Computing at Fermilab

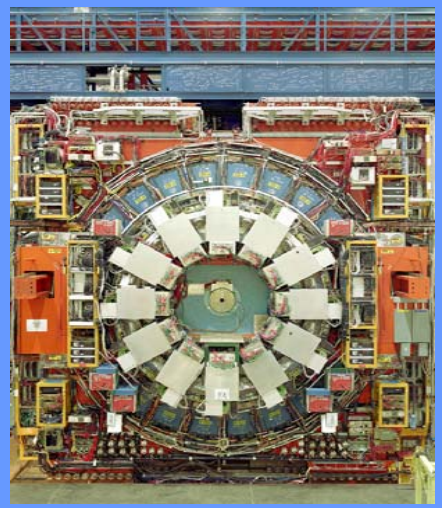
David J. Ritchie
Computing Division
May 27, 2008

Computing At Fermilab is…

- In support of the mission…
  - to advance the understanding of the fundamental nature of
    - matter and
    - energy

The Accelerator

The DZero Experiment

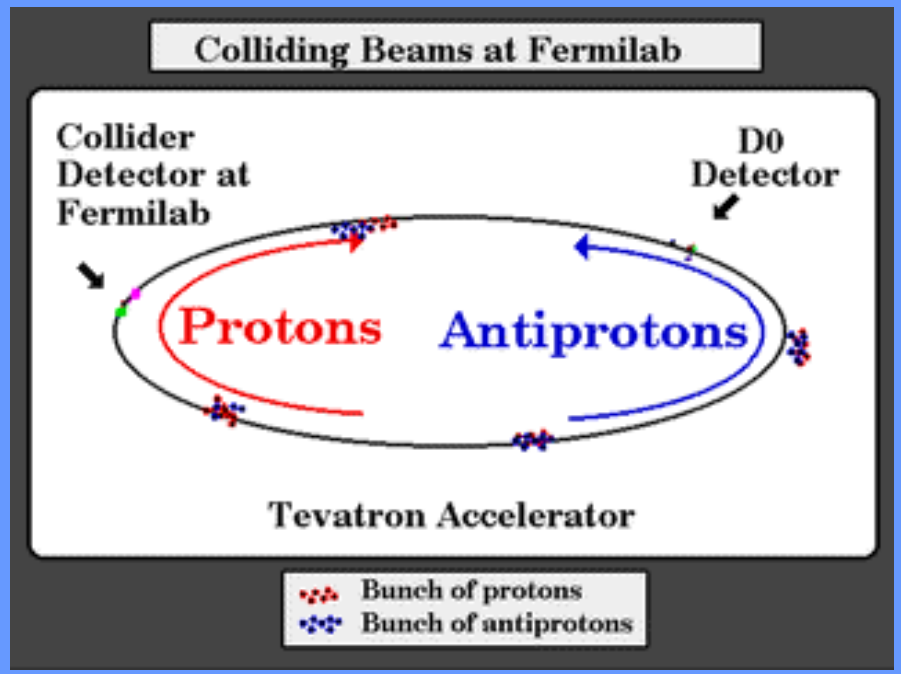


The CDF Experiment

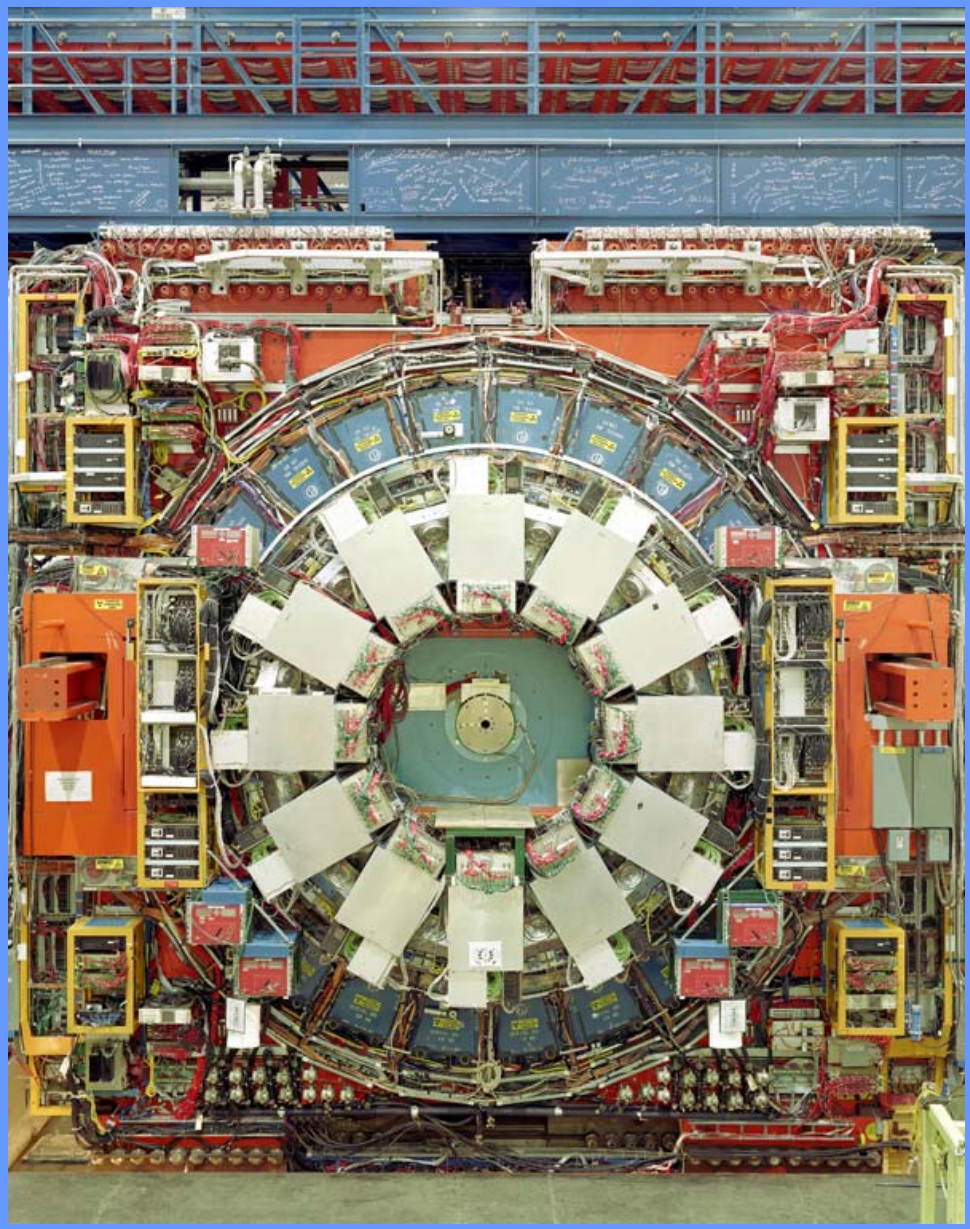
And Many Others
KTev, Minos, SDSS, …, P929.
I.e. more than 929 proposals to date.

The Accelerator

The Accelerator

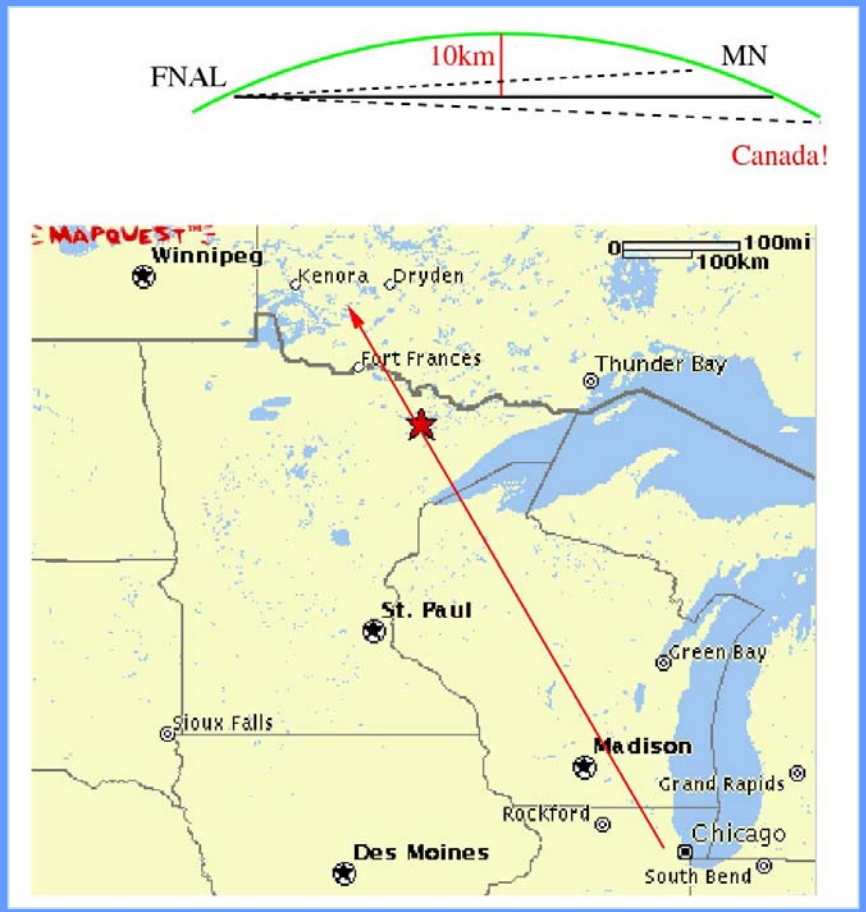


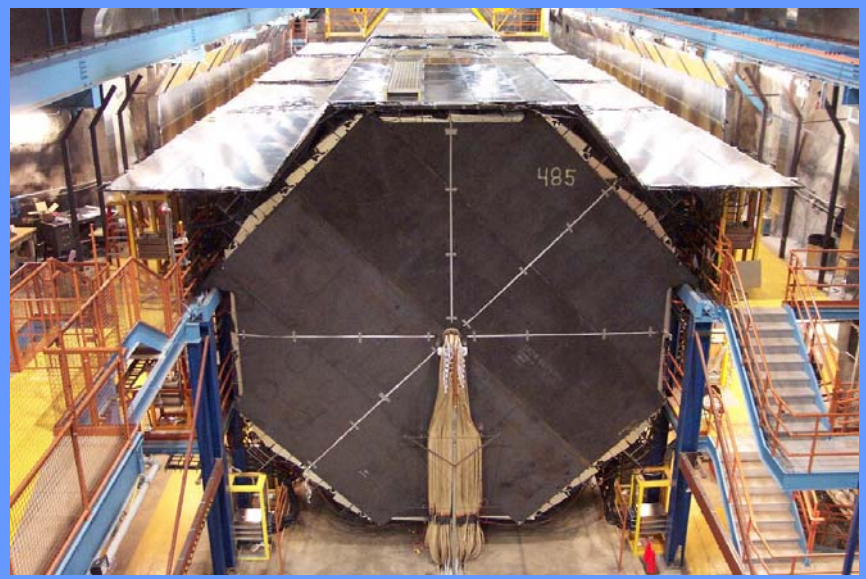
The DZero Experiment



The CDF Experiment

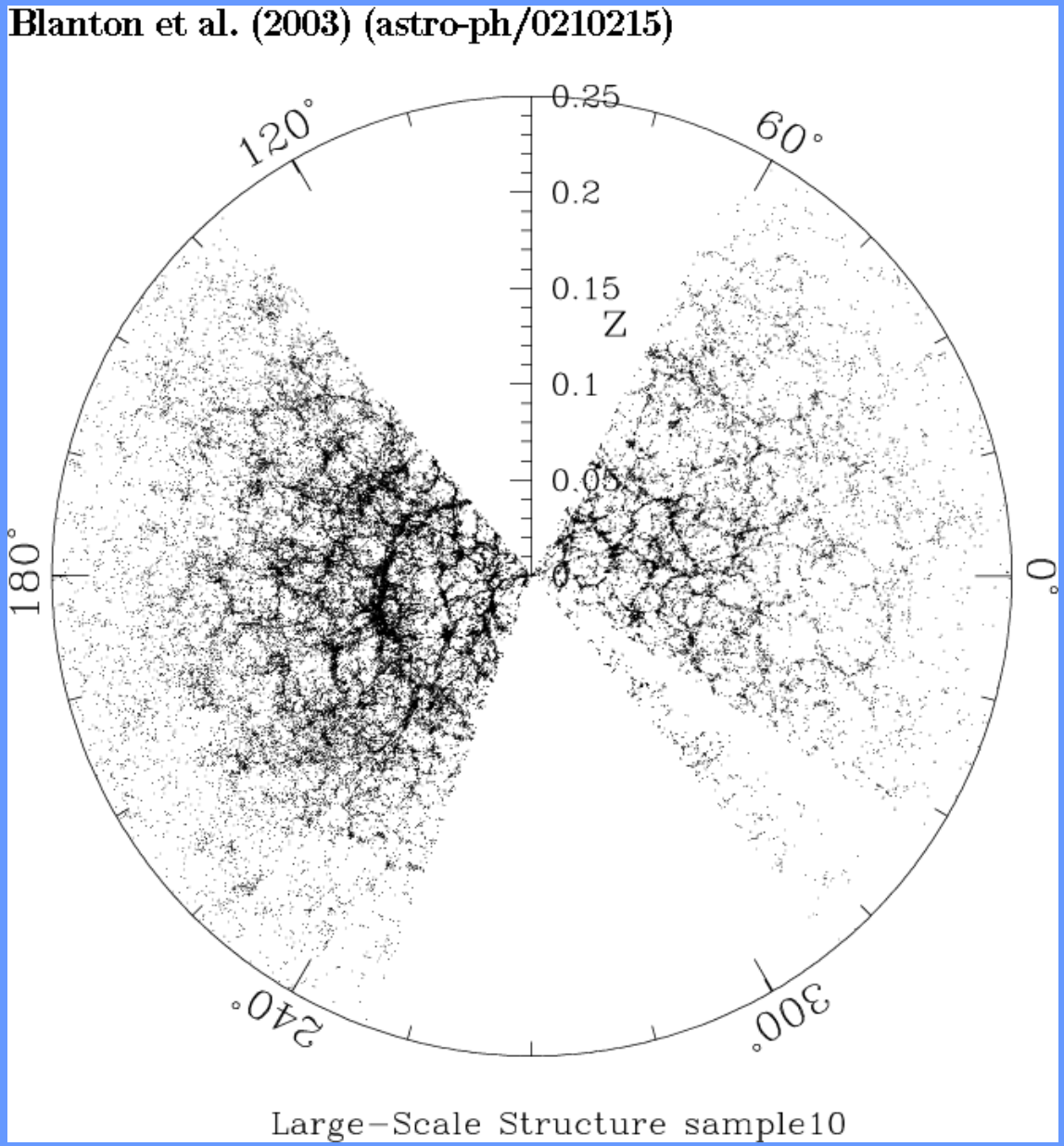
MINOS Near and Far





The Far Detector

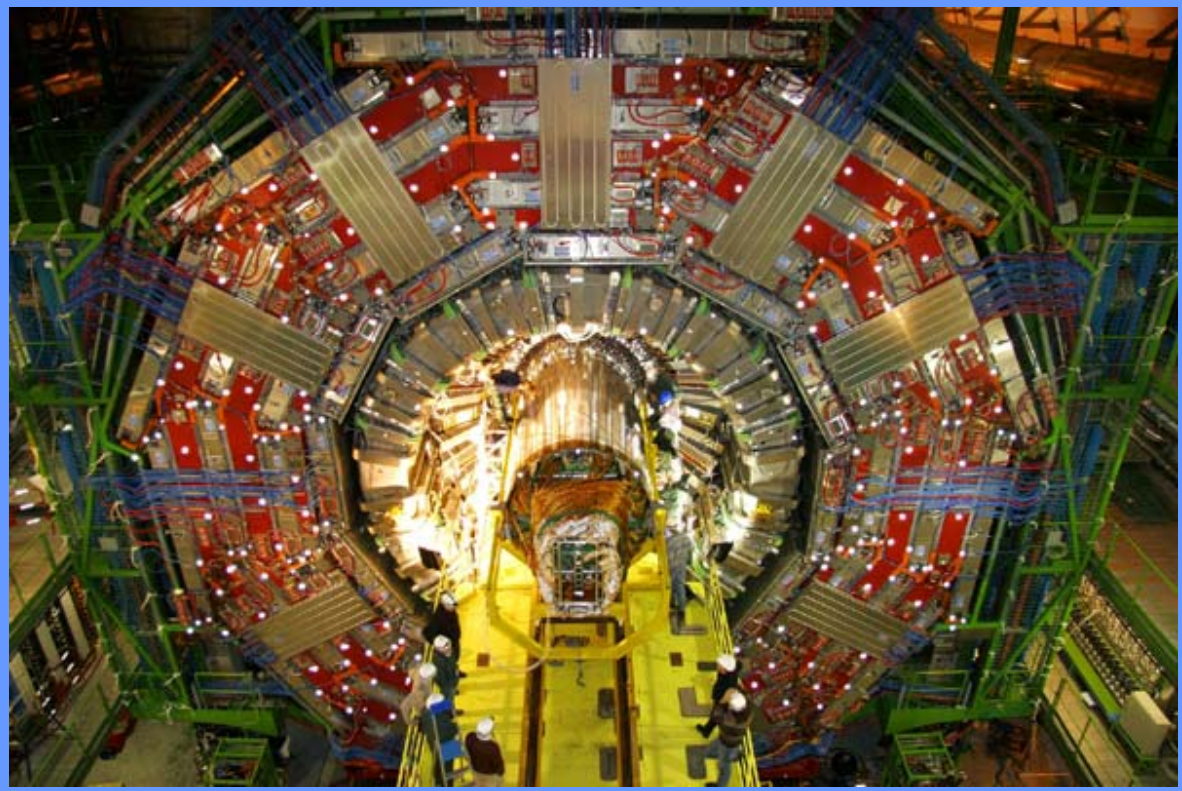
Sloan Digital Sky Survey



An SDSS Result

CMS Experiment

The CMS Detector…



How Do You Do Science?

(elementary school version)

1. Observe **phenomena**.
2. Develop a ***hypothesis***.
3. Use hypothesis to make ***predictions***.
4. Devise ***experiment*** to look for predictions.
5. Obtain ***results*** demonstrating (or not) predictions.
6. Draw ***conclusions*** about correctness of hypothesis.
7. If correct, add to accumulated hypotheses which are the **theory**. If not, revise hypothesis and …

Repeat

How Do You Do Science?
(Large Scale Science Version)

Each stage of scientific process has different computing needs.

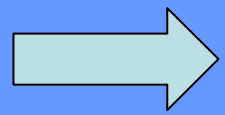
- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan and Design **Experiment**
- Build / Install Equipment
- Acquire / Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**

Repeat

Computing Needs
Phenomena and Hypothesis

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**



Need: access previous **results**

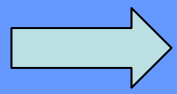
- PC's (Windows, Linux), UNIX, Macs
- Printers – color and B/W
- World-wide networking
- MSWord, TeX, Acrobat
- Web preprint repositories
  - Arxiv.org at Cornell U.
  - Spires at Stanford U.

Repeat

Computing Needs
Organize Collaboration

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**



Need:

- PC's (Windows, Linux), UNIX, Macs
- World-wide networking
- E-mail
- Web Pages
  - DZero "Top Group"
  - MINOS

Repeat

Organizing a Collaboration means interacting with ~400 colleagues

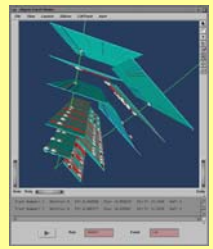
Computing Needs
Propose Experiment

Each stage of scientific process has different computing needs.

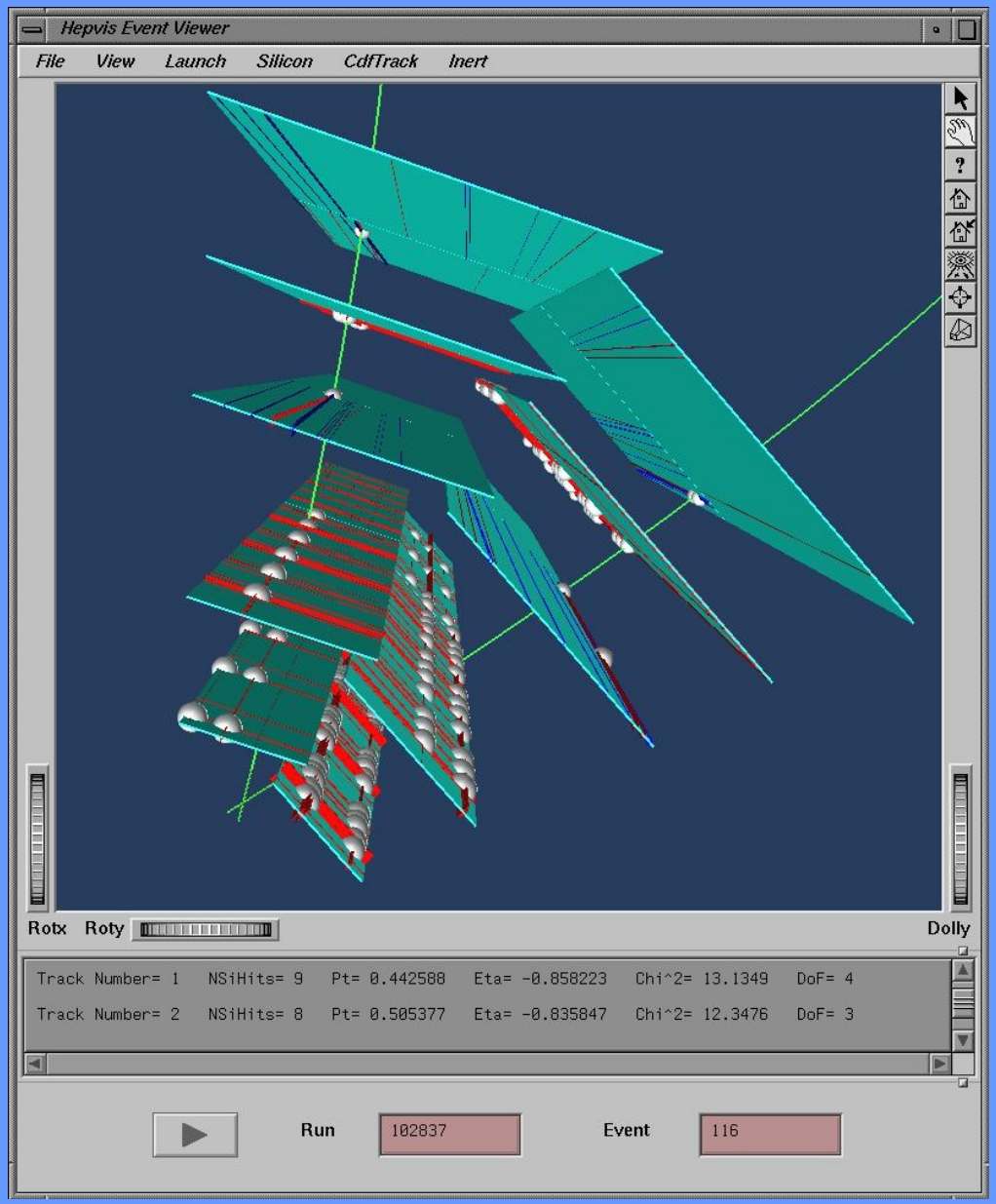
- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**

Need: with colleagues show the feasibility of a proposal.

- PC's (Windows, Linux), UNIX, Macs
- World-wide networking
- Web pages
  - Dark Energy Camera Project
  - Nova Proposal
- Simulations…



Repeat



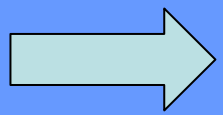
CDF Silicon Vertex Detector

A 3D modeling program simulating particles in the detector.

Computing Needs
Get Approved and Funded

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Acquire / Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**



Need: Project Management

Funds come from DOE, NSF, NASA, Congress, Taxpayer—YOU.

- PC's (Windows, Linux, Macs)
- Printers – color and B/W
- World-wide networking
- MSProject, Power Point…
- Proposal review…
  - NUMI Proposal Review

Repeat

NuMI Project

Project Overview

Greg Bock

DoE Office of Science Review of the NuMI Project

January 30, 2002

Fermilab

- Introduction
- Management Items
- Status and Technical Progress Since September
- Cost and Schedule Status
- Project Goals for May
- Summary

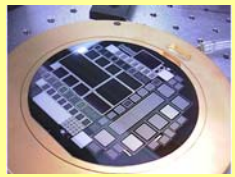
Computing Needs
Plan / Design, Build / Install

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Acquire / Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**

Need: simulation, sensor design, engineering software…,

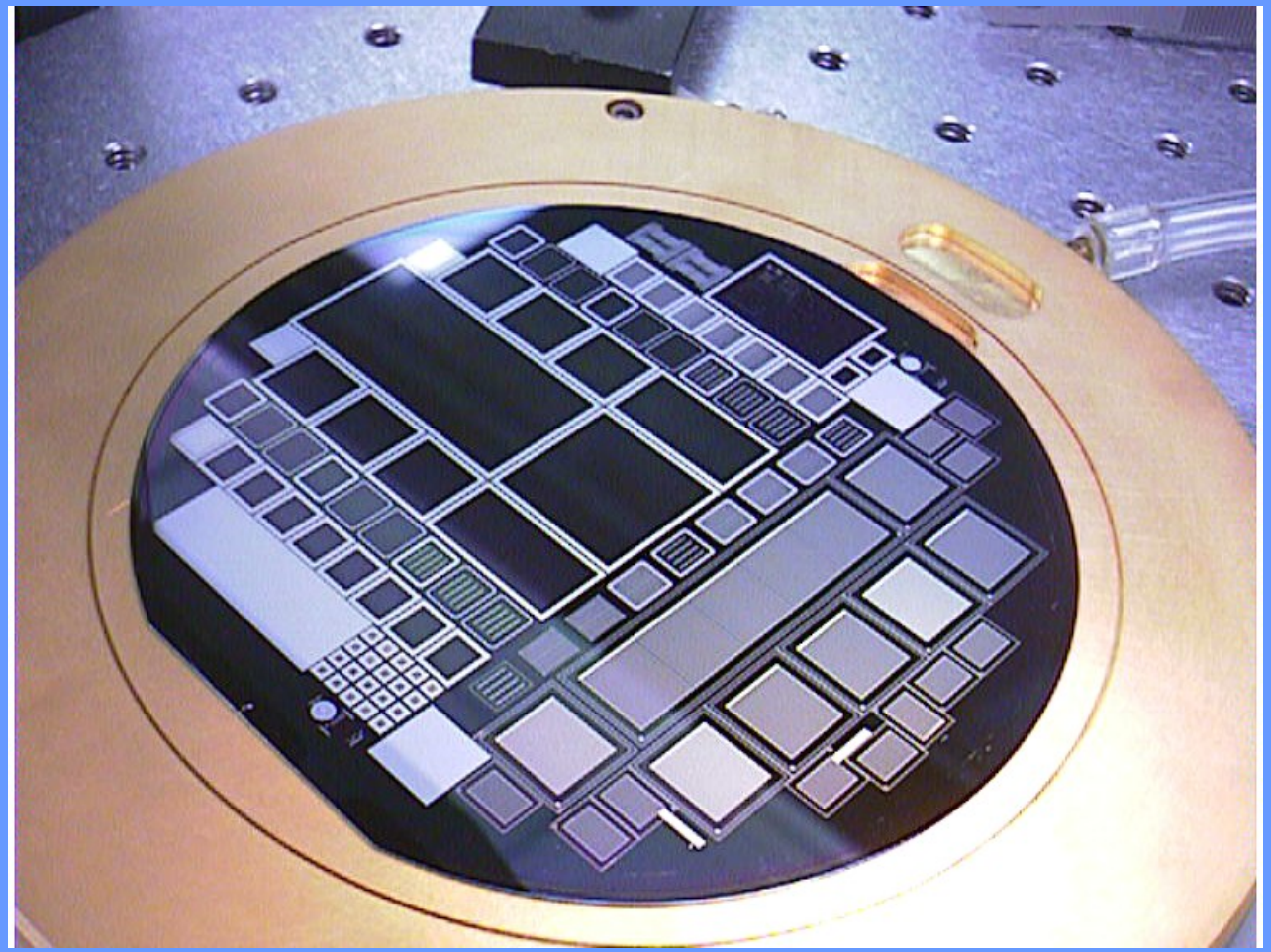
- Unusual sensors

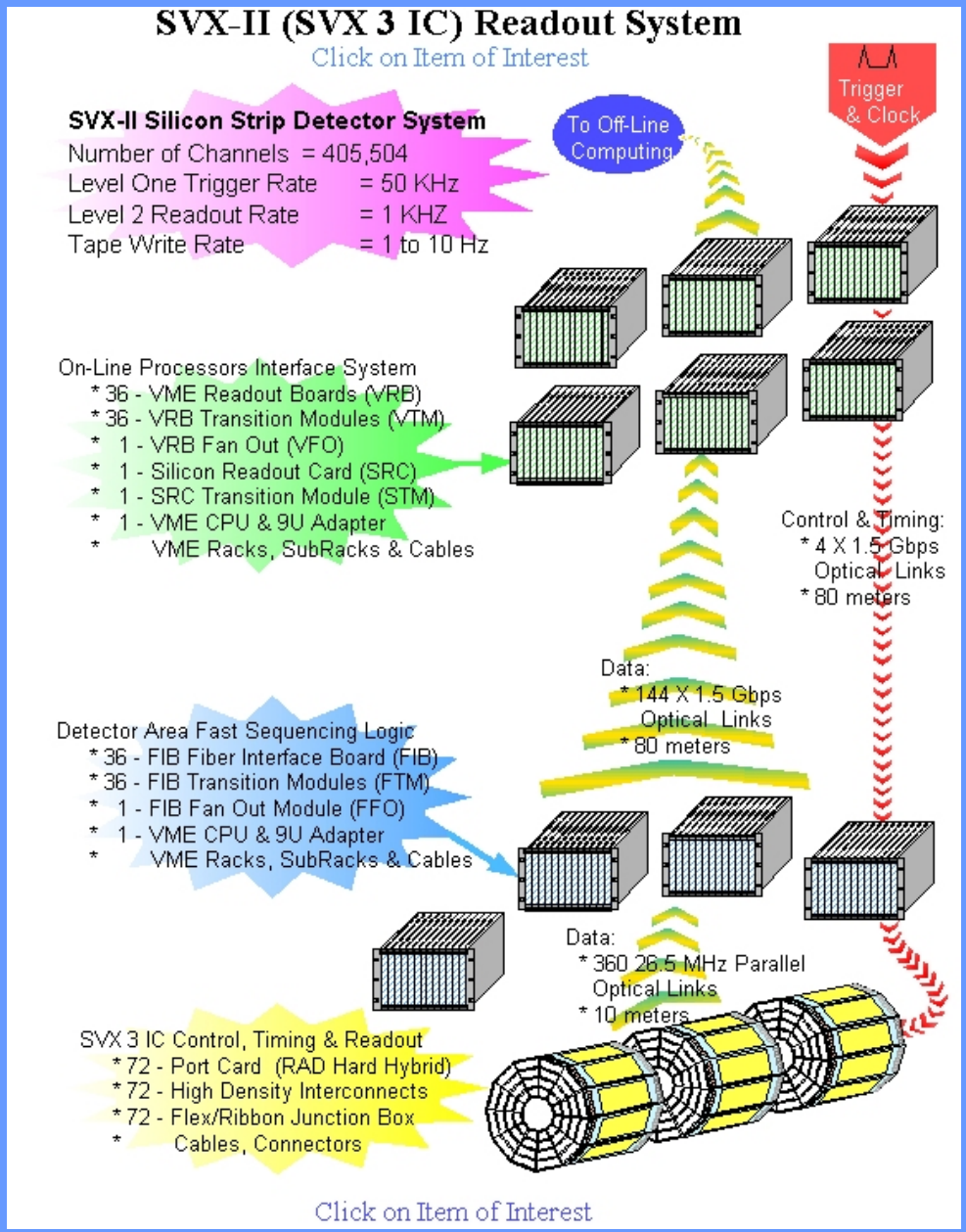


- Custom electronics
- Computing at its core: microprocessors, silicon logic

Repeat

Unusual Sensors





Custom Electronics

- Example: CDF Silicon Vertex Detector Readout
- VRB Board being passed around was a part of this.

Computing Needs
Acquire / Record / Store Data

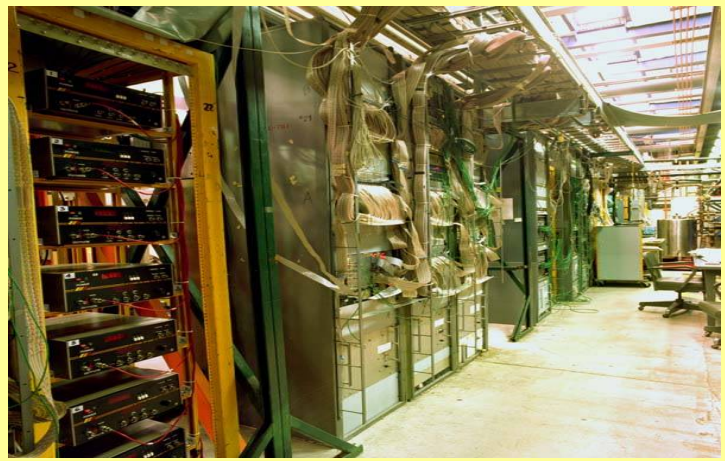
Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Acquire / Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**

Repeat

Need: Acquire, record, store

- Use sensors and custom electronics



Acquire, Record, Store Data

Recording Data

- Record data (“events”)
  - The sensors and custom electronics convert the voltages and currents into 300,000 bytes of 1’s and 0’s: N*ow, we are in the computer realm.*
  - Get an event every few hundred microseconds.
  - Raw data rate — hundred’s of GB/second.
- Either trigger on interesting events or filter out the uninteresting ones – ~10 MB/s.
- What gets through, one records to disk and eventually to tape.

Computing Needs
Analyze Data

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Acquire / Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**

Need:
- Reconstruct 1's and 0's into tracks and particle identification
- Do physics analysis on resulting collections of tracks

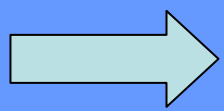
Means:
- PC farms, lots and lots
- Disk backed up by tape with robotic tape mounts Enstore Usage.

Repeat

Computing Needs
Obtain Results

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**



Repeat

Need: Do additional physics analysis.

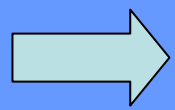
Means: Use analysis farms with their disk caching capabilities to speed data analysis.

CDF Usage

Computing Needs
Publish Conclusions

Each stage of scientific process has different computing needs.

- Identify **Phenomena**
- Develop **Hypothesis**
- Organize Collaboration
- Propose **Experiment**
- Get Approved
- Obtain Funding
- Plan / Design **Experiment**
- Build / Install Equipment
- Record / Store Data
- Analyze Data
- Obtain **Results**
- Publish **Conclusions**



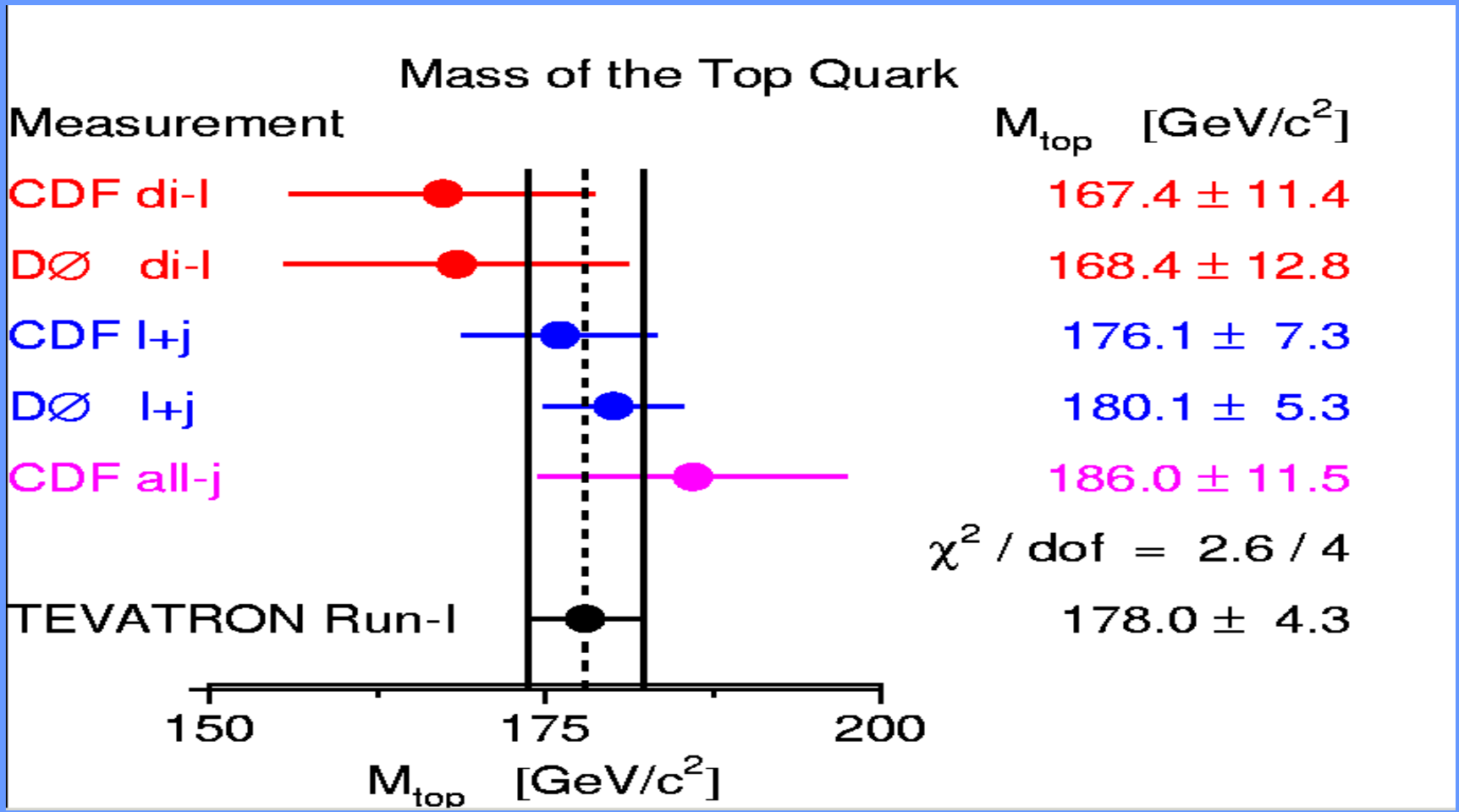
Repeat

Need: Show evidence of progress in scientific work.

Means: Publish papers and conferences.

Top Quark Mass

Mass of the Top Quark



Vital Statistics

- FNAL has >10,000 network connected devices
- They are connected by >1,000 miles of cabling
- We have over 6 PB of data on tape
- Half of it is less than 2 years old

Computing Facilities

- The activities map into facilities at the laboratory:
  - Networks
  - Mass Storage robotics / tape drives
  - Large computing farms
  - Databases
  - Operations
  - Support

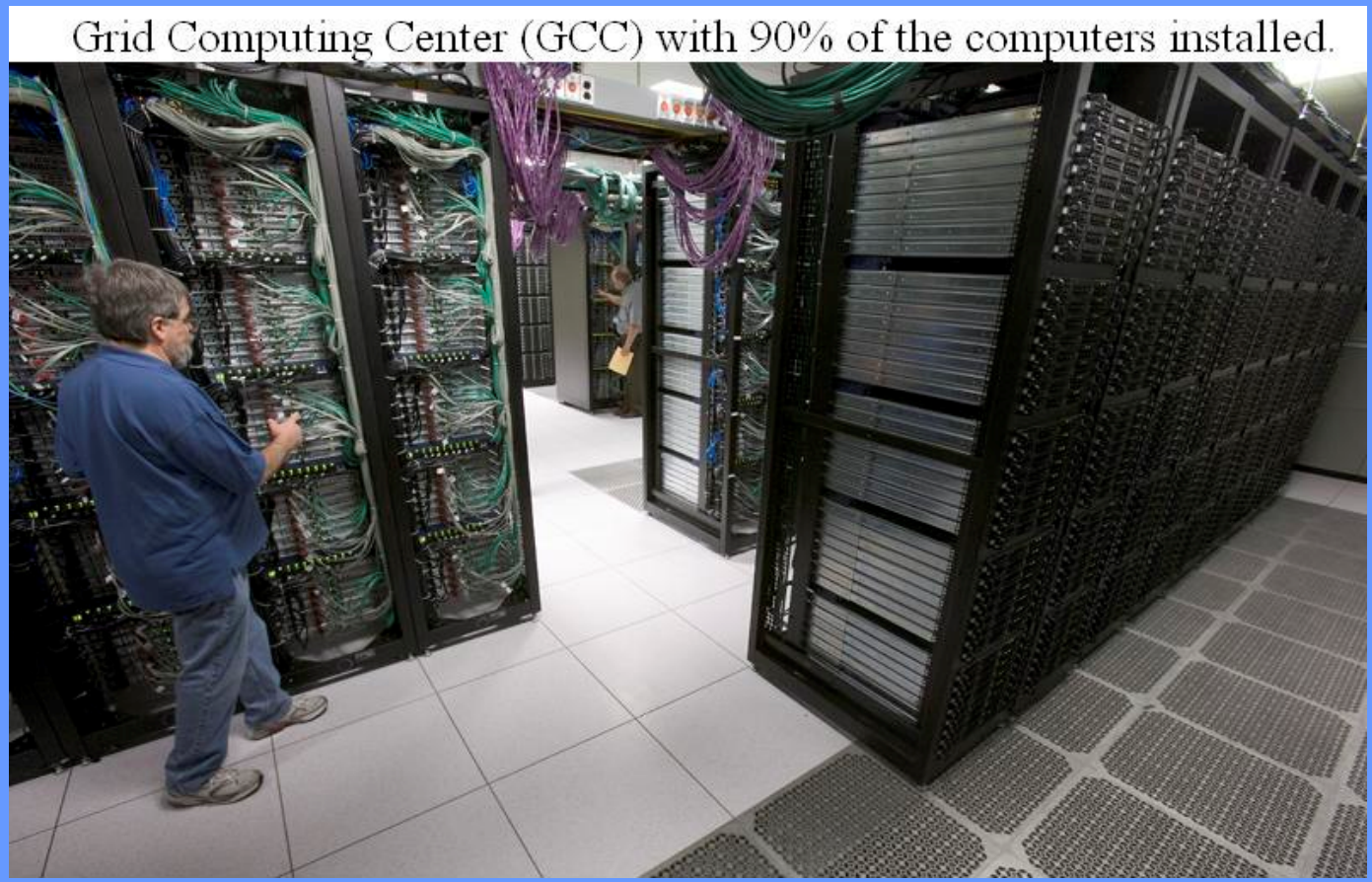
PCs and PetaBytes

FCC

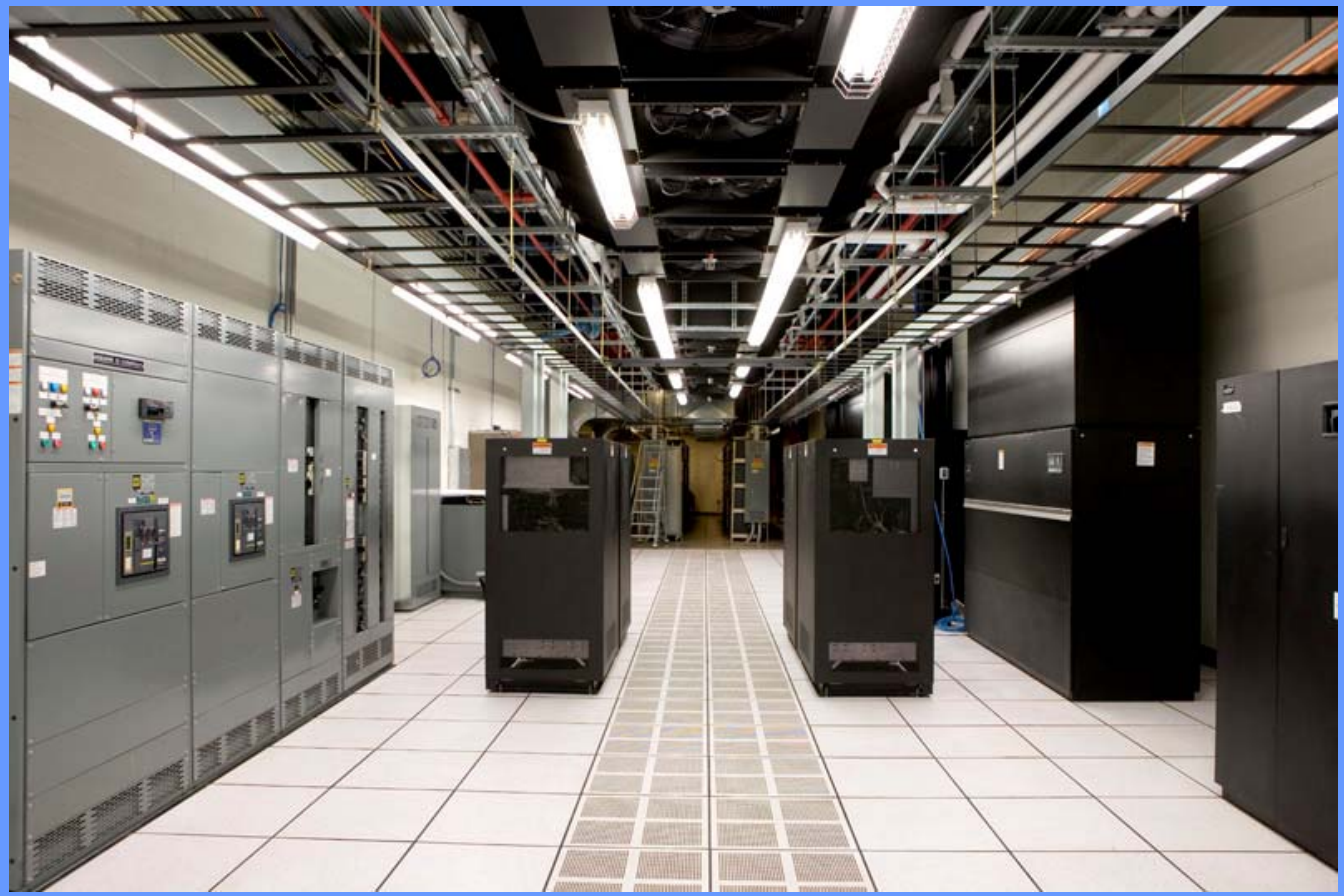
Facilities - GCC

GCC

Facilities - GCC

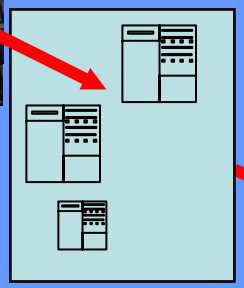


Facilities - LCC



D0 computing systems

"data-logger"



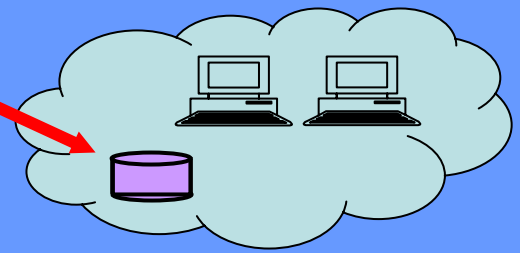
12-20 MBps

"d0-test" and
"sam-cluster"

"central-analysis"

100+ MBps

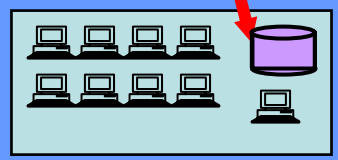
400+ MBps



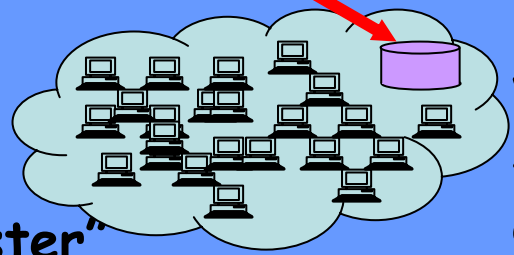
"linux-analysis
-clusters"

12-20 MBps

"farm"

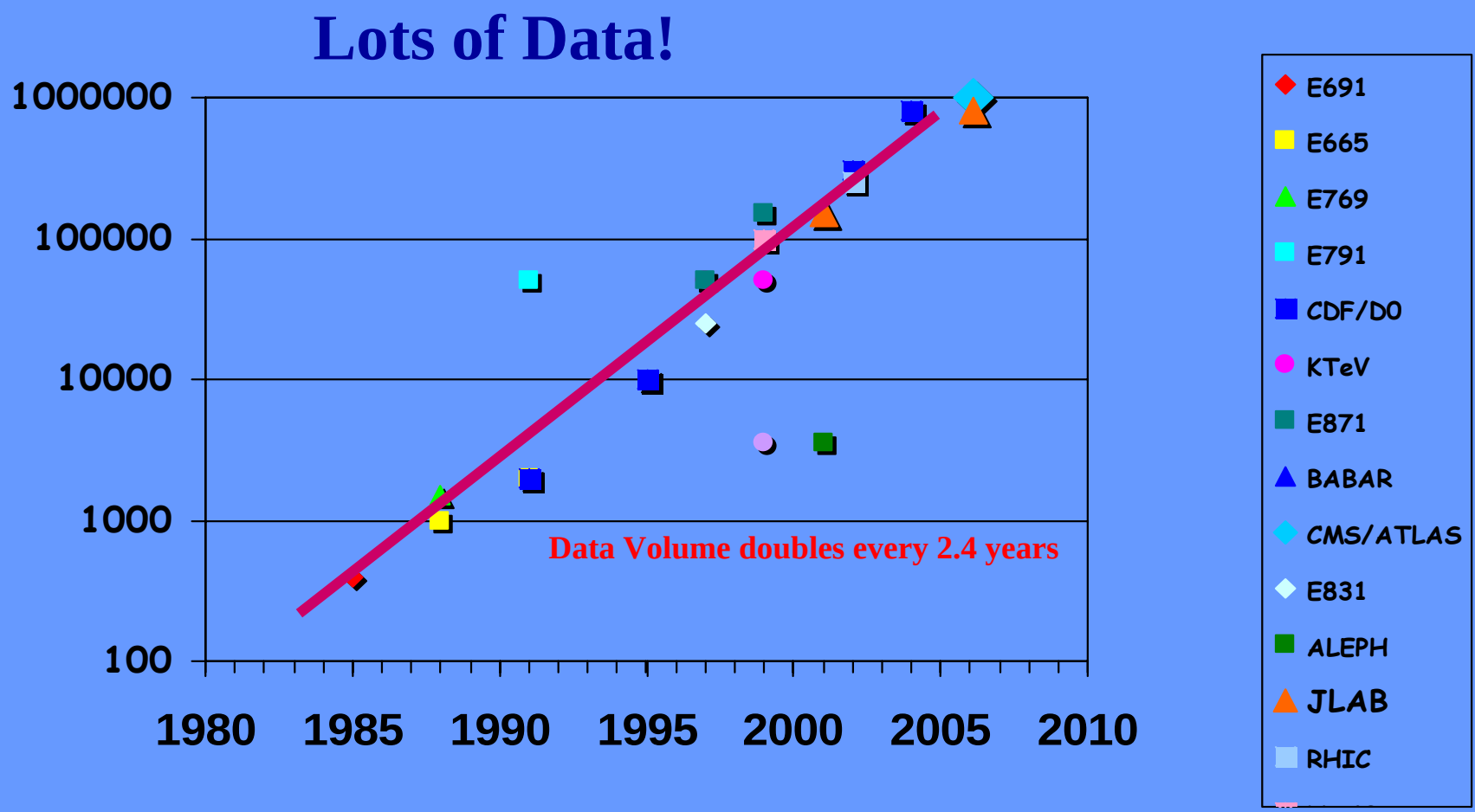


"linux-build-cluster"



"clueD0"
~100
desktops

Moore's Law


Lots of Data!
1000000
100000
10000
1000
100
1980
1985
1990
1995
2000
2005
2010
Data Volume doubles every 2.4 years
E691
E665
E769
E791
CDF/D0
KTeV
E871
BABAR
CMS/ATLAS
E831
ALEPH
JLAB
RHIC

GRID Computing

- Fermilab is participating in many GRID initiatives:
  - ppdg (DOE SciDAC)
    - Particle physics data grid
    - Fermilab, SLAC, ANL, BNL, JLAB, Caltech, UCSD, Wisconsin, SDSC
  - Griphyn
    - Grid physics network
  - iVDGL
    - International virtual grid laboratory

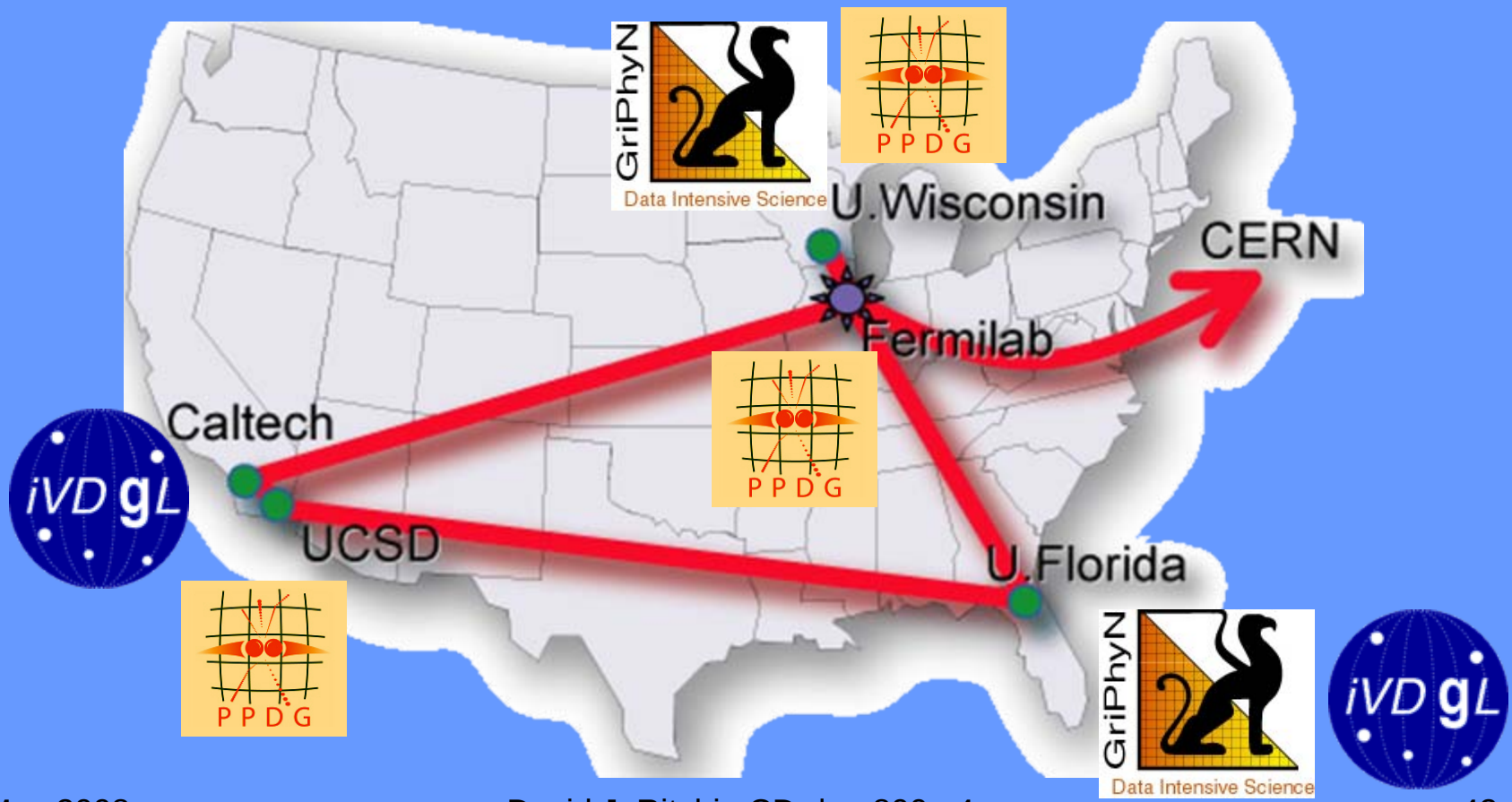
- GRID activities have been a very natural outgrowth of distributed computing of the large collaborations.

US CMS Computing

- Fermilab is the host lab of U.S. CMS experiment which will begin taking data at CERN in Geneva, Switzerland in ~2008
- Fermilab hosts the project management for the U.S. CMS Software and Computing Program in DOE
- U.S. Physicists will participate in this research.

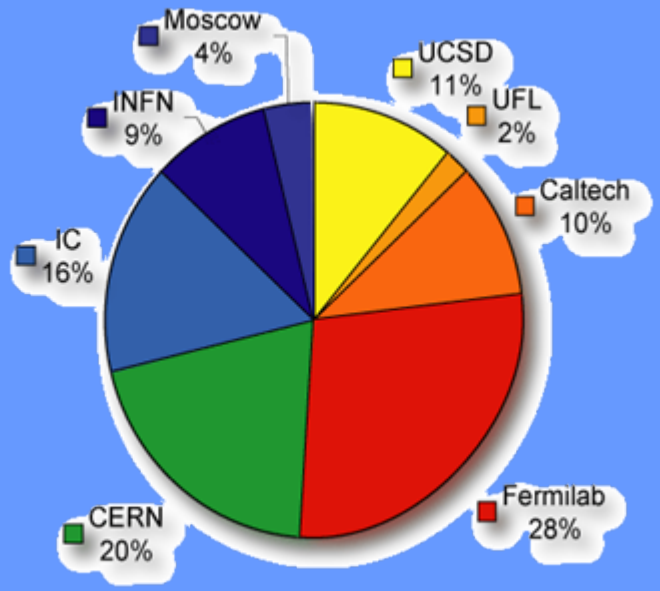
US CMS Computing

Tier 1/Tier 2 Centers and Prototypes



US CMS Computing

- Contribution to CMS is substantial
- Series of Data Challenges
  - Increasing complexity
  - >70 million events fully simulated



Computer Security
The Dark Side of the Force

- The Internet is now ~100 Million computers/users
  - "one in a million" events happen every day 'somewhere on the net'
  - Only with the net 'somewhere on the net' is always just next door.
- We are learning how to live in a world where not everyone is friendly -- but most are.

Computer Security
The Dark Side of the Force

- This means that computer security is an ever present task—particularly, when studies have shown that a hack attempt is made once every 13 minutes.
- We use extensive automated scanning software to detect problems and isolate them typically within a few hours if not a few minutes.

Acknowledgements

As always in a laboratory, one gains from the work of one's colleagues. This talk is no exception. I would particularly like to thank Dane Skow for the use of some of his slides from a previous talk.