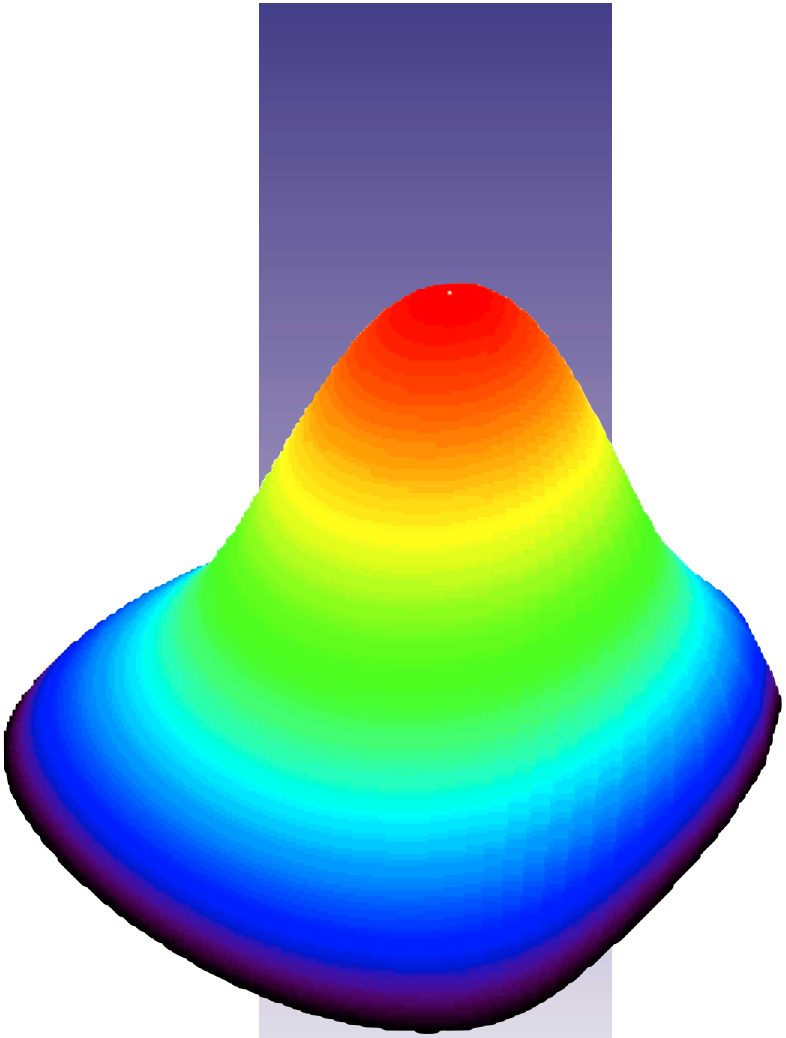


NIMROD simulated pressure stored in MDSplus and visualized with IDL

DATA MANAGEMENT AND VISUALIZATION TO ENHANCE SCIENCE DISCOVERY THROUGH ADVANCED COMPUTING IN THE PSACI PROJECT

D.P. Schissel, J. Schachter, General Atomics

S. Kruger, SAIC

X. Tang , Princeton Plasma Physics Laboratory

Presented to

42nd Annual Meeting of the Division of Plasma Physics

October 23–27, 2000

Quebec City, Canada

GENERAL ATOMICS

OUR GOAL IS TO MAXIMIZE THE EFFECTIVENESS IN EXTRACTING SCIENTIFIC UNDERSTANDING IN PSACI

- Advanced computing to solve critical physics problems requires a capable infrastructure and the necessary tools
- Allow the scientists to concentrate on their science, not computer science
  - Provide them with their required infrastructure
- Strengthen coupling between modeling teams & theory and experiment
- Minimize duplication of effort

PLASMA SCIENCE AND ADVANCED COMPUTING INITIATIVE (PSACI) RESEARCH PROGRAM

- Funded by the USDOE/OFES during FY00
  - Compliments new DOE Office of Science Initiative "Science Discovery through Advanced Computing" (SDAC) which has replaced SSI
- Enhance physics capabilities in scientifically advanced simulation codes
- Research, development, & deployment of better mathematical models & computational methods for optimal utilization of modern supercomputing
  - Parallel programming for scalability of modern MPP's
  - Advanced visualization for higher–dimensionality data
  - Object–oriented architecture for better community access
- Build advanced, shared diagnostics to provide a better bridge between simulation, theory & experimental communities

WE HAVE TEAMED WITH THE TWO FY00 PSACI PILOT PROGRAMS TO ENHANCE THEIR SCIENTIFIC PRODUCTIVITY

- **Microscopic Turbulence and Transport Simulation (W. Nevins as PI)**
  - With LLNL, PPPL, U. Maryland, GA, UCLA, U. Colorado
  - Mutually benchmarked, well diagnosed, microturbulence codes
  - Better understanding of turbulent transport to aid interpretation & planning of experiments
- **Macroscopic Simulation of Fusion Plasmas (S. Jardin as PI)**
  - With PPPL, SAIC, LANL, GA, U. Wisconsin, NYU, MIT, U. Colorado, SNL
  - Scalable 3D nonlinear MHD simulation capability with a wide user base
  - Mutually benchmarked comprehensive physics model with efficient & accurate solutions
- **Advanced data analysis and visualization capabilities**
  - Prototype national data repository for storing code output
  - Prototype visualization to be used as routine part of data analysis
- **Firm base for further expansion with possible support from new DOE SDAC**

ACHIEVING OUR GOAL TO SUPPORT PSACI REQUIRES A UNIFIED EFFORT TO ADDRESS COMPUTER SCIENCE ISSUES IN SUPPORT OF ALL FUNDED PROGRAMS

- A common, shared, secure network–enabled interface to all data
  - All data means from both experimental and modeling sources
- Shared methodology for organization, archival, and retrieval of all data
- Development of data browsing and visualization applications
  - These will exploit the common data interface for wide applicability
- Development of common, shared tools for data manipulation & analysis
- Effective shared development and reuse of software components

MDSplus PROVIDES A COMMON, SHARED, SECURED NETWORK ENABLED INTERFACE TO ALL DATA

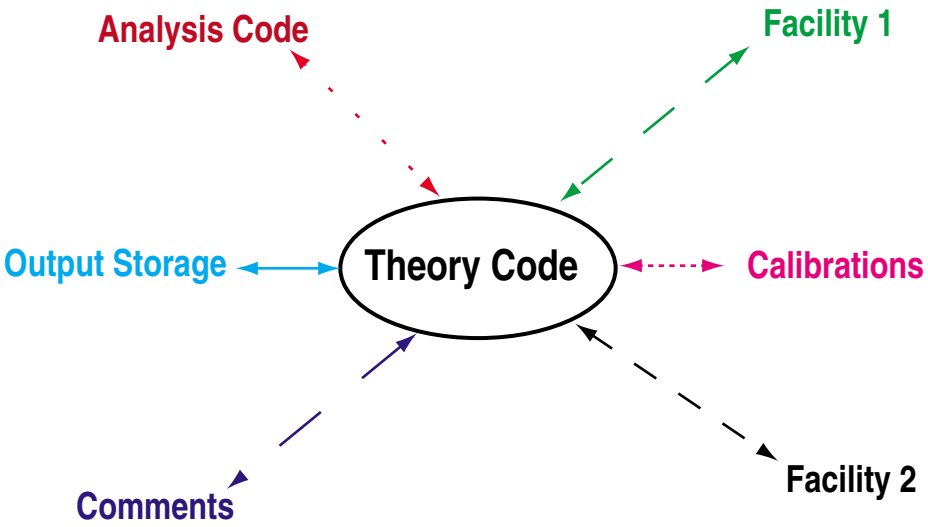
- MDSplus is a data system jointly developed by MIT, LANL, & Padova Italy
  - Provides for acquisition, storage, access, and organization of data
  - Secure client/server system utilizing TCP/IP
  - Can store simulation and theoretical data
- Presently used to serve data at 4 sites in US and 8 worldwide
  - Includes the 3 large US tokamaks
  - Clients at many sites
- Many physics analysis codes have been or are being integrated with MDSplus
  - Transport: MIST, ONETWO, TRANSP
  - MacroStability & Equilibrium: DCON, EFIT, M3D, NIMROD, PEST
  - MicroTurbulence: GS2

MDSplus CHARACTERISTICS

- Hierarchical and self descriptive
- Supports a variety of primitive data types
  - Byte, word, long, float, double, complex, string, expressions, actions
  - Built in expression evaluator (TDI)
- Remote access available currently from
  - Fortran, C/C++, Java, IDL, Matlab, and labview
- Tools exist for quick display of data and structure
  - X–windows and Java scope & traverser
  - IDL tools (ReviewPlus, Pslice, etc.)
  - Matlab tools
- Supported platforms (so far)
  - AIX, Cray Unix, Digital Unix, HP/UX, Irix, Linux, Mac OS, Sun OS, VMS, Win32 (windows 9x, NT, 2000)

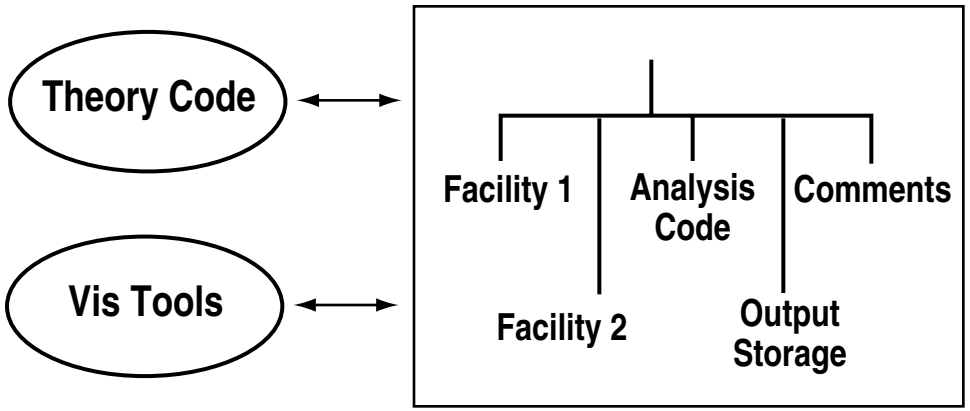
MDSplus CAN UNIFY DATA ACCESS FOR THE THEORETICAL COMMUNITY JUST AS IT HAS FOR THE EXPERIMENTAL COMMUNITY

Conventional Storage



- Each code needs it own interface
- Must know data format and file location
- Each code has its own graphics tool
- Hard to share results

MDSplus



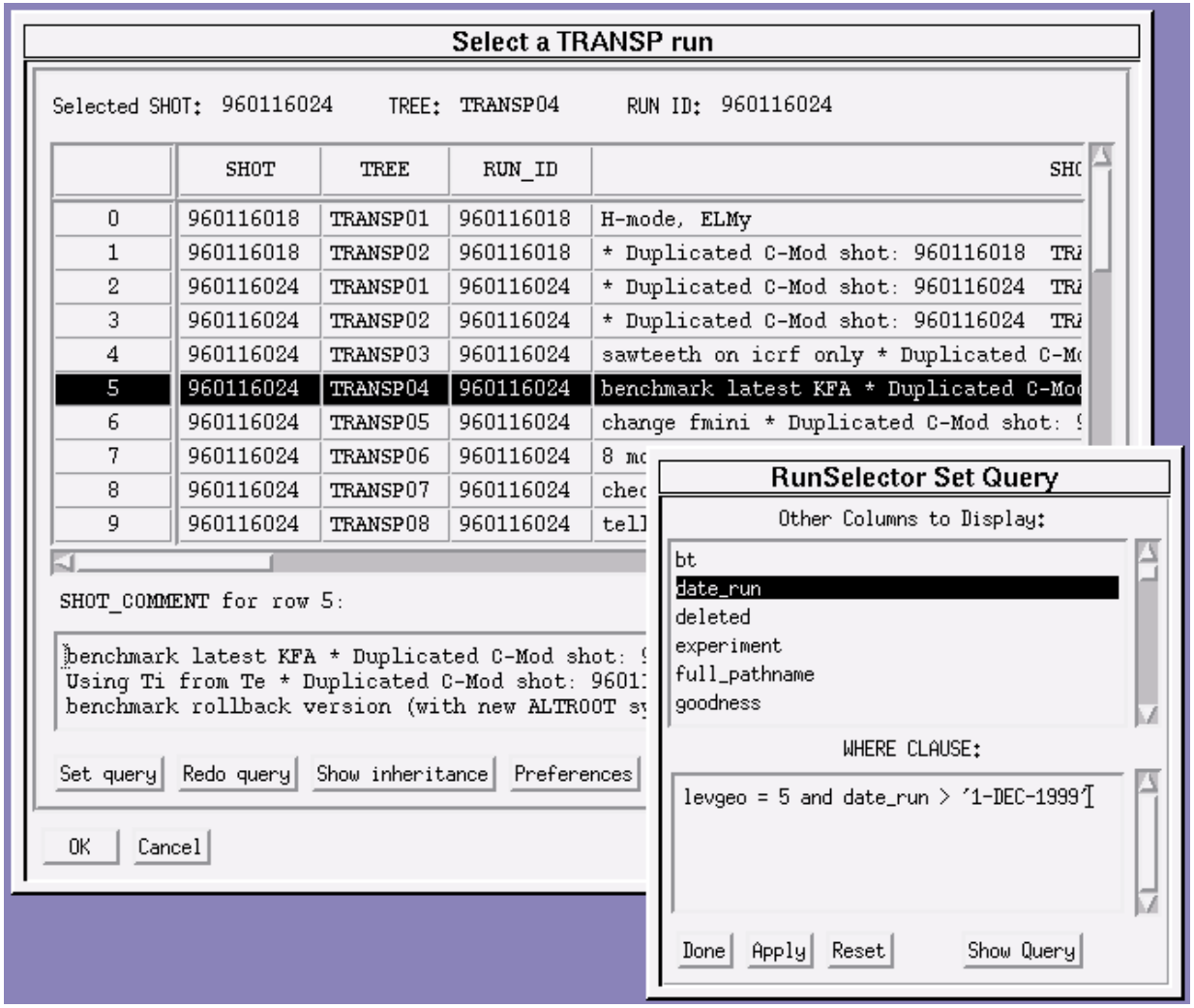
- One interface to many data types
- Only need location of data in tree
- Utilize existing visualization tools
- Easy to share results with both the theoretical and experimental community

GENERAL ATOMICS

A RELATIONAL DATABASE WORKS IN CONCERT WITH MDSplus

- **MDSplus stores all the data**
  - Not optimized for queries across multiple shots
  - Approximately 10 TB stored in the experimental programs
- **A relational database stores highlights of the data**
  - Optimized for queries
  - Drill down to a smaller dataset for more detailed examination
- **Essential requirements of a relational database**
  - Archival storage of highlights of the data
  - Keeps track of code runs (e.g. which NSTX pulses have NIMROD runs?)
  - Provides a convenient interface to the data visualization software
  - Efficient and flexible method to support schema evolution
  - Transparent access across the wide area network
- **Relational database is in use at DIII–D (see Burruss poster NP1.106)**
  - A system to track code runs has been implemented & is available to all collaborators
  - Highlights of DIII–D experimental data is also being stored

CODE RUN DATABASE LETS USERS TRACK PHYSICS ANALYSIS CODE RUNS



- GUI displays information from the code run database
- Provides a formalized method for tracking code runs
- Hierarchical organization
- Easy to integrate into existing tools

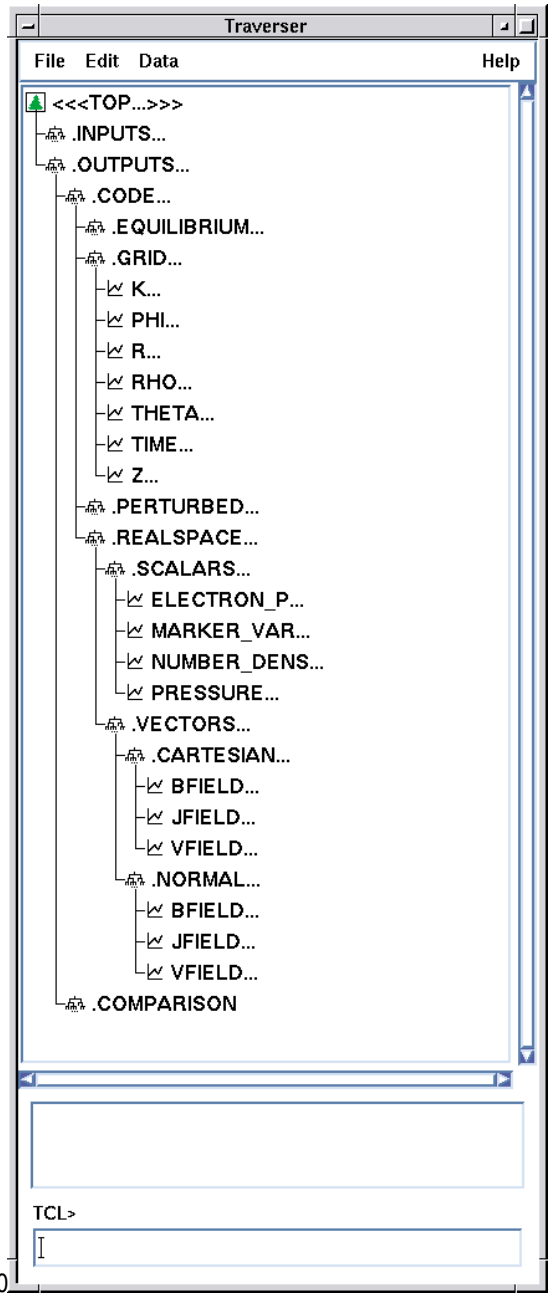
WORK SCOPE IS REDUCED BY LEVERAGING EXISTING EFFORT

- **Integrate simulation codes into the existing data sharing methodology**
  - MDSplus for storing all the data, MS SQL RDB for storing highlights
  - New storage for NIMROD and M3D have been created
- **Utilize existing DIII–D MDSplus server hardware for storing simulation data**
  - NIMROD and M3D data stored on the GA server
  - Work began immediately on the new NIMROD and M3D software
  - Allows accurate hardware specification to be set for a subsequent purchase
- **Utilize existing visualization tools for simulation data**
  - Generalized interactive plotting capability brings immediate benefit (ReviewPlus)
  - Tools more specific to simulation needs can be rapidly developed by using existing code libraries: NIMROD & M3D needs being determined

THE MDSPLUS SERVER AT THE DIII–D NATIONAL FUSION FACILITY

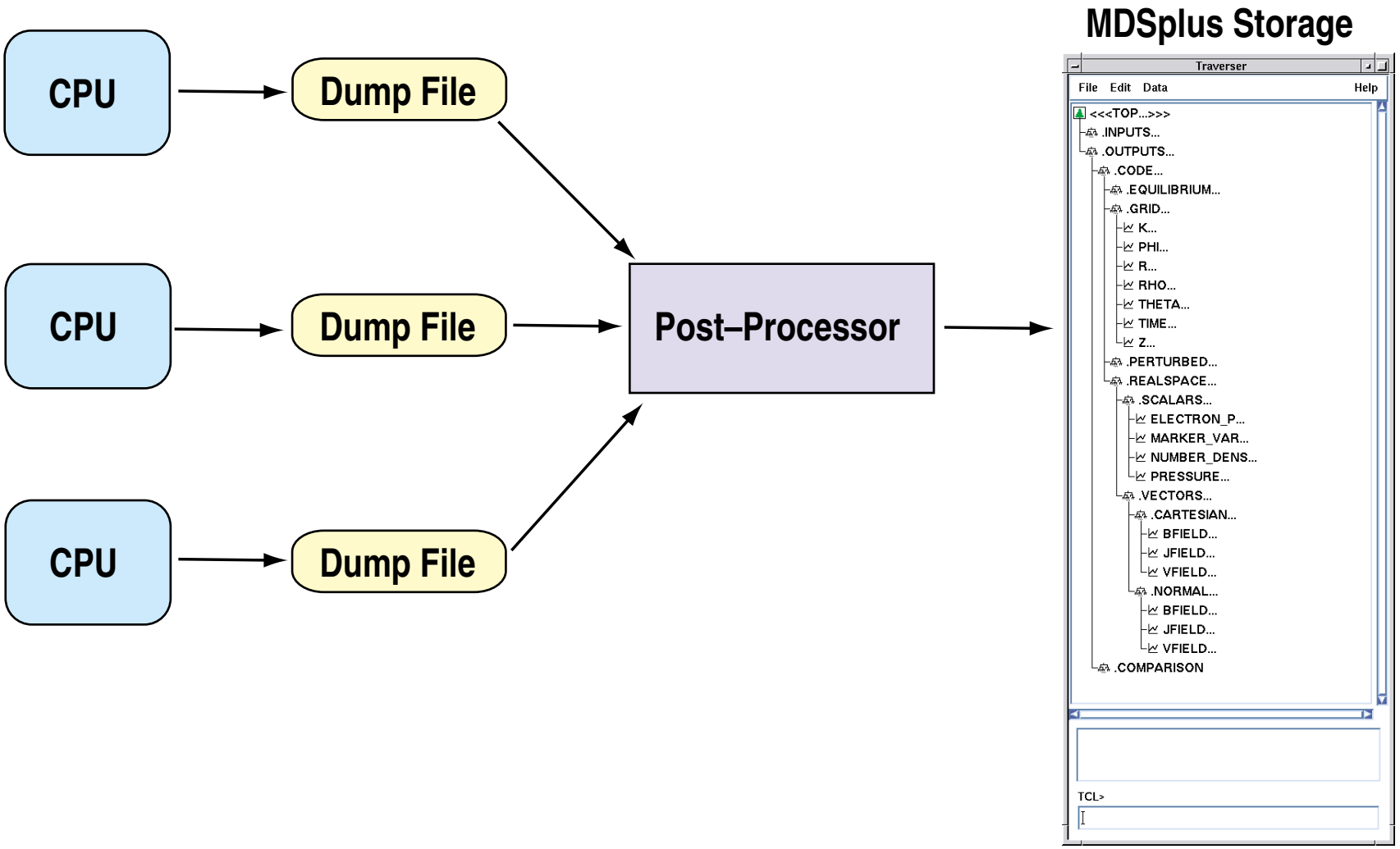
- Compaq AlphaServer 800 5/500 running True64 Unix
  - 512 MB RAM, 9 GB disk
  - 1999 purchase price of $11K
- nStor GigaRAID AA RAID Disk System
  - 100 GB at RAID 5 with a hot spare and dual–ultrawide SCSI
  - 1999 purchase price of $22K
- An additional 200 GB was added in 2000 for $10K
- This system heavily used at DIII—D
  - 100 on–site users, numerous off–site users
  - 8900 DIII–D pulses totaling 70 GB
  - Test system for NIMROD, M3D, and GS2

NIMROD AND M3D CODE RESULTS ARE BEING STORED IN MDSPLUS



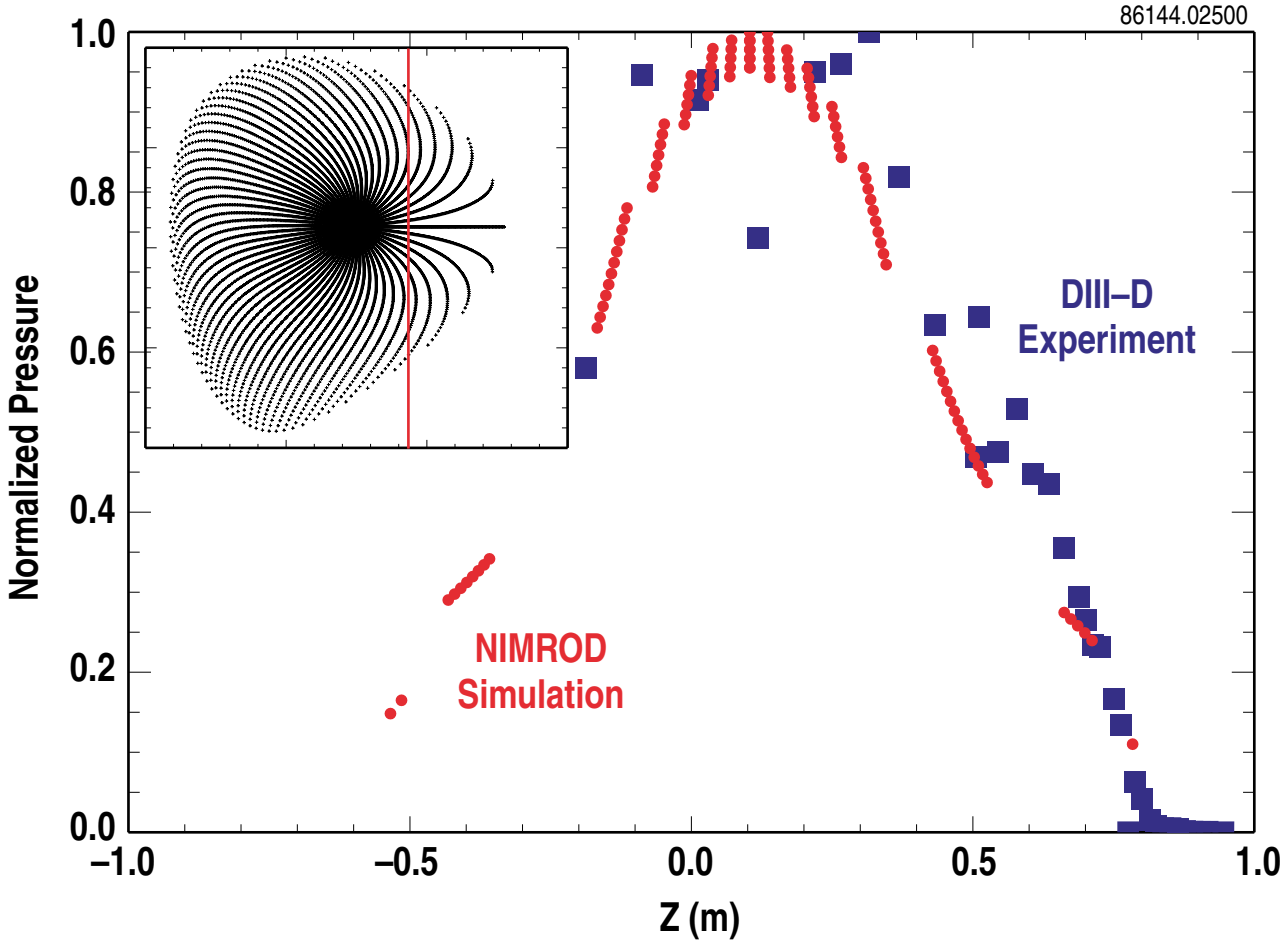
- **Both inputs and outputs can be stored**
  - Presently only outputs have been written
- **Structure of tree facilitates code comparison**
  - Same quantities on grid with same name
- **Several runs for NIMROD, one for M3D store**
  - Has been used for testing purposes
  - Users have writen in their own code runs
  - MDSplus client & ReviewPlus installed at SAIC
- **MDSplus storage provides a factor of 4 compression for M3D**
  - NIMROD is less due to the binary dump file which stores only 1 representation of the data

DATA IS WRITTEN INTO MDSplus FROM A POST–PROCESSOR



COMPARISON OF NIMROD SIMULATION TO DIII–D EXPERIMENTAL DATA IS FACILITATED WITH MDSplus

- NIMROD P_e overlayed onto DIII–D measurement
  - Synthetic diagnostics
- NIMROD data in MDSplus
  - Widely accessible
- Data easily accessible to visualization tools
  - These plots with IDL

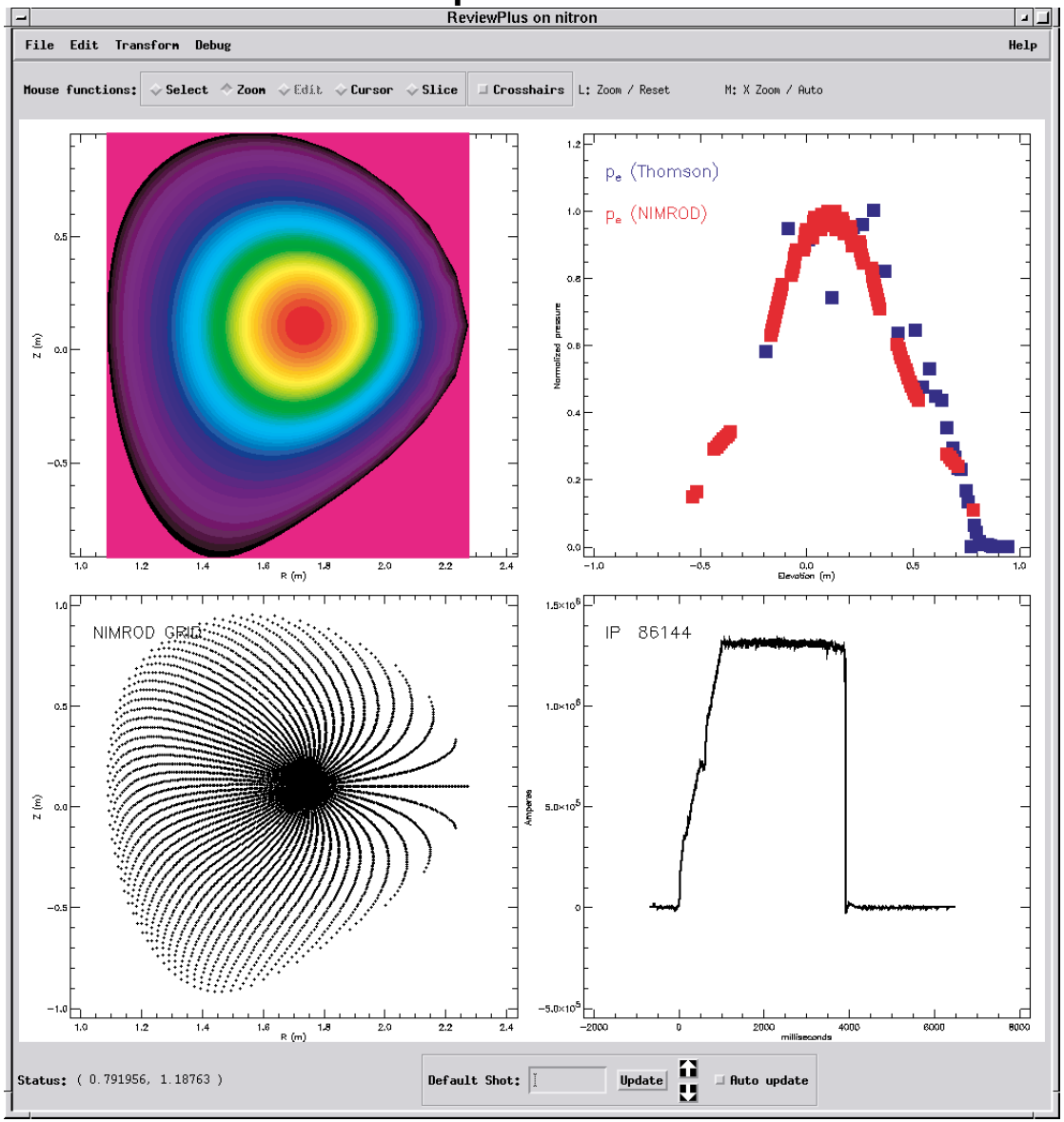


- With MDSplus the data is easy to retrieve & the plot is easy to create
  - Today other simulation data is hard to retrieve and therefore the plot is difficult
  - Requires phone calls and emails & takes a day to create

GENERAL ATOMICS

UTILIZE EXISTING VISUALIZATION TOOLS FROM THE EXPERIMENTAL COMMUNITY WHERE POSSIBLE

For example: ReviewPlus



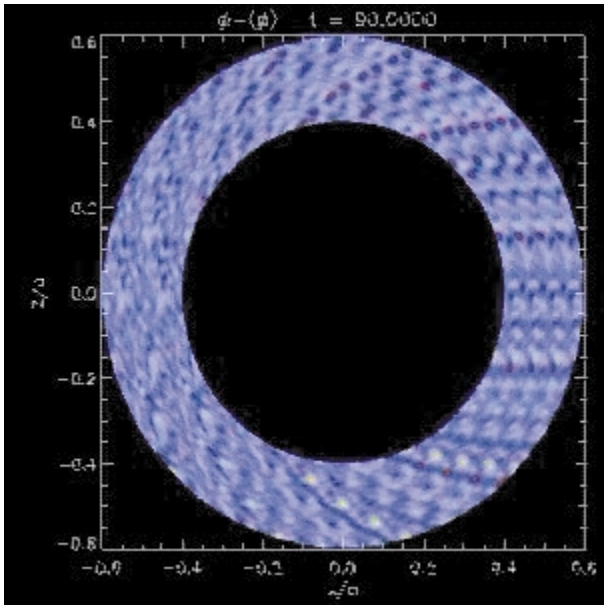
- Some tools are capable of plotting any data regardless of origin
  - Experimental & simulation data
- Any data stored in MDSplus is immediately available with these tools
- Simulation community gains rapid access to interactive visualization

GENERAL ATOMICS

SCIENTIFIC VISUALIZATION OF LARGE DATASETS

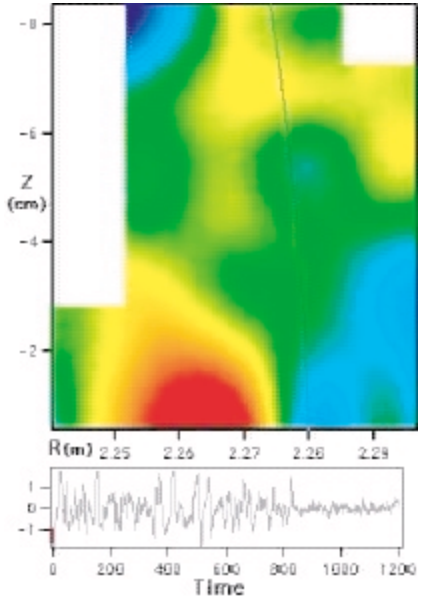
- Development of common, shared tools for data browsing, visualization, data manipulation, and analysis
- Create interactive multidimensional capability allowing browsing data in time, space, and context
- Ability to combine and use data from different repositories in an integrated product
- MDSplus API allows access from any visualization tool

Density Fluctuations



Microturbulence Simulation via the Gyro code - J. Candy

Density Fluctuations



2D Tokamak Edge Turbulence Data via BES – G. McKee

GENERAL ATOMICS

INTERACTIVE SCIENTIFIC VISUALIZATION OF MICROTURBULENCE DATA

- Using VTK with OpenGL on a Linux PC with hardware graphics rendering

Interactive Cutting Plan

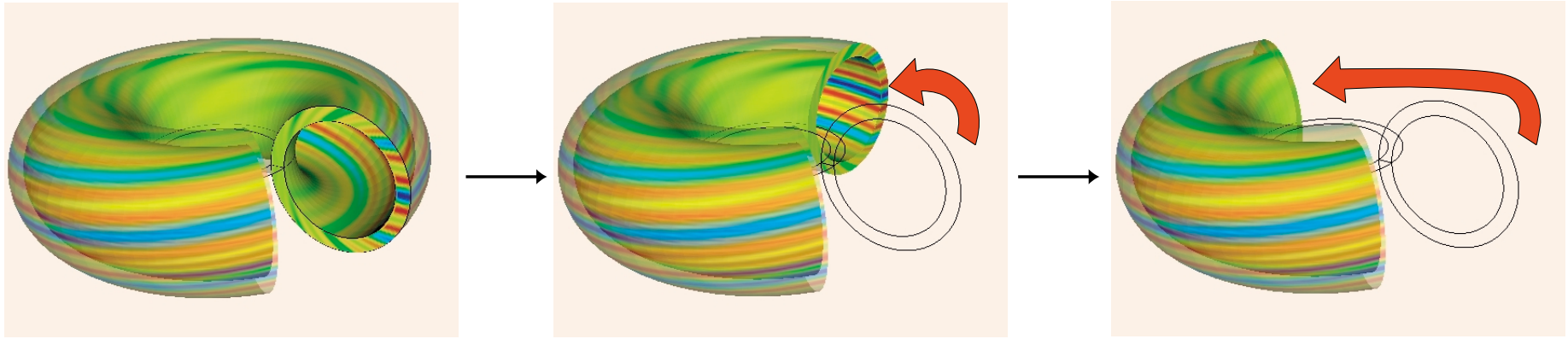
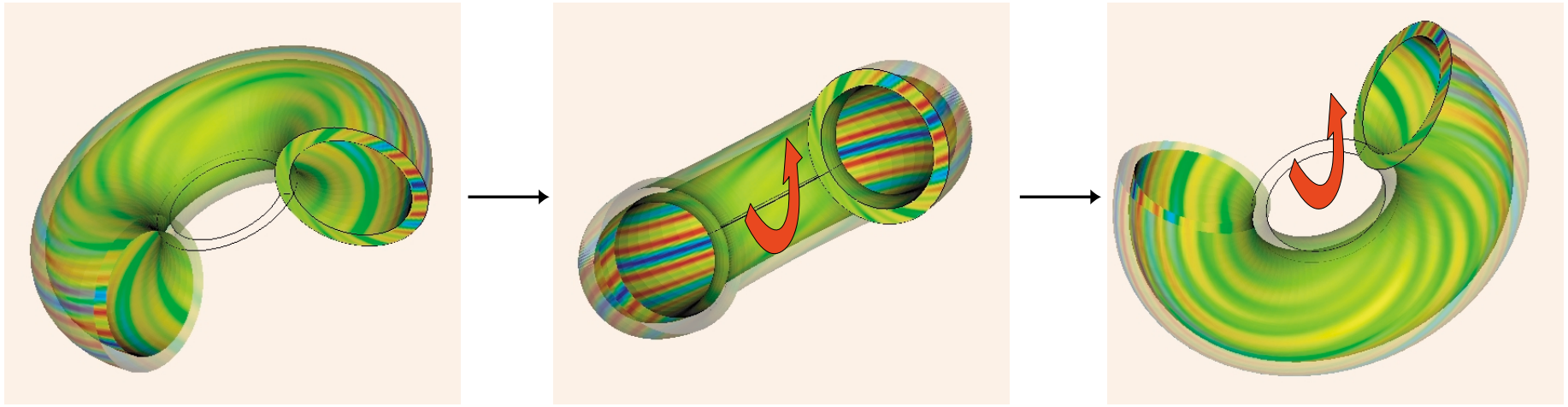


Image Rotation



Microturbulence Simulation via the Gyro code - J. Candy

FUTURE WORK WILL EXAMINE WHETHER OUR SOLUTION CAN SCALE TO TERAFLOP COMPUTING

- It has been estimated that simulations will produce on the order 100 TB/year
  - Experimental community has a total of 10 TB
  - Accumulating at an expanding rate: today it is at 1 TB/year
- MDSplus will be tested on an existing multi–100 TB mass storage system
  - HPSS has been used successfully in the Nuclear Physics community
  - Data retrieval speed both within HPSS and across the ESNET will be evaluated
  - We will develop a plan to support TB datasets in future years
- Advanced scientific visualization to efficiently examine larger datasets will be pursued
  - Develop state of the art graphics packages utilizing technology such as OpenGL
  - Examine the use of display wall technology for collaborative exploration
  - Inexpensive desktop graphics versus centralized systems: What are the efficiency issues of transporting the data versus the visualization?

A SMALL SCALE DEDICATED EFFORT WILL HAVE A LARGE SCALE BENEFIT

- This work will have an immediate benefit to the simulation community
  - Allow scientists to concentrate on their science
  - Facilitate the comparison of experiment and simulations
  - Interactive visualization of simulation data
- It will demonstrate the path to Terascale data management & visualization
  - This will be a long term benefit to the entire science community
- We will continue to reach out to the expertise outside of plasma sciences
  - Kwan–Lu Ma (UC Davis): visualization of large datasets
  - C. Johnson (U. of Utah): interactive large scale computation
  - Kai Li (Princeton U.): display wall
  - R. Mount (SLAC): HPSS
- Full work scope should require on the order 1–2 FTEs & minimal capital expense
  - With leveraging from ongoing activity within the MFE experimental program