

# REPORT FROM THE DATA ANALYSIS PROGRAMMING GROUP

Presented to  
the DIII-D STAFF

*Mission statement is to increase the DIII-D data analysis  
throughput rate and data retrieval rate*

March 18, 1998

GENERAL ATOMICS  
San Diego, California

# ACKNOWLEDGMENT

---

- **Kristi Keith and Bill McHarg**
- **Joe Freeman, Kathy Greene, Carl Parker, and Sheila Miller**
- **The Data Analysis Group**
  - Qian Peng, Jeff Schachter, and 30% of Cathy Campo
  - Bill Davis and Ted Terpstra from PPPL
  - Tom Casper, Bill Meyer, and Jeff Moeller from LLNL
- **The DIII-D User Community**
- **From MIT/C-Mod – Martin Greenwald, Tom Fredian, and Josh Stillerman**

# MEETING AGENDA

---

- **Future computational resources – Schissel (20 minutes)**
  - CPU, storage, and network
- **Automatic in between shot analysis (20 minutes)**
  - Computation and storage in MDSplus – Schachter
  - Serving data – Schissel
- **Tool summary – Schissel (5 minutes)**
  - Kinetic EFIT, Interactive EFIT, PSLICE, TSA
- **EFIT Viewer demo – Peng (10 minutes)**
- **DISSPLA – Schissel (5 minutes)**
- **MDSplus Demo – Schachter (10 minutes)**
  - From IDL, REVIEW, and 4D
- **Web help Demo – Peng (5 minutes)**

# SCOPE OF WORK FOR FY98

---

- Purchase and install a mass storage system for DIII–D raw and analyzed data
- Increase available CPU power through computer upgrades and new systems
  - Factor of 6 for Unix systems (15.4 to 88.8 SPECfp95)
- Enhance existing tools and provide web based documentation
- Common storage of analyzed data via MDSplus
- Long term computational vision
  - Distributed computing with greatly enhanced CPU power
  - Network design and data serving speed enhancement
  - Relational database capabilities
- Relational database implementation delayed due to budget constraints

# A MULTIWORKSTATION ENVIRONMENT IS THE OPTIMUM CONFIGURATION

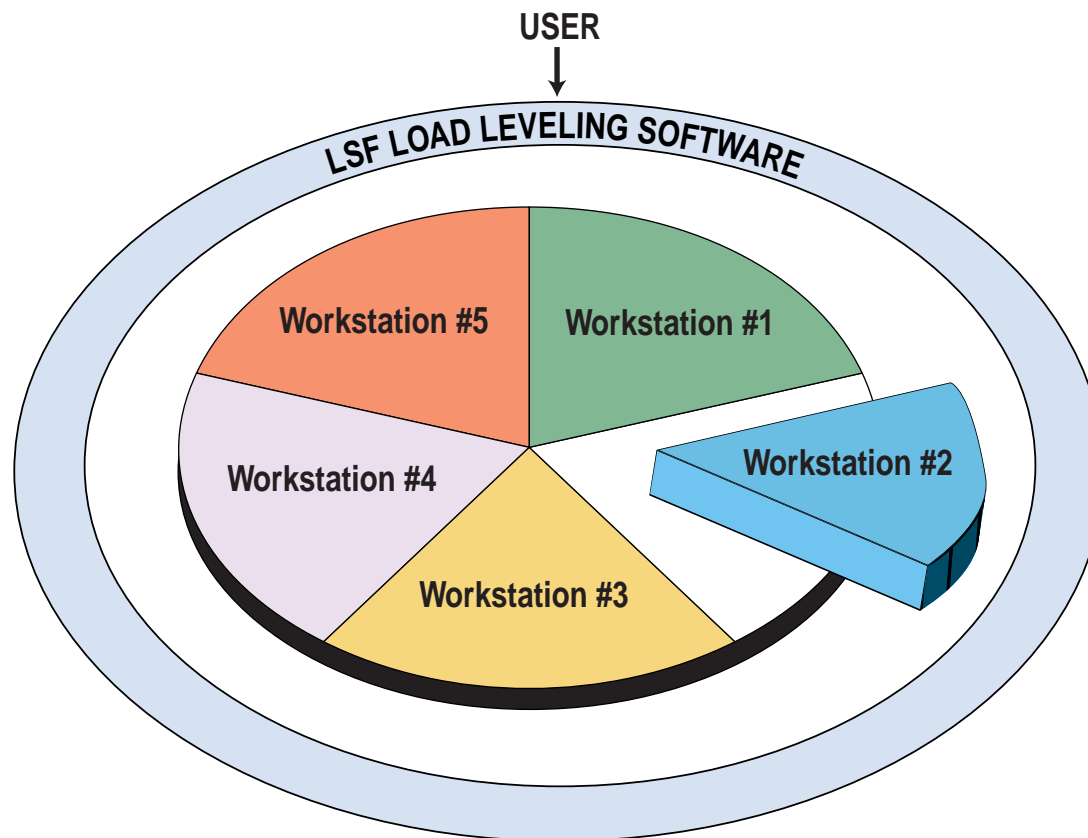
---

- Choice is adding processors to Hydra or adding more workstations
- Hydra has 3 processors and room for 9 more
  - all processors must be the same
  - buying old technology 3 years down the road
  - very expensive to replace Hydra in 5 years
- Multi-workstation environment allows greater flexibility

# BETTER UTILIZATION OF COMPUTING POWER WITH LOAD LEVELING SOFTWARE

---

- LSF handles both interactive and batch processing in heterogeneous environment
  - planning 4 new workstations per year
  - costs \$1k per CPU



# PRESENT AND FUTURE USER FILE STORAGE REQUIREMENTS WILL REQUIRE OUR OWN FILE SERVER

---

- **Present system is by AUSPEX and serves the entire GA site**
  - handles 1000 simultaneous NFS mounts
  - total capacity of 90 GB with 30 GB reserved for fusion
- **GA will be replacing the AUSPEX this summer**
  - \$400k for 200 GB
  - additional 100 GB for fusion will cost \$100k
- **We need on the order of only 20 simultaneous mounts**
  - \$45k for 60 GB and a CPU for NFS
  - \$25k for the next 100 GB
- **When we upgrade we will institute a user file archival capability**

# WE ARE WORKING WITH IRD TO UPGRADE BUILDING 13/15 NETWORK TO FAST ETHERNET

---

- **Building 13/15 network committee**
  - IRD and Fusion personnel
  - will make the technical decision
- **Bill Meyer has been doing network timing tests**
  - During non-work hours our ethernet is OK
  - Hydra to Athena FDDI connection about a factor of 4 slow
  - Busy segments observed for group's involved in operations
  - Some ethernet segments show greater than 70% peak utilization during operations
- **Upgrading to Fast Ethernet requires rewiring Building 13 and 15**
- **Hope to have an answer by the summer**
  - Schedule the work after APS?

# **AUTOMATIC BETWEEN-SHOT ANALYSIS COMPUTATION & STORAGE IN MDSplus**

by  
**Jeff Schachter**

Presented to  
**The DIII-D Staff**

**March 18, 1998**



# OUTLINE

---

- Describe system
  - MDSplus courtesy MIT (Alcator C-Mod)
  - DIII-D: stores analyzed data
- Performance from last run period
- Future issues



# BETWEEN-SHOT ANALYSIS STORED IN MDSplus

---

## CURRENT ANALYSIS

- **EFIT (“automatic control room”)**
  - Non-MSE (usually JT)
  - MSE
- **Calibrations**
  - Neutral beam powers
  - ASDEX gauges
  - Gas flows
  - Filterscopes
- **In-depth calculations**
  - Impurity density
  - Global confinement ( $dW/dt$ ,  $W_{fast}$ ,  $P_{OH}$ ,  $\tau$ 's)
    - ★ Coming soon!



# ACCESS TO ANALYZED DATA

---

- Data can be retrieved from
  - REVIEW
  - 4D
  - Direct from IDL
- Context stored with data
  - Snapfiles
  - Comments
  - Units
- Future datasets automatically available



# COMPUTING

---

- Analysis of each dataset triggered by IPCS events
  - Events declared when raw data available
- Analysis and storage for each dataset happen together
- Multiple datasets loaded in parallel based on events
- Analysis ran on helios
  - Exception: EFIT (DCE system)
- Can run anywhere
  - Unix writes via client/server mode



# PERFORMANCE

---

- **Time to data availability**
  - From when raw shotfile written
  - EFIT: 4 min computation, 2–3 min loading
  - Other data: < 6 min after shot
  - Most are faster: (gasflow ~20 s)
- **Load on HELIOS**
  - Out of 21 run days, problem on only one
  - Load increased over time
- **Disk usage**
  - Total MDSplus usage: 4.2 GB
  - For nine weeks of operations: 8–9 GB
  - 6–8× savings for EFIT's
- **Reliability**
  - Load errors on ~30/465 shots = < 6%
  - Includes shots with PTDATA errors, testing
  - Good feedback



# FUTURE

---

- **Add datasets**
  - Thomson scattering
  - CERquick
  - SPRED
  - Between-shot 4D
- **Unix write access**
- **Reduce availability time**
  - Automatic analysis on multiple, faster CPU's
  - Faster network
  - Faster PTDATA retrieval



# HEAVY DATA DEMAND DURING TOKAMAK OPERATIONS SLOWS DOWN PTDATA SERVING BY UP TO A FACTOR OF 10

---

- **Hydra–Athena on the weekend using PTDATA**
  - 9 Mbits/s for 40 PCS points (0.4 seconds total)
- **Hydra–VAXS during operations using PTDATA**
  - 0.8 Mbits/s for 40 PCS points (4.6 seconds total)
- **PTDATA search path**
  - Local system, VAXN, VAXS, Athena, AEG1

# WITH MODERATE SOFTWARE AND HARDWARE ENHANCEMENTS WE CAN SERVE PTDATA UP TO 80 TIMES FASTER

---

- **Hydra–Athena on the weekend using PTDATA**
  - 9 Mbits/s for 40 PCS points (0.4 seconds total)
- **Hydra–Athena on the weekend comparing PTDATA to GADAT**
  - 0.3 Mbits/s for 40 PCS points (12.0 seconds total) using GADAT
- **Hydra–VAXS during operations using GADAT**
  - 0.1 Mbits/s for 40 PCS points (32.6 seconds total)
- **Solution should be implemented for FY99 operations**
  - Serve data from a fast CPU (VAXS is 100 times slower than Nitron)
  - Connect this new CPU on Fast Ethernet
  - Utilize the new mass storage system for shot archival
  - Repair GADAT and upgrade some tools to call PTDATA directly

# ANALYSIS TOOLS ARE BEING DEVELOPED IN SEVERAL AREAS

---

- **Kinetic EFIT**
  - Completed and working in 4D but can be operated independently
  - Kinetic EFIT in <1 hour
- **Interactive EFIT**
  - In development phase
  - Mirror the interactive feel of 4D's Profile Fitter
- **PSLICE (general profile plotting)**
  - Being adapted from Schachter's work at C-Mod
- **TSA**
  - General Fourier analysis code
- **4D re-write**
  - Start with EFIT viewer and PTDATA Viewer (1 month each)
  - Create more features and taking less than 5% of our time to maintain

# DEMO: NEW EFIT VIEWER

---

- **New IDL compound widget to choose an EFIT**
  - from files or from MDSplus
  - MDSplus has description of EFIT run
- **Contour plot of equilibrium**
  - improved overlay features
  - rapid retrieval of shape control points (80 times faster than during ops)
- **New profile graphs**
  - includes most information from file PLOTOUT.OUT
  - new profile information utilizing MSE diagnostic (new measurements file)
  - rapid switching between  $\rho$ , R, and  $\Psi$  representation
- **Plot-Objects graphical routine**
  - developed in conjunction with consultant
  - improved cursor functionality with (x,y) readout for each graph
  - rapid transition from many to one graph
  - zooming, data selection within a graph

# DISSPLA HAS BECOME TOO COSTLY TO USE

---

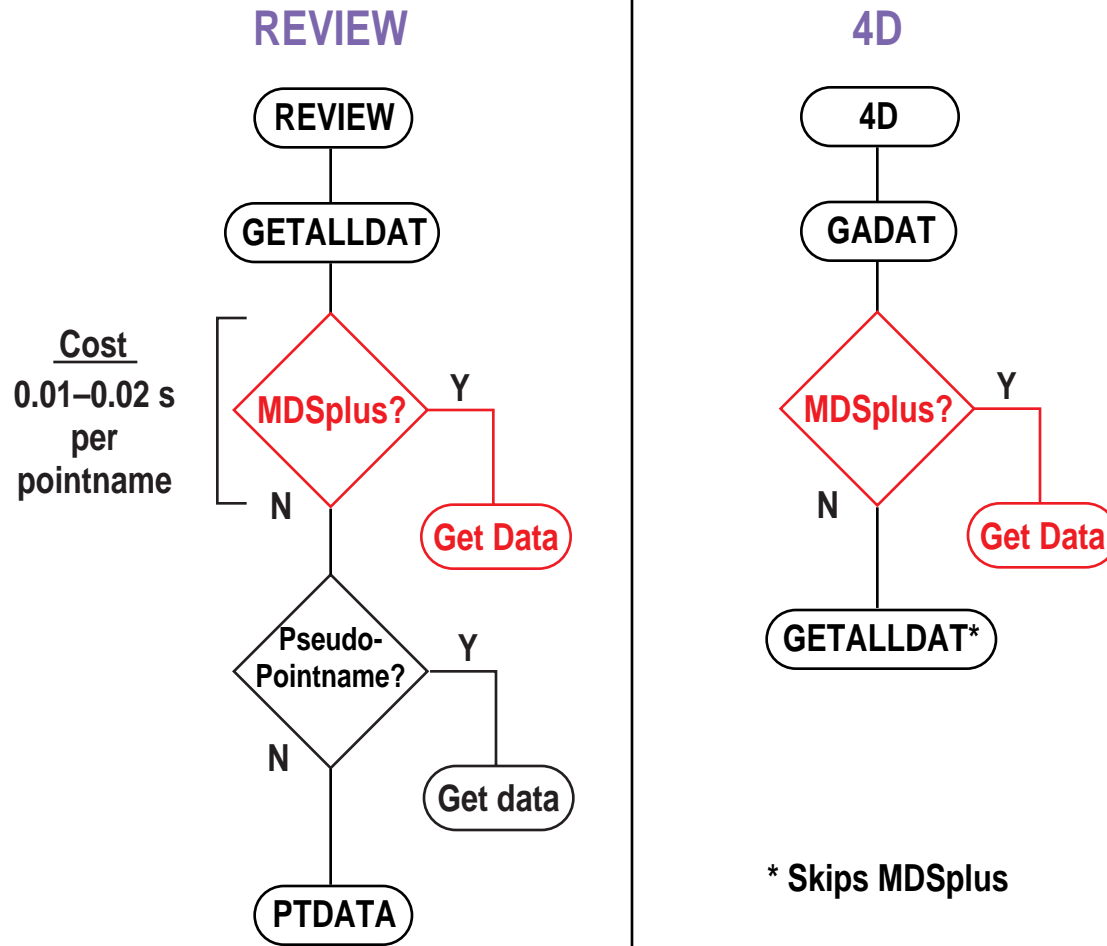
- **Quoted at \$29.6K per HP C-180 class workstation**
  - Nitron in the control room is an HP C-180
  - Disspla will cost more than the workstation
- **We anticipate purchasing 4 workstations per year for 5 years**
  - Total Disspla purchase outlay will be \$592.7K
  - At the 5 year mark, \$77.1K/year in maintenance
- **How large is the people cost of converting legacy code?**
  - The FTE cost of converting in the past has been perceived to be much larger than the cost of a new graphic software license

# THE NASADIG SOFTWARE PACKAGE IS AN ECONOMICAL WAY TO REPLACE DISSPLA

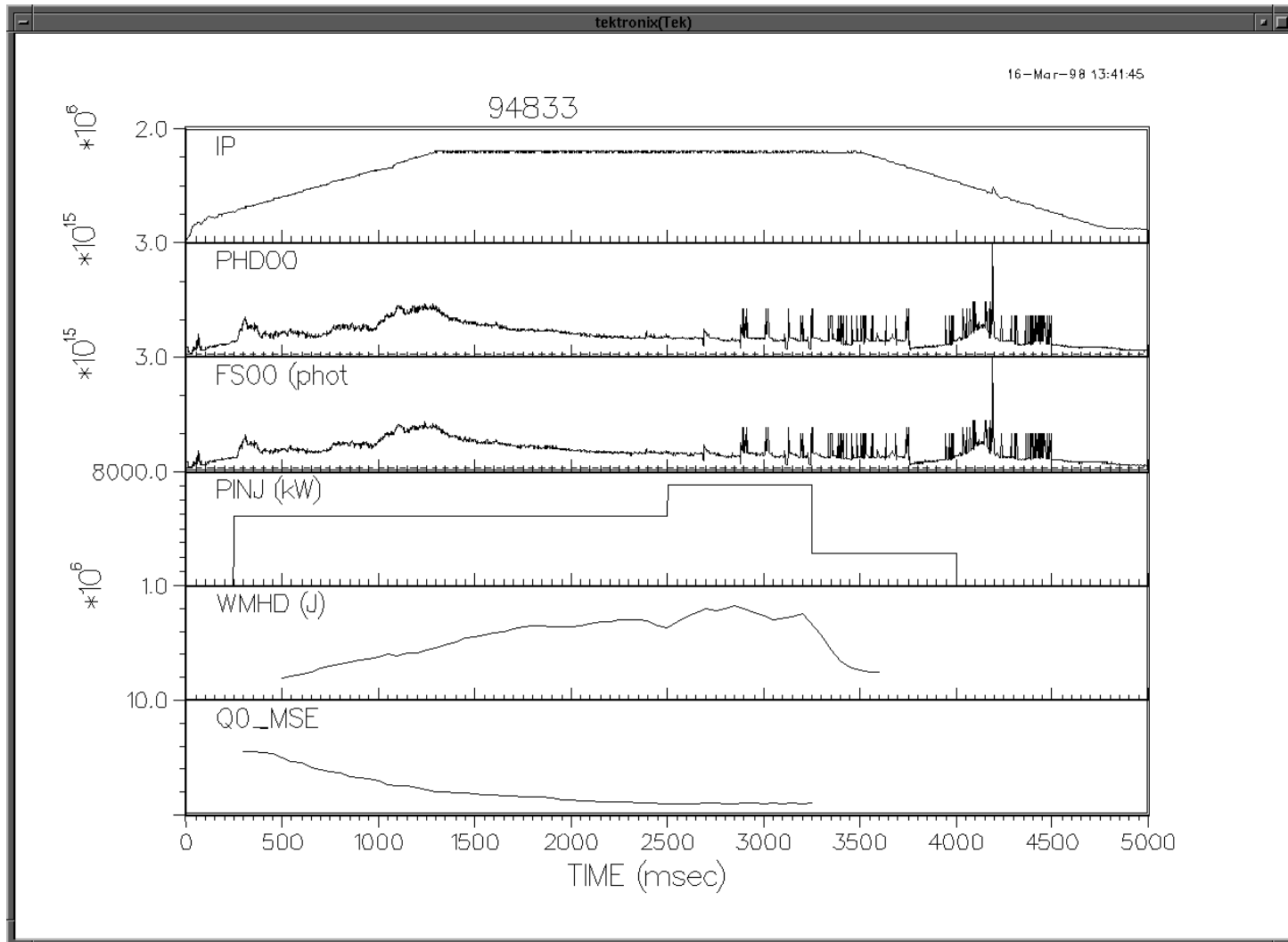
---

- A UNIX site license costs \$3k with no annual maintenance cost
- For our usage NASADIG has the same functionality as Disspla
- Converting legacy code will take no more than 1 person-year
  - Represents a \$200K converting cost
  - REVIEW took only 2 weeks
  - NASADIG is a Disspla clone
- Switching to NASADIG will save close to \$500K

# LOGIC FOR DATA RETRIEVAL



# MDSplus DATA ACCESSED IN REVIEW



PTDATA

Raw PTDATA

Calibrated MDSplus

MDSplus & PTDATA

EFIT MDSplus

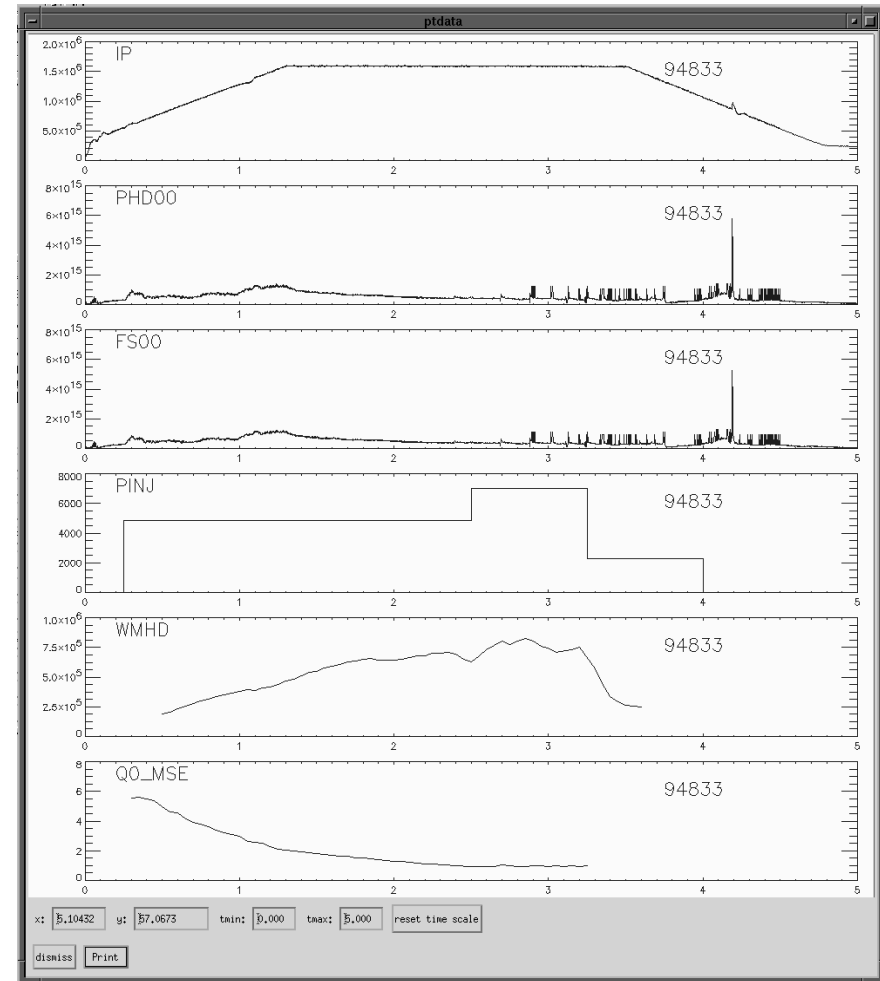
MSE EFIT MDSplus

# MDSplus DATA ACCESSED IN 4D

## Point Data Selection Widget



## Point Data Output Widget



# DEMO: MDSplus DATA RETRIEVAL FROM IDL

---

- **Time history of EFIT signal WMHD**

- retrieve data and time base, units, and snapfile

- Commands:

```
mds$open, 'EFIT01', 94833
y = mds$value ('_S = \WMHD')
x = mds$value ('DIM_OF (_S)')
plot, x, y
units = mds$value ('UNITS_OF (_S)')
snapfile = mds$value (' \TOP:NAMELIST ')
```

- **Safety factor profile at one time**

- Commands:

```
y = mds$value ('_S = \QPSI')
x = mds$value ('DIM_OF (_S,0)')
plot, x, y (*,index)
```

- **Retrieve PTDATA from MDSplus**

- Commands:

```
y = mds$value ('_S = PTDATA("IP")')
x = mds$value ('DIM_OF (_S)')
plot, x, y
```

# DEMO: WEB BASED HELP

---

- **Proof of concept using EQVIEW application**
- **Rapid help**
  - displayed within the tool
  - one line short description
  - source of one line from Signal documentation: Brief Description
- **Detailed information from Web**
  - double click on variable to launch or bring up Netscape with appropriate URL
- **One location for help information that is viewable by multiple methods**