the DIII-D National Team

Presented to

41st Annual Meeting of the Division of Plasma Physics

November 15-19, 1999

Seattle, WA

ACKNOWLEDGMENTS

- **The Data Analysis Applications Group**
 - Qian Peng, Jeff Schachter, Dave Schissel from GA
 - Ted Terpstra from PPPL
- **DIII-D Computer Staff**
 - J. Freeman, K. Keith, B. McHarg, and C. Parker
- **The DIII-D User Community**
- **From LLNL — Tom Casper, Ray Jong, 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

ANL
 INEL
 LANL
 LLNL
 ORNL
 PNL
 PPPL
 SNLA
 SNLL

Industry Collabs

CompX
 CPI (Varian)
 GA
 Gycom
 Orincon

Universities

Alaska
 Alberta
 Cal Tech
 Chalmers U.
 Columbia U.
 Hampton U.
 Helsinki
 Johns Hopkins U.
 Lehigh
 MIT
 Moscow State U.
 RPI
 U. Maryland
 U. Texas
 U. Wales
 U. Washington
 U. Wisconsin
 UCB
 UCI
 UCLA
 UCSD

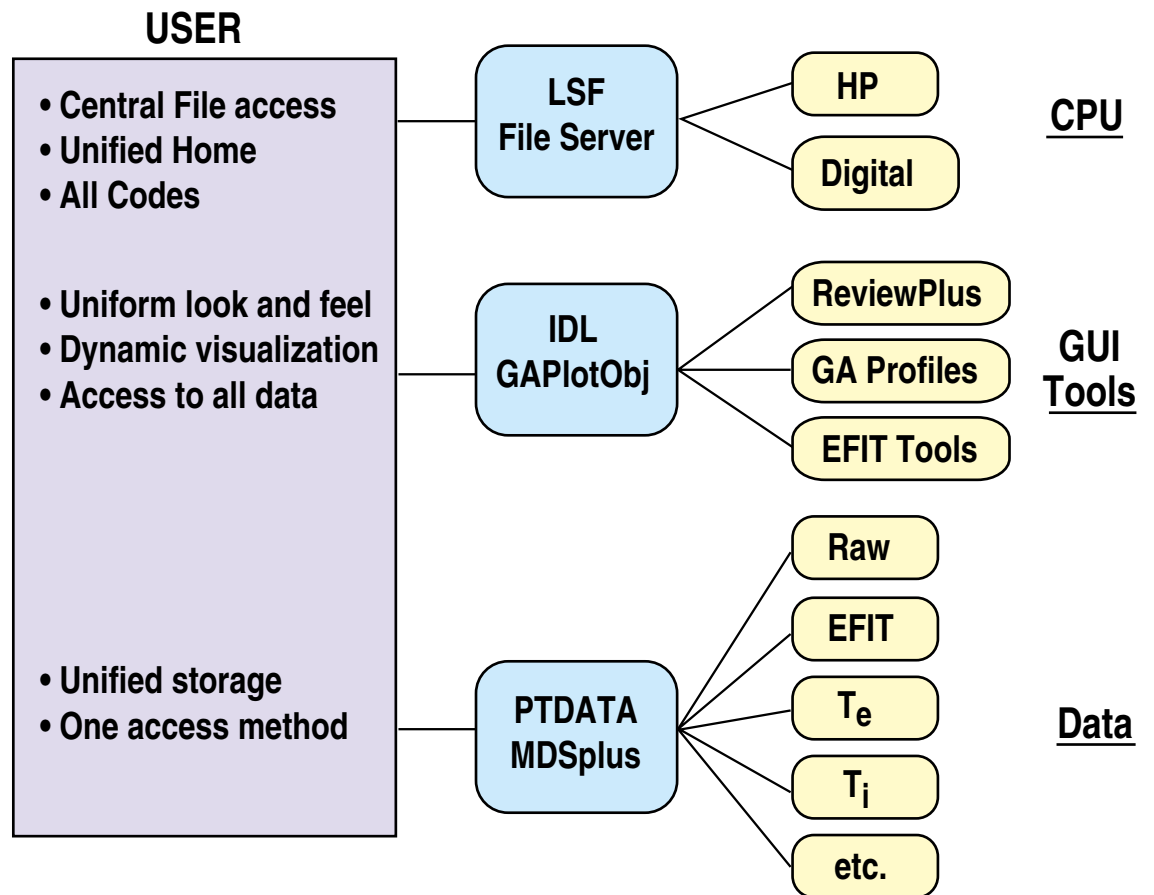
International Laboratories

Academia Sinica (China)
 Cadarache (France)
 CCFM (Canada)
 Culham (England)
 FOM (Netherlands)
 Frascati (Italy)
 Ioffe (Russia)
 IPP (Germany)
 JAERI (Japan)
 JET (EC)
 KAIST (Korea)
 Keldysh Inst. (Russia)
 KFA (Germany)
 Kurchatov (Russia)
 Lausanne (Switzerland)
 NIFS (Japan)
 Troitsk (Russia)
 Southwestern Inst. (China)
 Tsukuba U. (Japan)

INCREASE DIII-D DATA ANALYSIS THROUGHPUT AND DATA RETRIEVAL RATE BY EASE OF USE

- Underlying philosophy is uniformity

- Look and feel of GUI tools
- Access methods to analyzed datasets
- Access to existing computer power



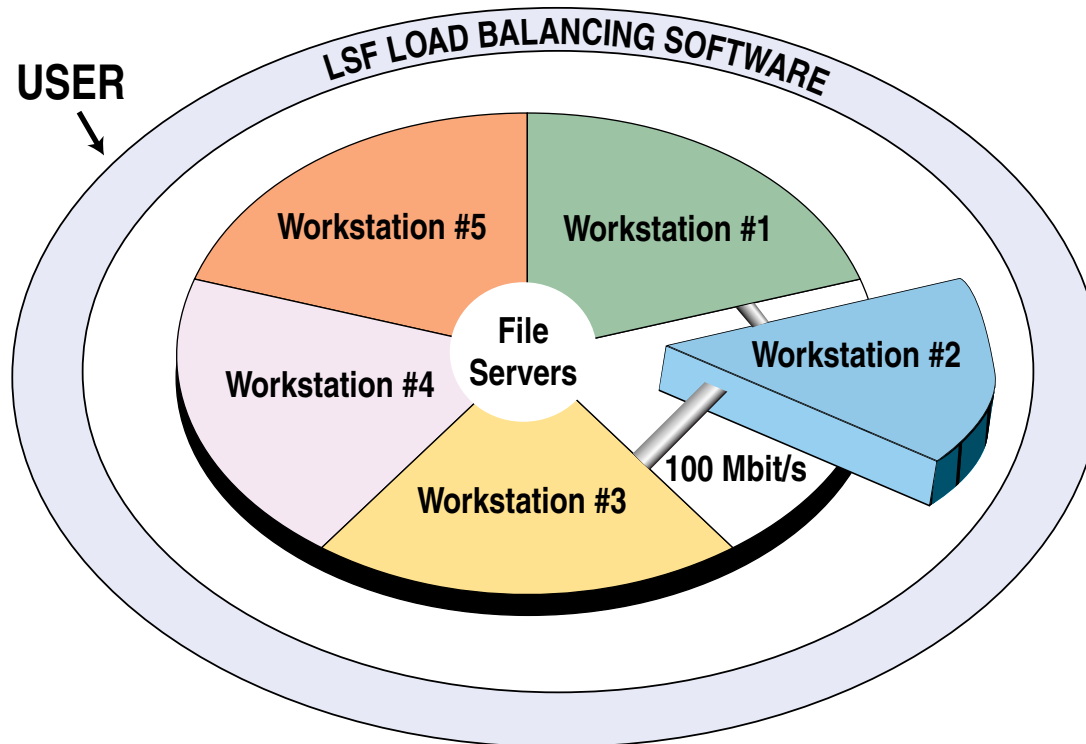
COMPUTATIONAL SUPPORT OF THE DIII-D RESEARCH TEAM

- **Computer**
 - Efficiently load balance onsite CPUs
 - Leverage use of remote CPUs
- **Access to data**
 - 7 day-24 hour (24/7) 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**
 - Remote meetings and experimental operations

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

LOAD BALANCING SOFTWARE PROVIDES SUBSTANTIAL COMPUTATIONAL POWER TO ONSITE USERS

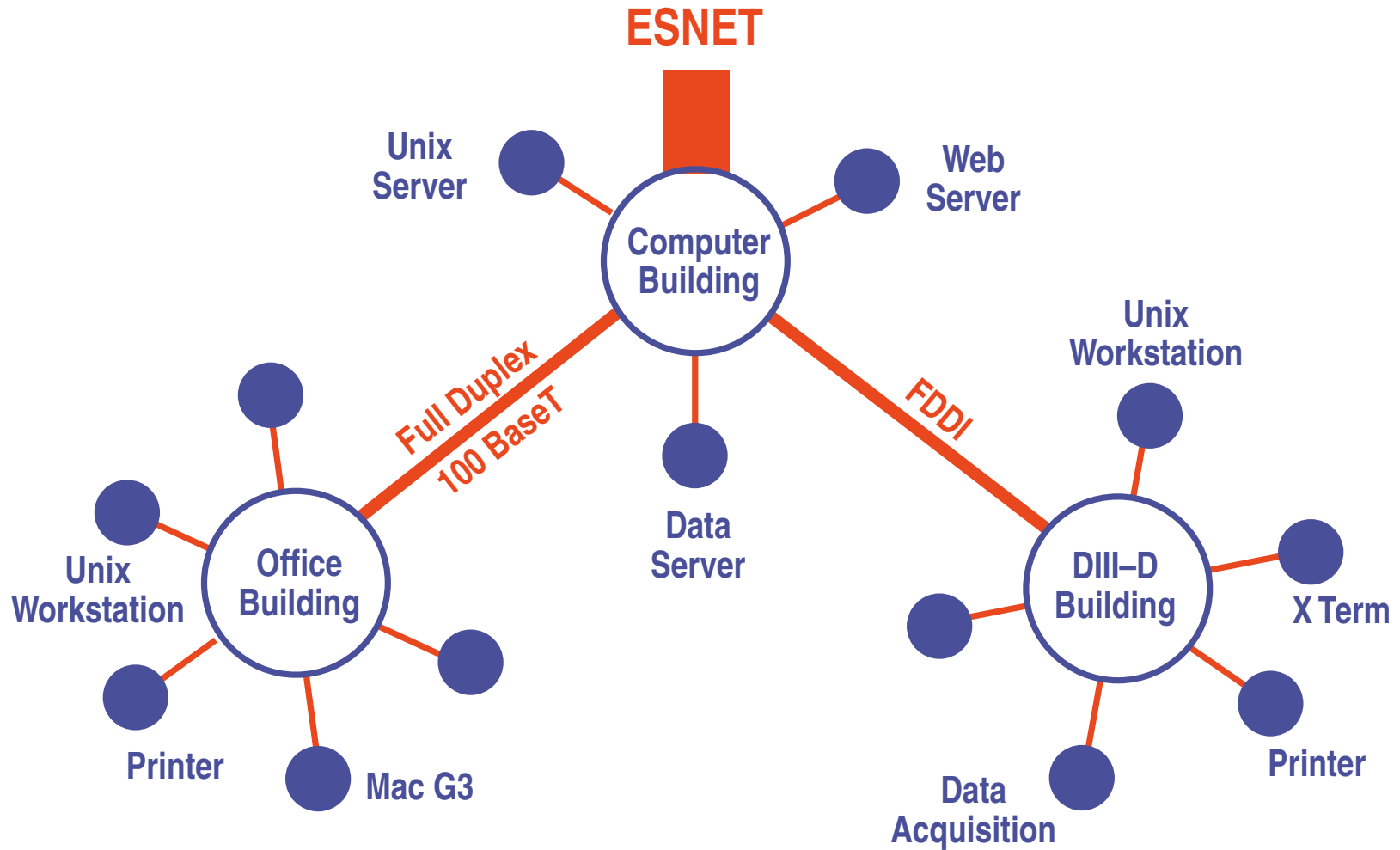
- Heterogeneous Unix environment with a large server and numerous workstations
- Platform Computing's LSF Suite performs interactive load sharing
- Cost effective for sharing CPUs (GA, LLNL, ORNL, U. Wisc.) and commercial software
- Central file servers for data and user files with fast network access



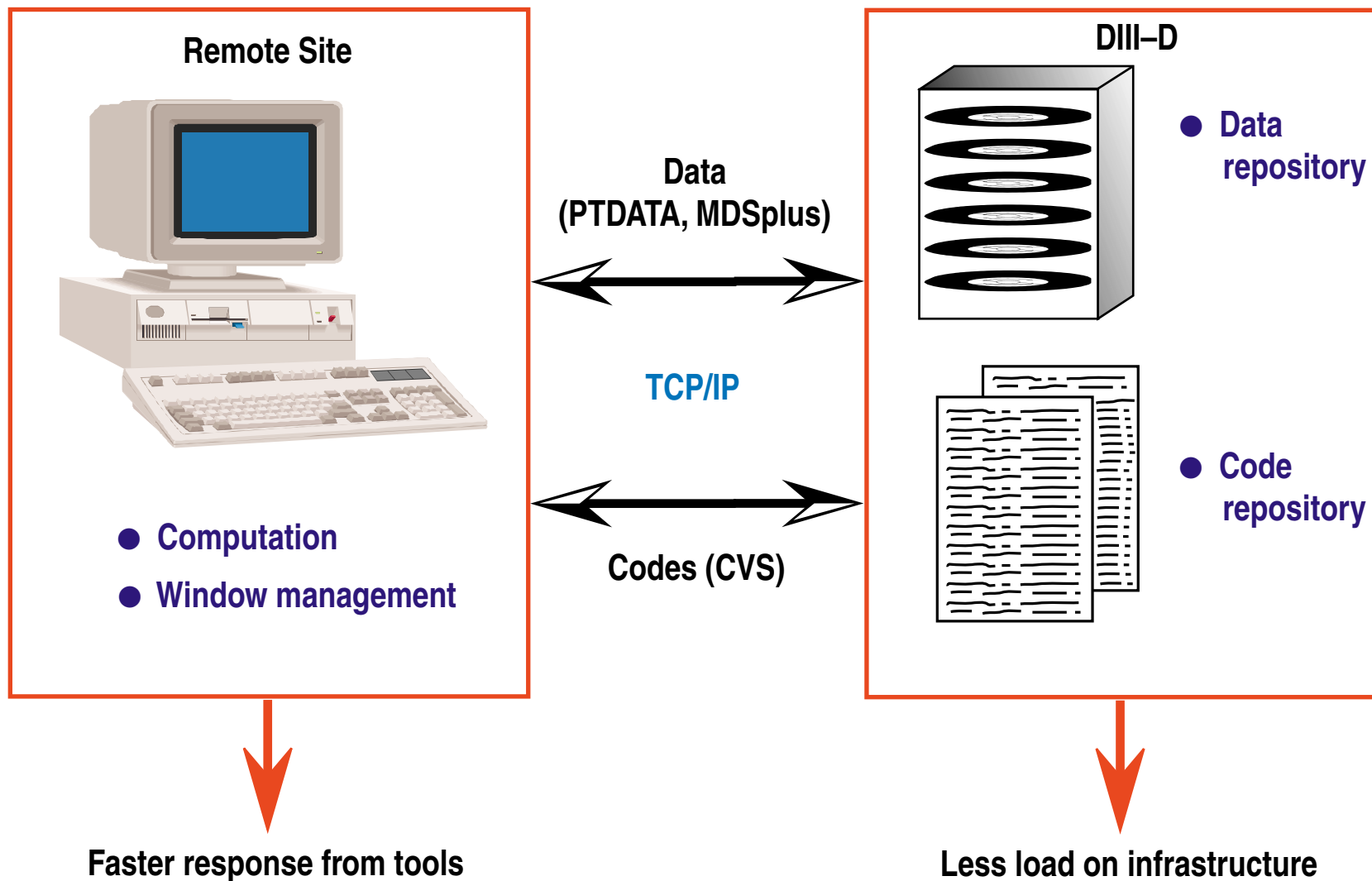
SUBSTANTIAL COMPUTER HARDWARE FOR DATA ANALYSIS

- **Large Unix Server – HP 9000 T600 Enterprise Server (45 SPECfp95 total)**
 - 3 CPUs with a total of 1GB RAM
 - Over 350 accounts including onsite and off-site collaborators
- **HP 9000 based Unix workstations (42 SPECfp95 total)**
 - Two C180 (each 18 SPECfp95) and one 735/125 (4 SPECfp95)
- **DEC 433au Unix personal workstations (113 SPECfp95 total)**
 - 5 machines placed on the desktop but in the LSF load balanced environment
 - Each workstation has a speed of 18 SPECfp95
- **In FY00 HP computer power will be doubled to 163 SPECfp95**
 - Three C240 workstations (each 25 SPECfp95) will be added to LSF cluster

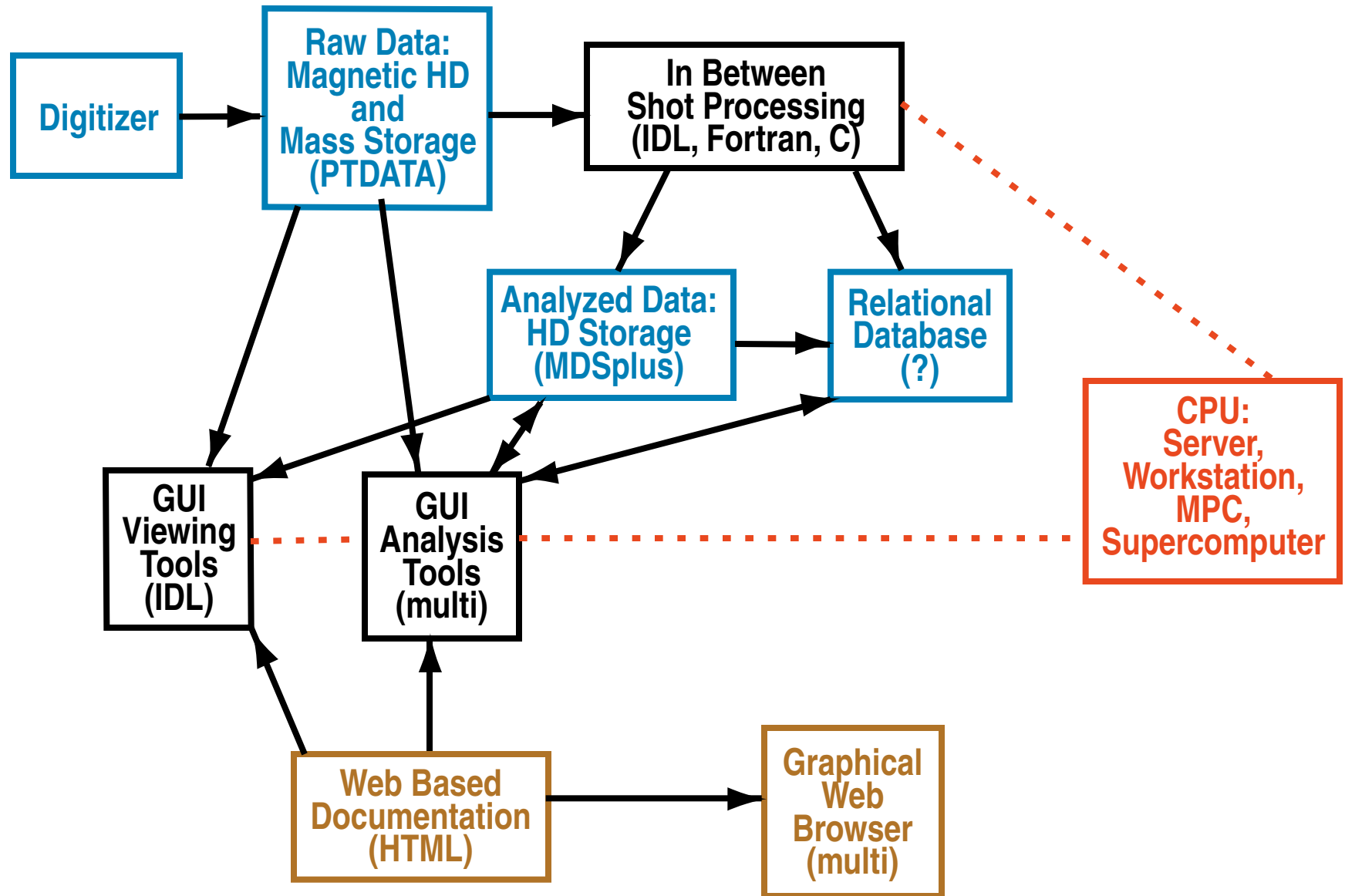
THE UPGRADED STAR NETWORK TOPOLOGY HAS AIDED UNIFICATION OF THE COMPUTING ENVIRONMENT



COMPUTATIONAL ARCHITECTURE ALLOWS COMPLETE DATA ANALYSIS WITH OFF-SITE COMPUTERS

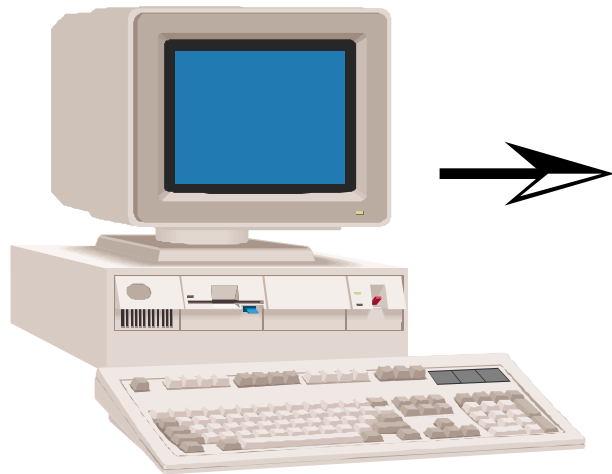


DIII-D DATA FLOW CHART



ALL DIII-D DATA AVAILABLE AROUND THE CLOCK

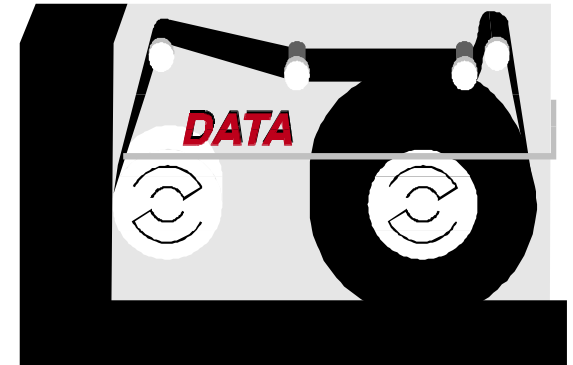
- All raw shot data (PTDATA) automatically available (24/7) for data mining
 - 820 GB interactively (15–30 sec minimum), ≈10000 raw compressed shots
 - 3.3 TB on a 2–5 minute minimum time scale per request
 - Present needs are 1.5 TB with about 0.3 TB added per year
- All analyzed shot data (MDSplus) on magnetic hard drive



Sun Ultra-10
200 GB Magnetic-Raid 5



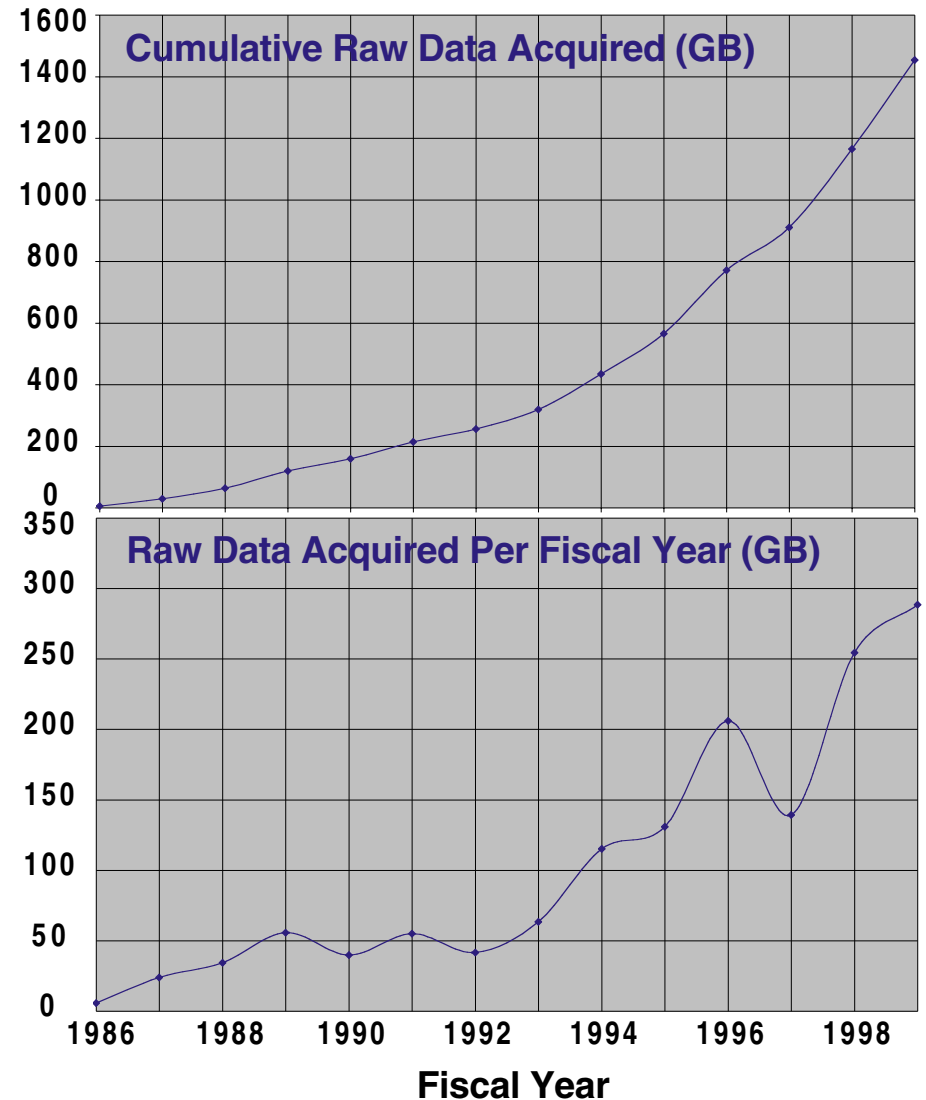
HP Magneto-optical 600fx
620 GB from 238 platters
2.3 GB/platter, 6 drives



ATL DLT 7000 drives
3.3 TB from 100 bays
35 GB/tape, 7 drives

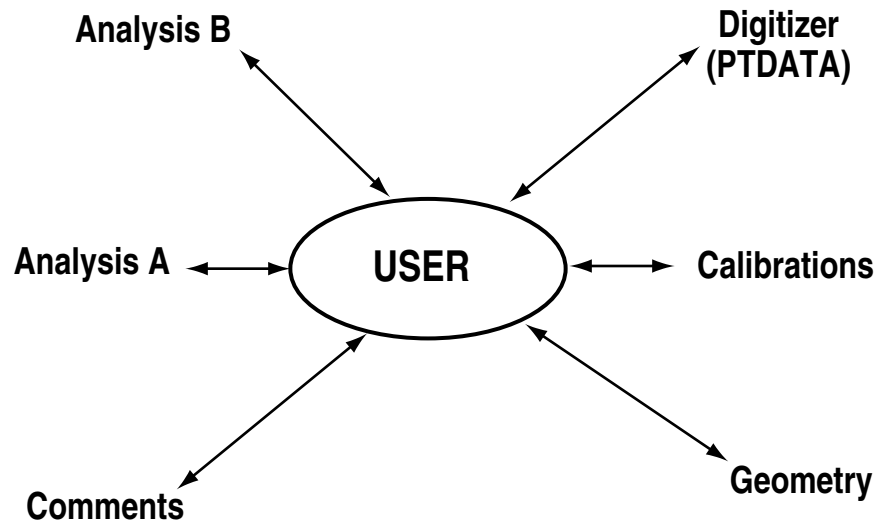
RAW DATA ACQUIRED PER FY CONTINUES TO INCREASE

- Compressed PTDATA files
- Quadratic increase in total raw data
- Extrapolating existing data shows a total of 3 TB in FY 2005
 - Will need a new storage system



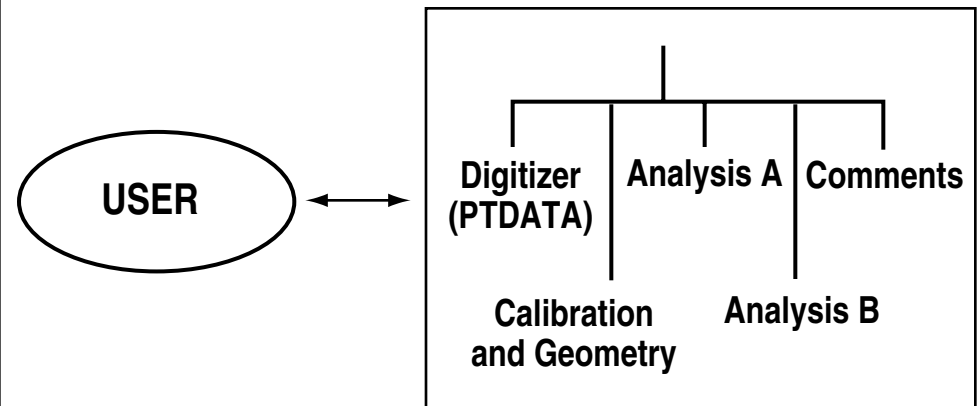
MDSplus UNIFIES DATA ACCESS

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**

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

- **Utilized to unify data storage and retrieval**
 - All 1998 & 1999 shots and selected older ones for a total of 4730
 - 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 and users' EFITs
 - CER and Thomson scattering data only available through MDSplus
 - Confinement related quantities, impurity density, auxiliary heating
- **DIII-D now operating with Unix MDSplus server**
 - Approximately 45 GB of data is being served
 - Digital Unix AlphaServer 800,100 GB of RAID 5 expanding to 300 GB
 - Unix port completed March 1999 (GA, LLNL, MIT, PPPL)
 - Exercised fully during June and July operations period

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

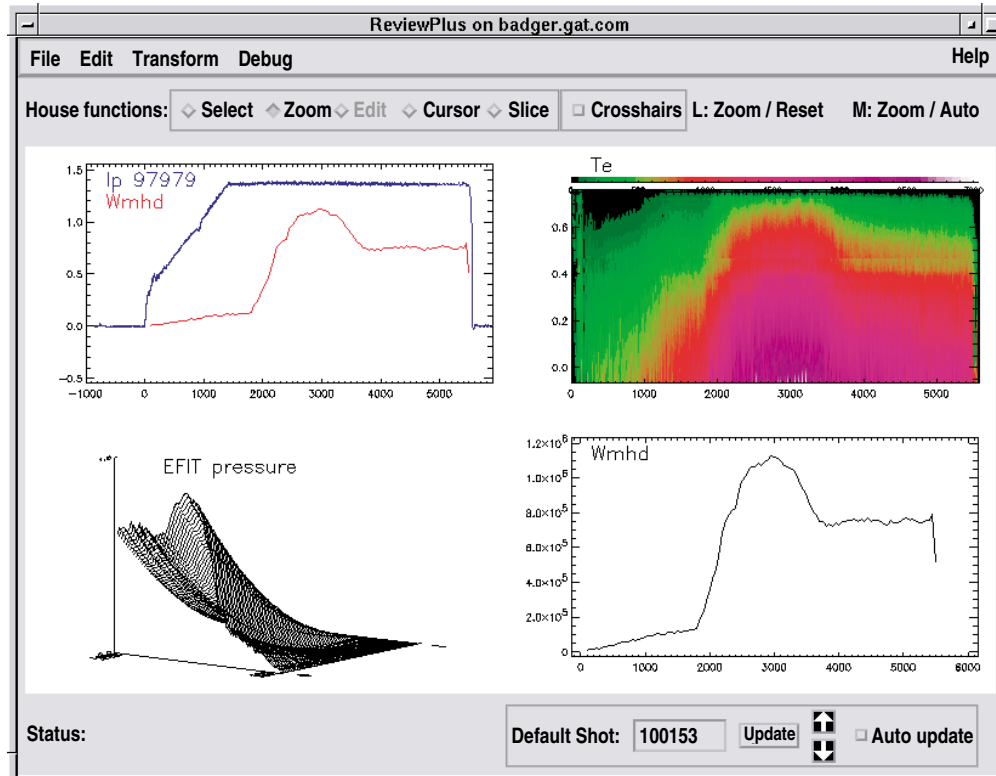
- **DIII-D raw and analysis data available without a local DIII-D computer account**
 - For PTDATA access DIII-D can supply the software
 - For MDSplus access the MIT group can supply the software
 - MDSplus data access security is by username and Internet address
- **Collaborators are offered full access 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 posted on the Web either at DIII-D or on the collaborator's site**

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

- **New system will work in concert with MDSplus**
- **Intel based system with Microsoft SQL Server 7**
 - 0.5 GB RAM, dual power supply, RAID 1, 100 BaseT
- **Use existing solutions to minimize effort (MIT)**
 - Electronic Notebook is operational
 - Run Management, Signal Documentation, Summary Database to follow
- **Multi-platform GUI tools connected to database engine**
 - Relational queries from the Web, Fortran, C or IDL
- **Searching for an Applications Programmer to implement databases**
 - Job is posted at this meeting

GRAPHICAL USER INTERFACES SIMPLIFY DATA VIEWING AND ANALYSIS

ReviewPlus



EFITtools

The EFITtools GUI displays various analysis parameters and plot selection options:

- Menu Bar:** File, Edit, Plot, Debug, Help.
- Tools:** interactive, kinetic, run, view EFIT.
- Path:** /f/mdsplus/jeff/testidl/reviewplus
- Input File / Snap Mode:**
 - shot: 97979
 - time: 4000
 - step: 100.0
 - steps: 1
 - snap: []
- Parameters:**
 - error: .005
 - itera: 50
 - Z Thomson constrain: 0.0
 - axial q: 0.95
 - polynomial: []
 - spline fitting: []
 - Include Er: []
- Profiles:**
 - 4: KFFCUR (FF')
 - 3: KPPCUR (P')
 - 0: KEECUR(Epol)
- Options:**
 - 0 at edge: []
 - 0 at edge: []
 - 0 at edge: []
- Buttons:** Start EFIT, Finish/Next
- Overlay:** on, off
- Select Plots:**
 - plasma equilibrium: []
 - fitting quality: []
 - profiles 1 (mse): []
 - ne,te,ti,vrot: []
 - profiles 2: []

- Same "look and feel" in all GUIs (IDL based GAPIotObj)
- Viewing/Analysis tools: ReviewPlus, EFITtools, Gaprofiles
- GUIs simplify use of FORTRAN analysis codes

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 activating the web browser

TOOLS RUN ON REMOTE CPUs

- **Can now run ReviewPlus, EFITtools, GAprfiles at LLNL, MIT, PPPL**
 - Retrieve DIII-D data via MDSplus
 - Any site with IDL and MDSplus client can use
 - Some tools will run on MacOS, Win95/NT
- **EFIT analysis between pulses is distributed for faster processing**
 - EFIT analysis on one shot shared by computers at DIII-D and LLNL

IDL IS LANGUAGE OF CHOICE FOR GUI TOOLS REQUIRING SIMPLE 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 software technology evolves**
 - e.g., A more robust Java with required graphics

ADVANCED DESKTOP GRAPHICS IS BEING PURSUED

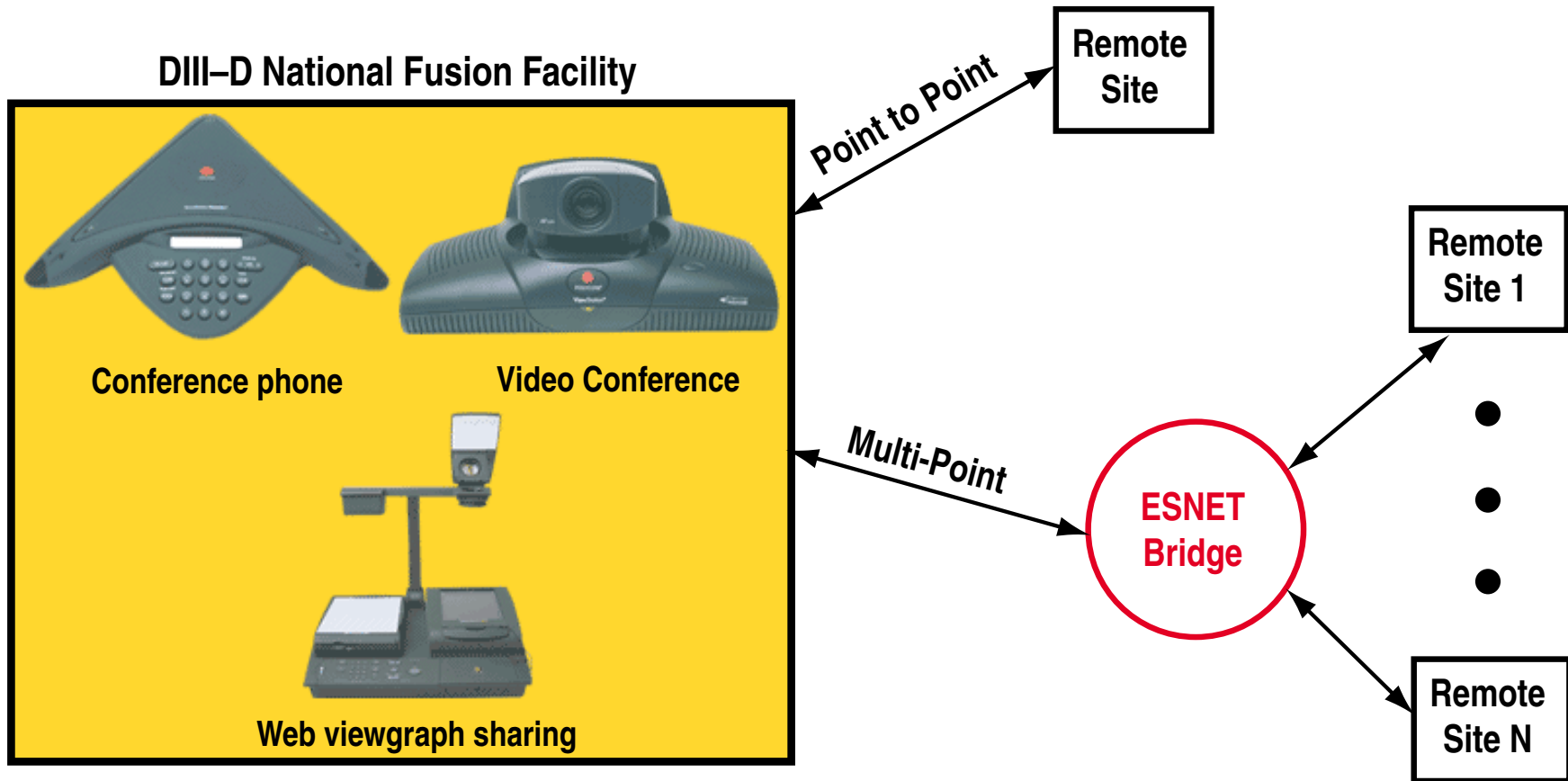
- **Hardware rendered graphics on the scientists' desktop**
 - Requires a graphics card with OpenGL support
- **Our existing X-Terminals have their limitations**
 - 10 BaseT, 8 bit color, no hardware rendering
- **IDL's Object Graphics**
 - Their interpretation of OpenGL
 - Surface fitting
- **Partnership with UC Davis CS Department**
 - 3D visualization of tokamak flux surfaces
- **Visualization Toolkit (VTK)**
 - Mutli-dimensional space problem

COLLABORATIVE ANALYSIS CODE DEVELOPMENT EFFICIENTLY UTILIZES EXISTING COMPUTER SCIENCE RESOURCES

- **Good history of collaborative**
 - TRANSP from PPPL
 - MDSplus led by MIT
 - EFIT from GA
 - Distributed EFIT computing from LLNL

- **Exploring new areas of collaboration (GA, MIT, LLNL, ORNL, and PPPL)**
 - IDL-based transport analysis results display tool
 - IDL-based data preparation software
 - Name translation service
 - Run management database
 - Database analysis applications and database tool kit

THE DIII-D NATIONAL FUSION FACILITY HAS A NEW ELECTRONIC CONFERENCE ROOM EQUIPPED FOR MULTI-SITE REMOTE MEETINGS



SUMMARY: INCREASED DATA ANALYSIS THROUGHPUT & DATA RETRIEVAL RATE AT THE DIII-D NATIONAL FUSION FACILITY

- **Uniform interface to a very heterogeneous environment**
 - Easy-to-use GUIs for data viewing and analysis
 - Efficiently utilize onsite and remote CPUs
 - 24x7 remote data access via MDSplus, PTDATA
 - Interactive remote communication with onsite DIII-D staff
- **Future will be built on current success**
 - Promote collaborative development of software
 - Remotely accessible relational database and tools
 - More effective remote communication