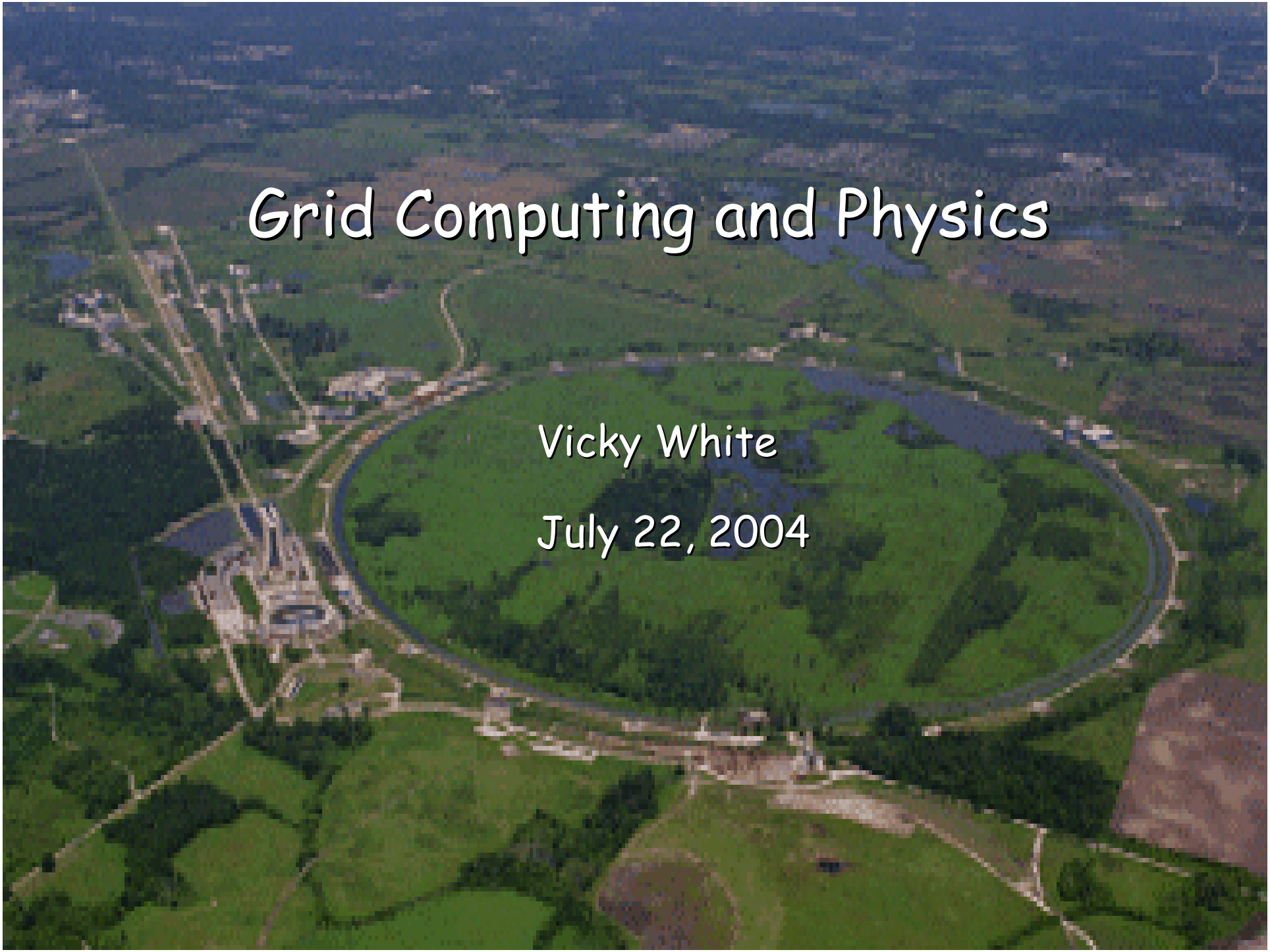


Grid Computing and Physics

Vicky White

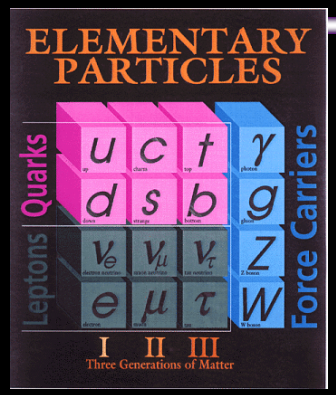
July 22, 2004

Outline of Talk

- Computing and Physics
- Computing at Fermilab
- Grid Computing – what and why
- Grid Computing Tour

Computing and Physics

Computing is Essential to Scientific Discovery

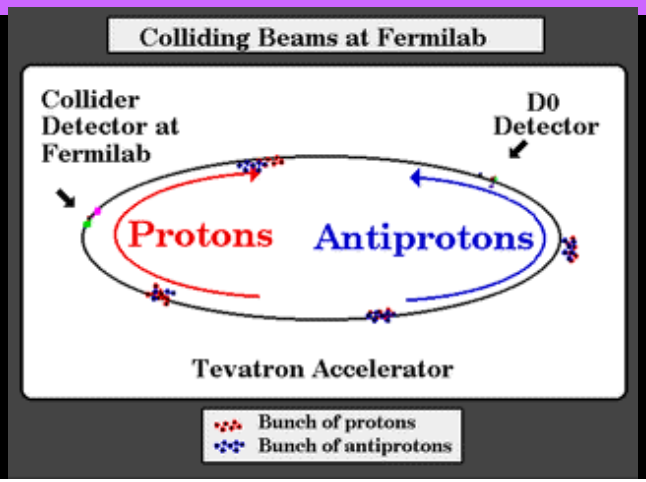


Theory

discovery by computation e.g. Lattice QCD, Nuclear Astrophysics, Cosmology

Accelerator Science

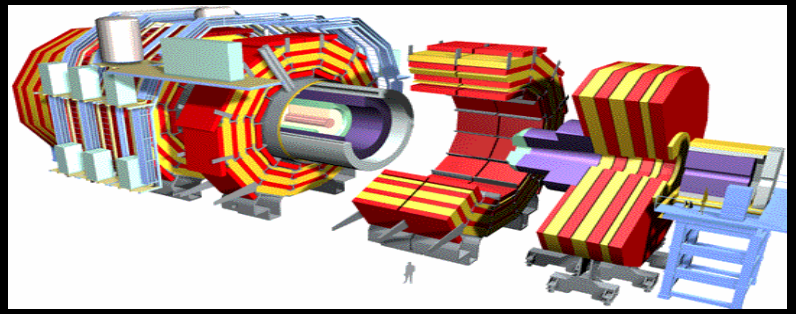
models and simulations



Experiments

simulation of detector and data

and DATA ANALYSIS



CMS

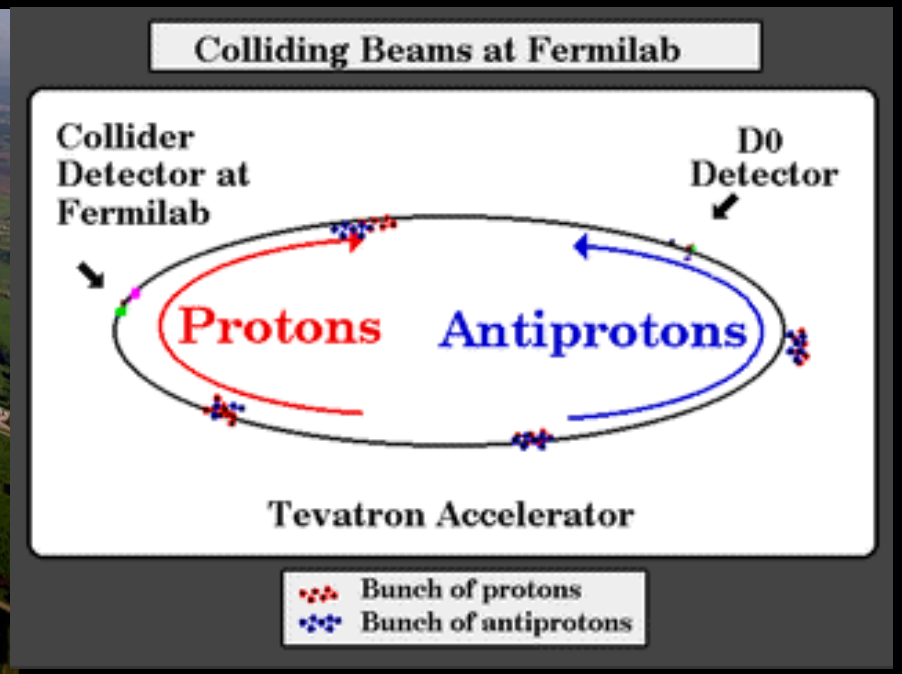
Computing At Fermilab is…

In support of the mission…

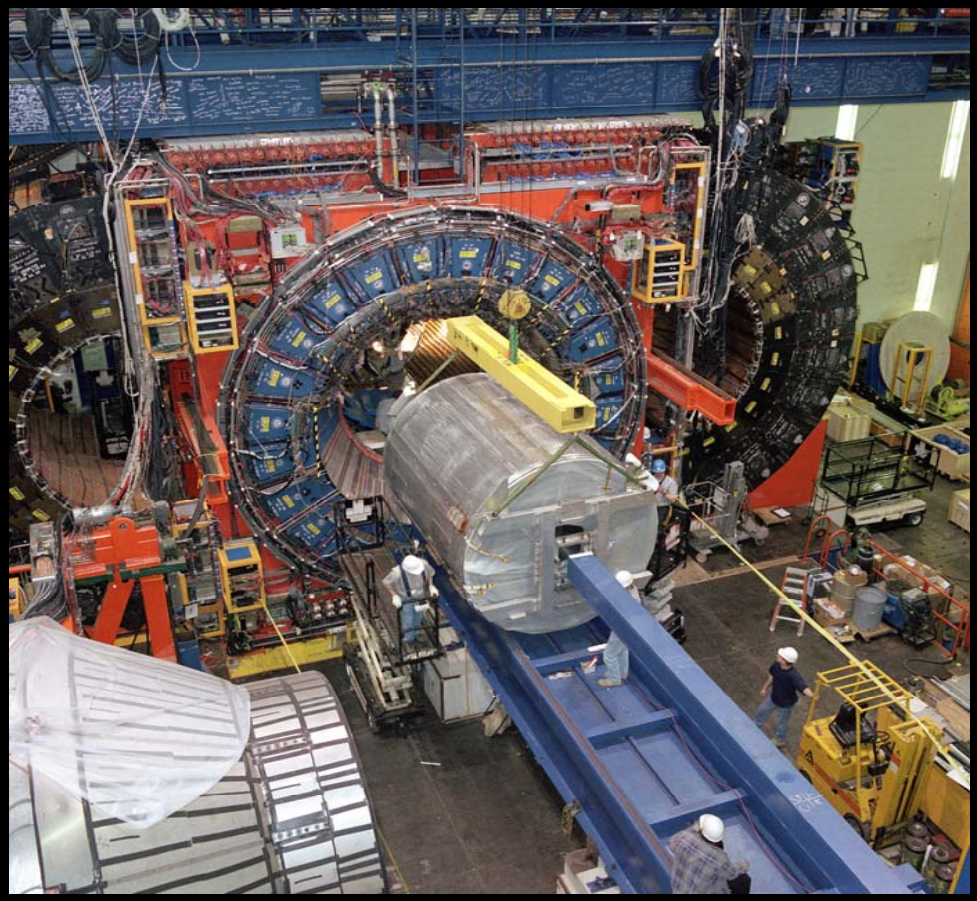
- to advance the understanding of the fundamental nature of
  - matter and energy

Computing Division
http://computing.fnal.gov

The Accelerator



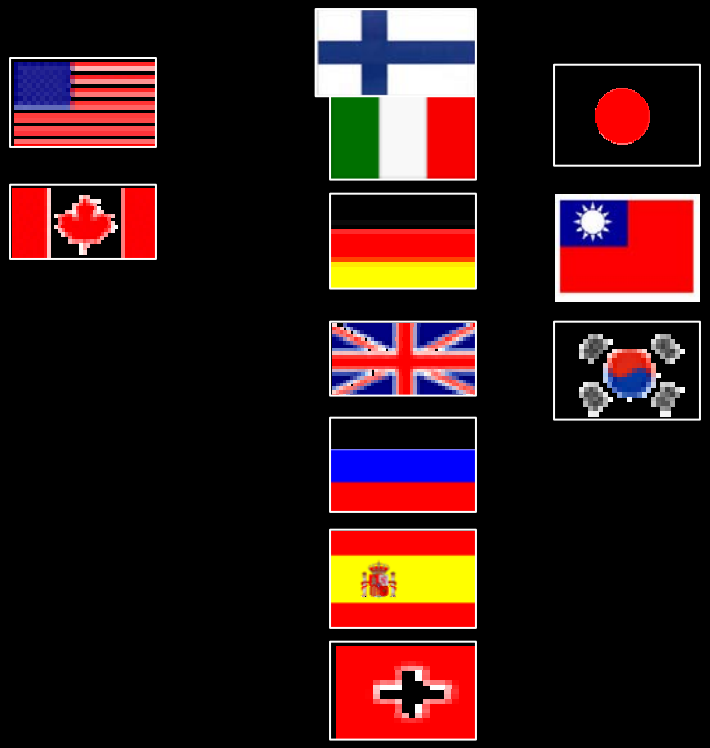
The CDF Collaboration

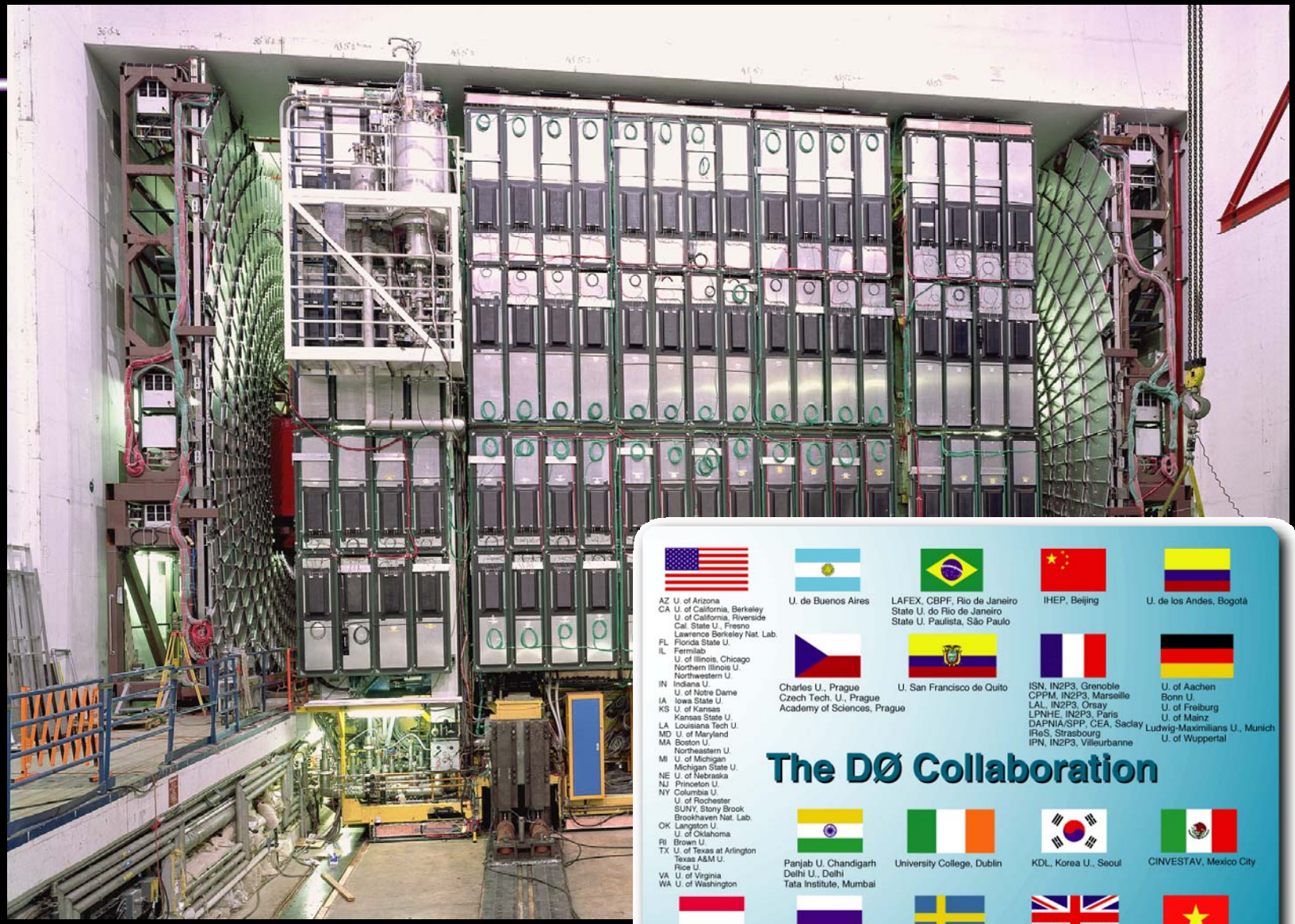


12 countries

59 institutions

706 physicists





The DØ Collaboration

AZ U. of Arizona
CA U. of California, Berkeley
U. of California, Riverside
Cal. State U., Fresno
Lawrence Berkeley Nat. Lab.
FL Florida State U.
IL Fermilab
U. of Illinois, Chicago
Northern Illinois U.
Northwestern U.
IN Indiana U.
U. of Notre Dame
IA Iowa State U.
KS U. of Kansas
Kansas State U.
LA Louisiana Tech U.
MD U. of Maryland
MA Boston U.
Northeastern U.
MI U. of Michigan
Michigan State U.
NE U. of Nebraska
NJ Princeton U.
NY Columbia U.
U. of Rochester
SUNY, Stony Brook
Brookhaven Nat. Lab.
OK Langston U.
U. of Oklahoma
RI Brown U.
TX U. of Texas at Arlington
Texas A&M U.
Rice U.
VA U. of Virginia
WA U. of Washington

U. de Buenos Aires

LAFEX, CBPF, Rio de Janeiro
State U. do Rio de Janeiro
State U. Paulista, São Paulo

IHEP, Beijing

U. de los Andes, Bogotá

Charles U., Prague
Czech Tech. U., Prague
Academy of Sciences, Prague

U. San Francisco de Quito

ISN, IN2P3, Grenoble
CPPM, IN2P3, Marseille
LAL, IN2P3, Orsay
LPNHE, IN2P3, Paris
DAPNIA/SPP, CEA, Saclay
IReS, Strasbourg
IPN, IN2P3, Villeurbanne

U. of Aachen
Bonn U.
U. of Freiburg
U. of Mainz
Ludwig-Maximilians U., Munich
U. of Wuppertal

Panjab U. Chandigarh
Delhi U., Delhi
Tata Institute, Mumbai

University College, Dublin

KDL, Korea U., Seoul

CINVESTAV, Mexico City

FOM-NIKHEF, Amsterdam
U. of Amsterdam / NIKHEF
U. of Nijmegen / NIKHEF

JINR, Dubna
ITEP, Moscow
Moscow State U.
IHEP, Protvino
PNPI, St. Petersburg

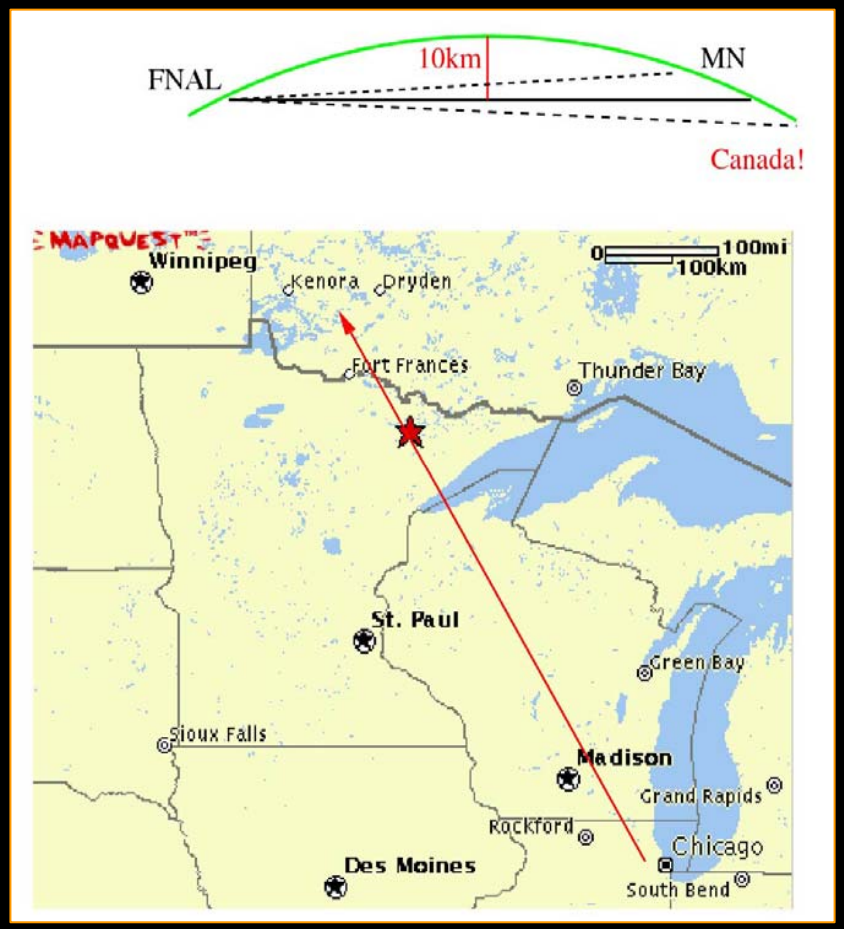
Lund U.
RIT, Stockholm
Stockholm U.
Uppsala U.

Lancaster U.
Imperial College, London
U. of Manchester

HCIP, Hochiminh City

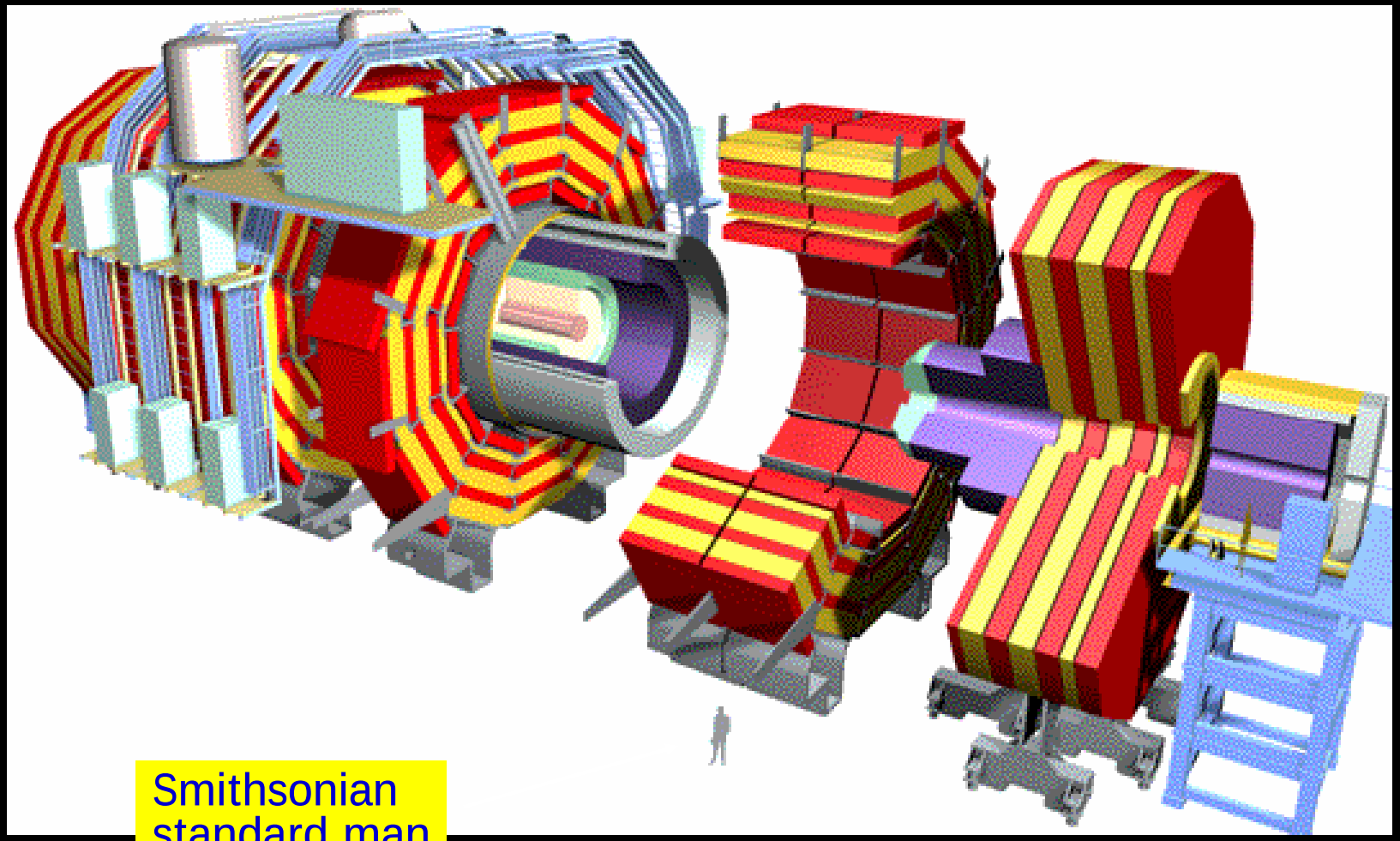
Ann Heinson, UC Riverside

MINOS Near and Far



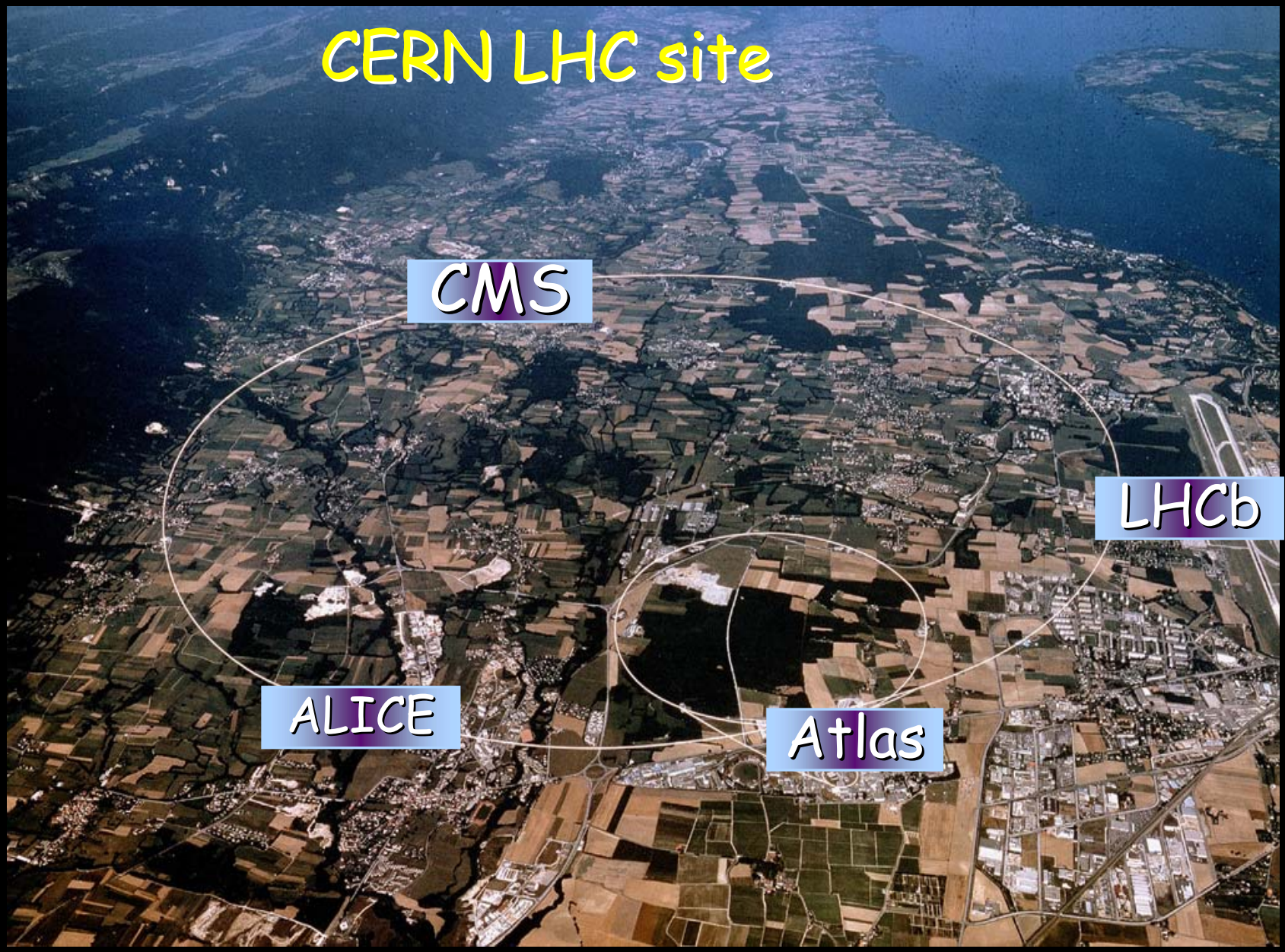
The Far Detector

"Compact" Muon Solenoid (CMS) at the LHC (CERN)



Smithsonian standard man

CERN LHC site



Astrophysics Experiments

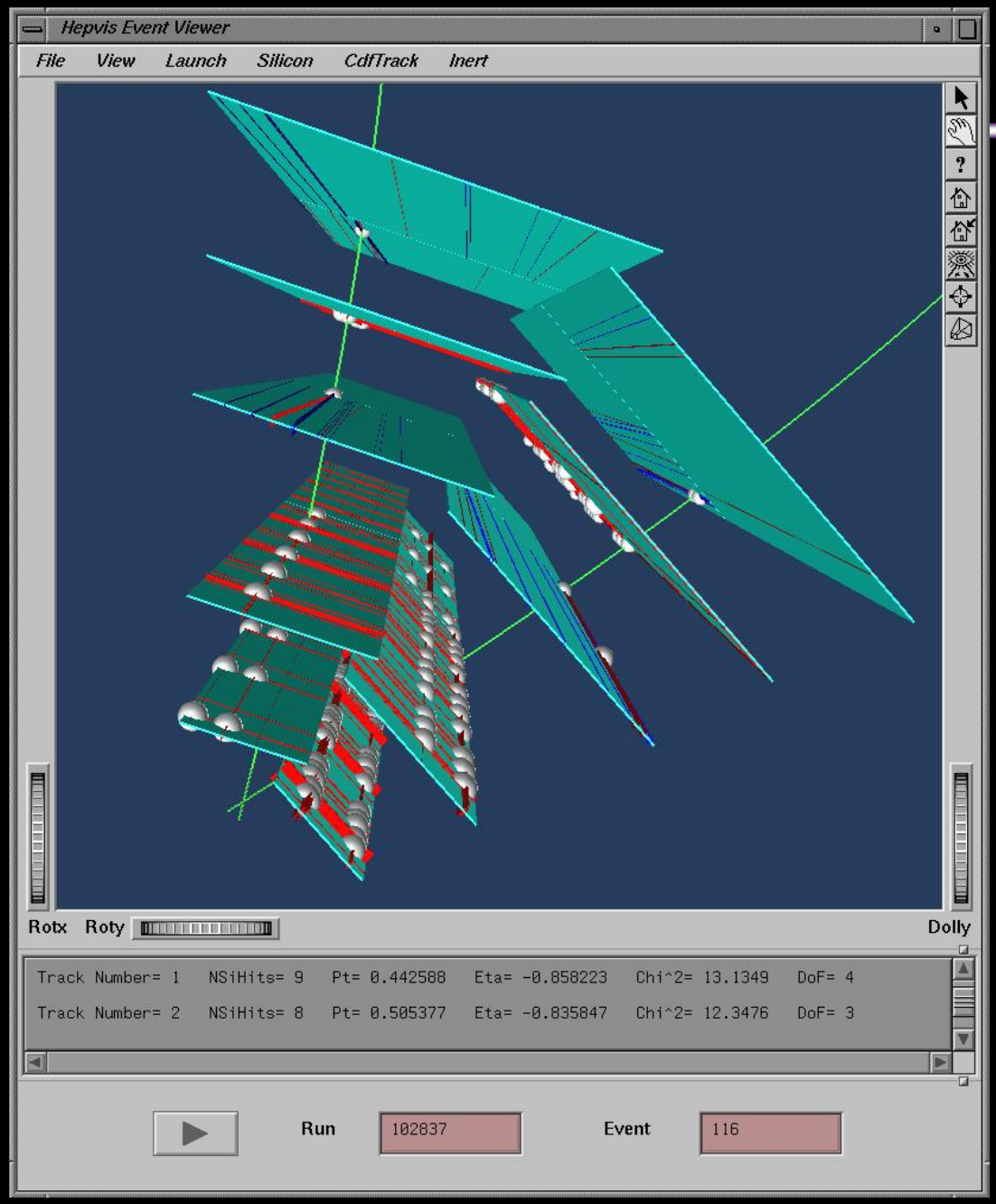
Sloan Digital Sky Survey

How Do You Do Science?

- 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**

Each stage of scientific process has different computing needs.

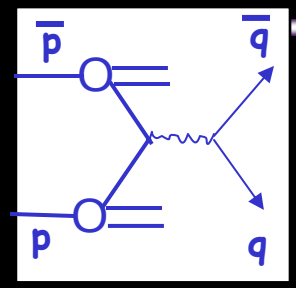
Repeat



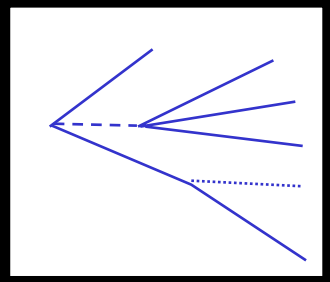
CDF Silicon Vertex Detector

A 3D modeling program simulating particles in the detector.

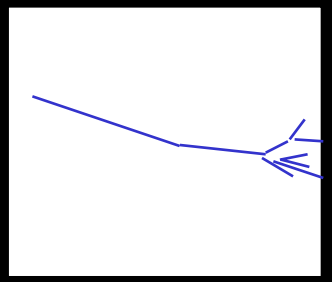
Physics → Raw Data



Basic Physics process

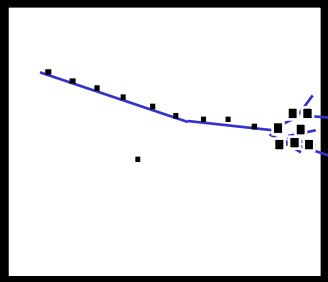


Fragmentation, Decay



Interaction with detector material

Multiple scattering, interactions



Detector Response

Noise, pile-up, cross-talk, inefficiency, ambiguity, resolution, alignment, temperature

2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

Raw data (Bytes)

channel addresses, ADC, TDC values, Bit patterns

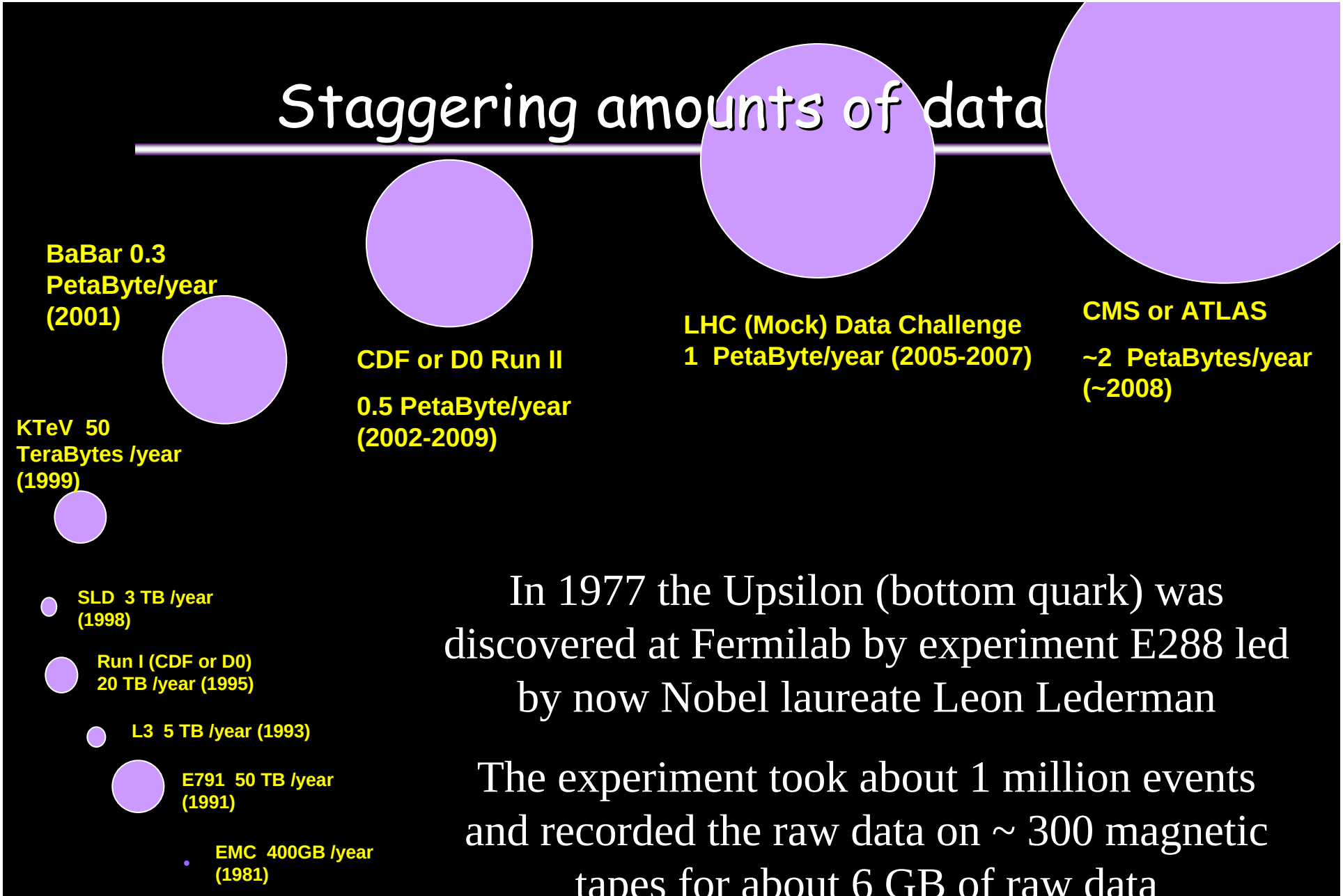
Recording Data

- Record data ("events")
  - The sensors and custom electronics convert the voltages and currents into several 100,000 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.

- We "trigger" on interesting events as well as filter out the uninteresting ones – ~few tens of MB/s.

- What gets through, is written to disk and eventually recorded on tape for permanent storage in a robotically controlled tape library

Staggering amounts of data



In 1977 the Upsilon (bottom quark) was discovered at Fermilab by experiment E288 led by now Nobel laureate Leon Lederman

The experiment took about 1 million events and recorded the raw data on ~ 300 magnetic tapes for about 6 GB of raw data

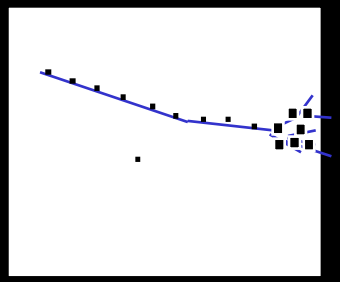
1 tape (2004) is ~200 GB

Raw Data → Physics

2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

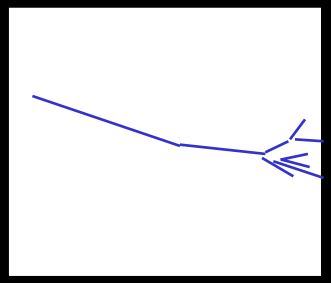
Raw data (bytes)

convert to physics quantities



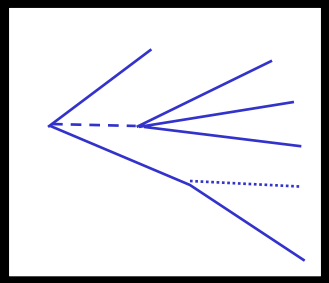
Detector Response

apply calibration, alignment,



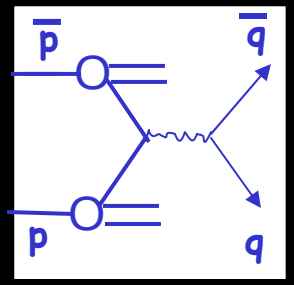
Interaction with detector material

pattern, recognition, particle identification



Fragmentation, Decay

physics analysis



Basic physics

results

Reconstruction

Analysis

Simulation (Monte-Carlo)

Computing at Fermilab

Computing Facilities

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

Vital Statistics

- FNAL has >10,000 network connected devices
- They are connected by over 1,000 miles of cabling
- There is more than 200TB of disk spinning now
- Our robotic tape libraries contain > 1.8 Petabytes of data

Learning the Lingo

Divide by		Multiply by
Milli	1,000	Kilo
Micro	1,000,000	Mega
Nano	1,000,000,000	Giga
Pico	1,000,000,000,000	Tera
Femto	1,000,000,000,000,000	Peta
Atto	1,000,000,000,000,000,000	Exa
Zepto	1,000,000,000,000,000,000,000	Zetta
Yocto	1,000,000,000,000,000,000,000,000	Yotta

Million million million million

Examples

- 1 page of text is approx 5 Kilobytes
- The complete works of Shakespeare is approx. 5 Megabytes
- A large home library contains approx 5 Gigabytes of text
- All the text is a college research library is approx 5 Terabytes
- If every person on the planet wrote as much as Shakespeare it would take approx 5 Petabytes
- All the data moved on the Internet this year is approx 5 Exabytes

Examples – going the other way

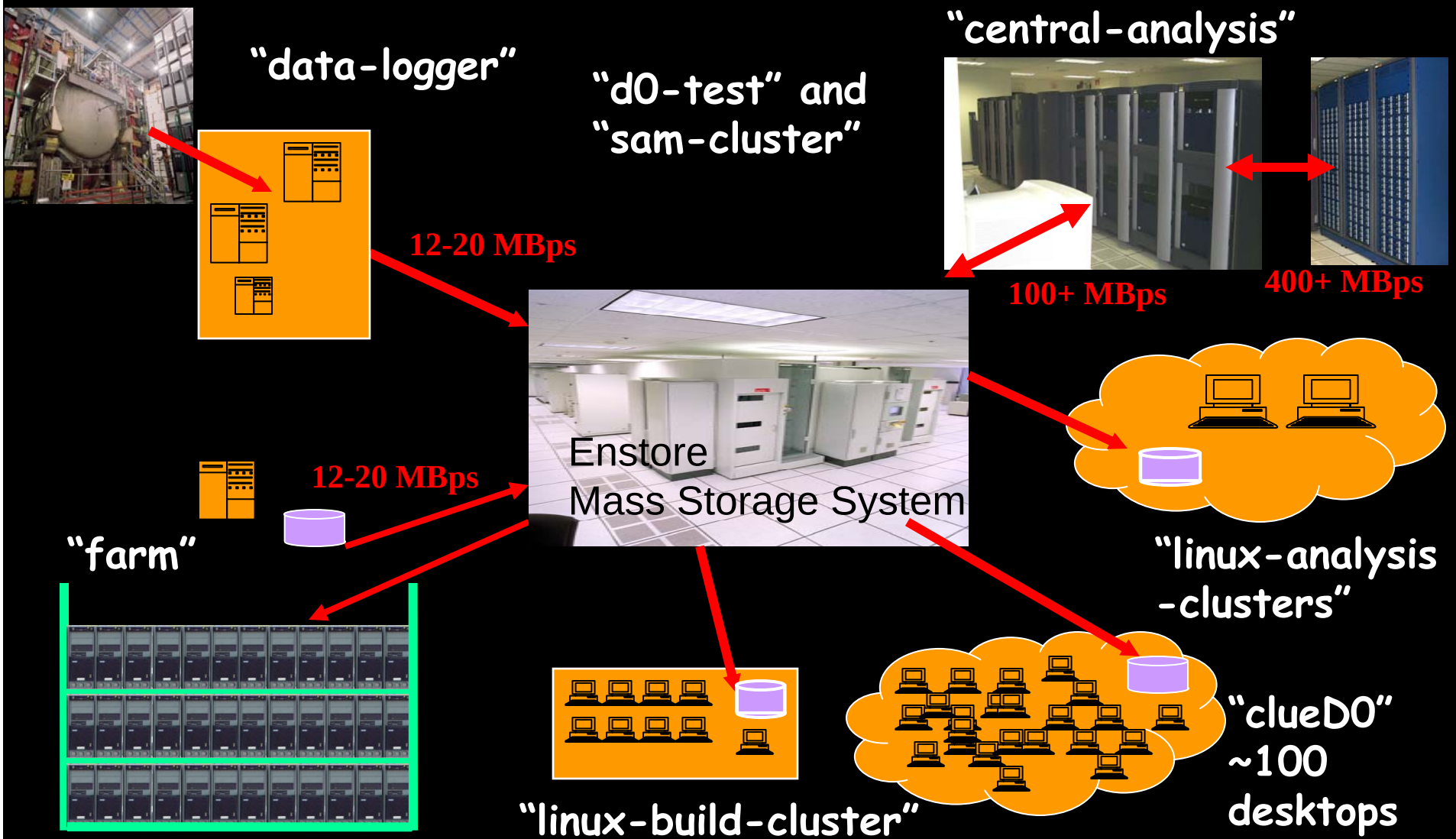
- A moderately fast camera lens click is 1 millisecond
  - (about the fastest thing the human eye can see)
- A redlined car engine piston moves noticeably in 1 microsecond
- Light travels about a foot in 1 nanosecond
- The fastest electronic switches operate in about 1 picosecond
- The fastest events measured directly are laser bursts of about 1 femtosecond
- (Shortlived particles in collisions live less than an attosecond)

PCs and PetaBytes

Mass Storage is "central"

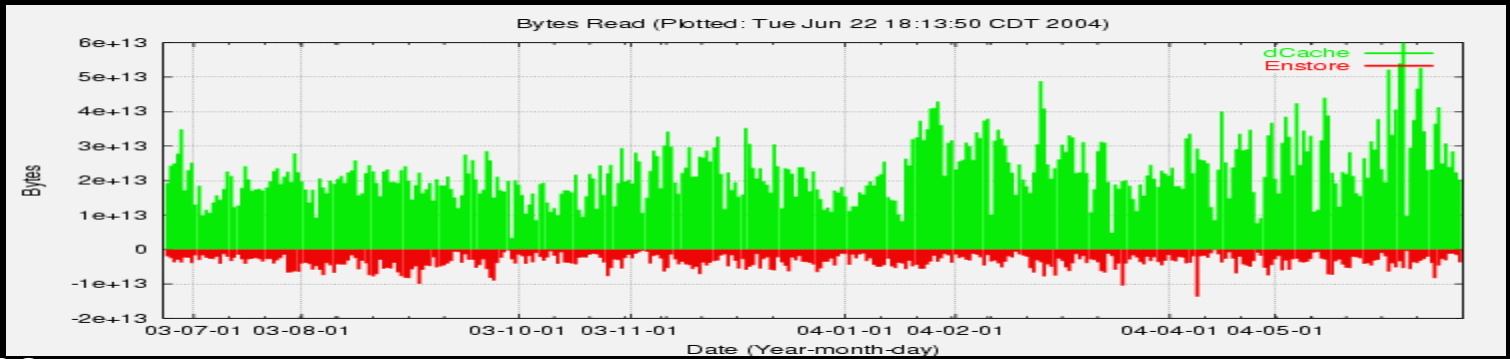
Scalable, IP network-attached storage systems that are widely accessible, including over the WAN, with "GRID" interfaces, are central to our data handling and data access strategies at Fermilab

D0 computing systems



State-of-the art Central Storage

CDF dCache
Data read
Often exceeds
40 TB/day –
Up to 60 TB/day

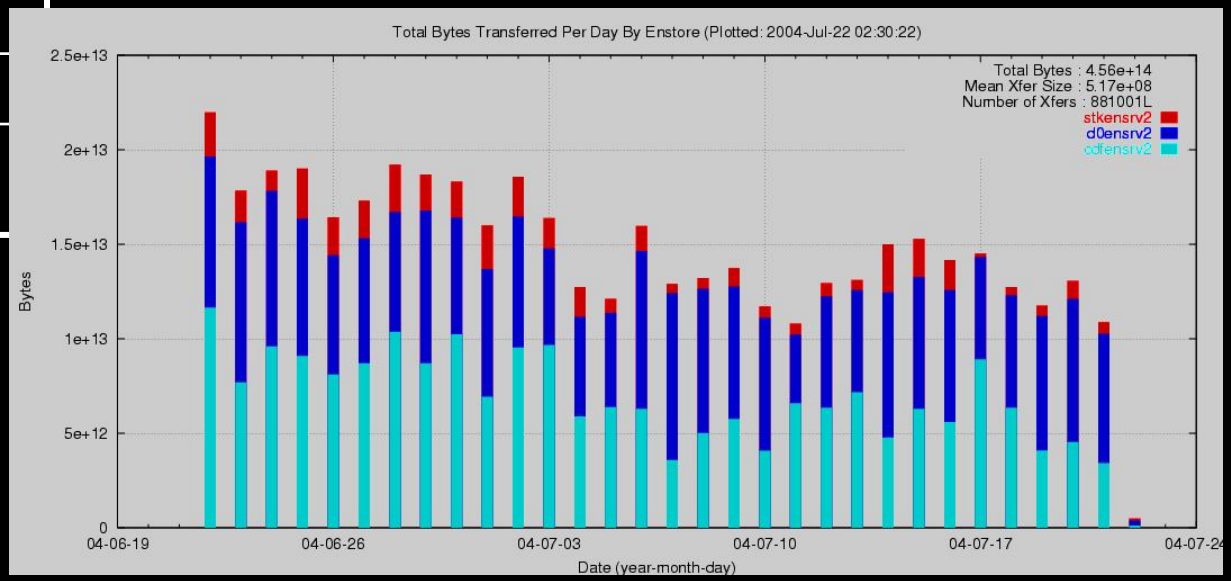


Data to tape

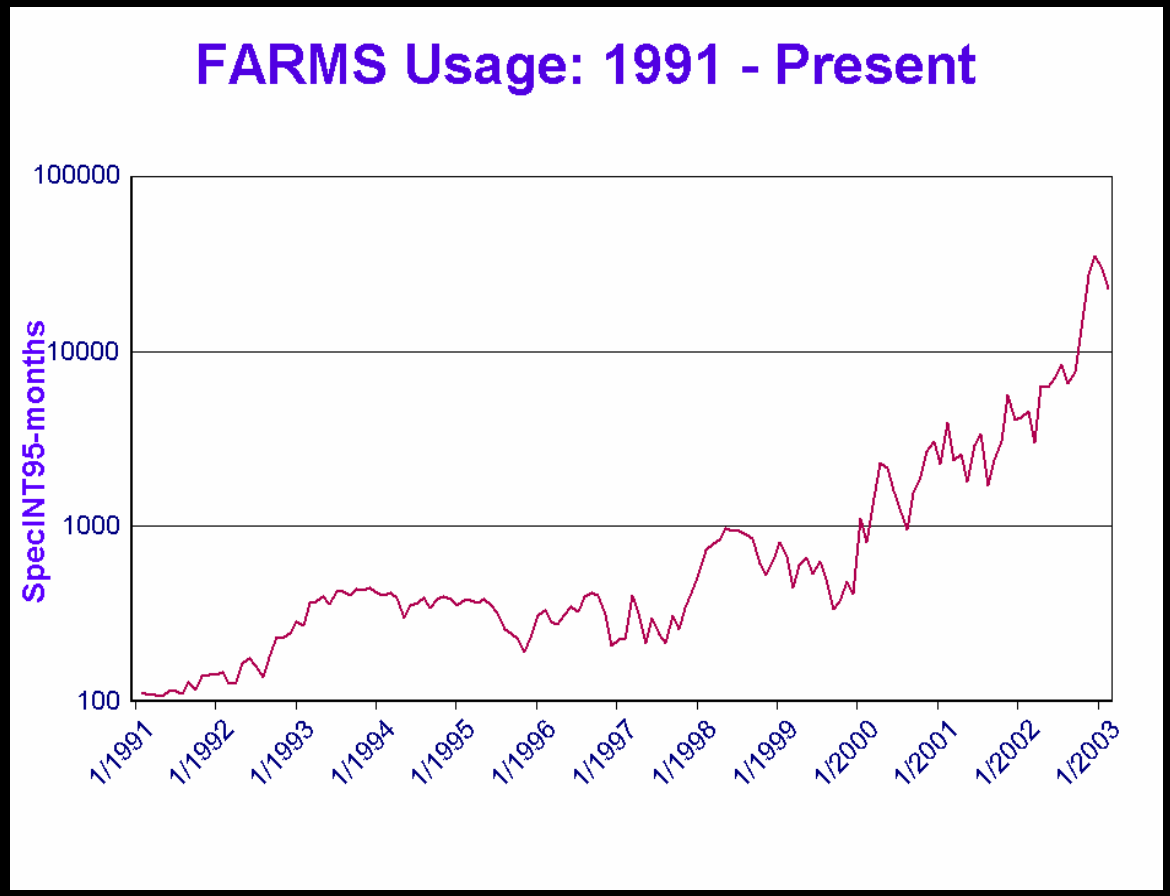
Experiment	TB
CDF	919
DO	729
CMS, MINOS,Miniboone, SDSS, Theory, legacy	145

Total Data on tape - 1.8 Petabytes

Total Tape I/O – peak of 22 TB/day



Farms (Production Processing resources)



CMS 1U

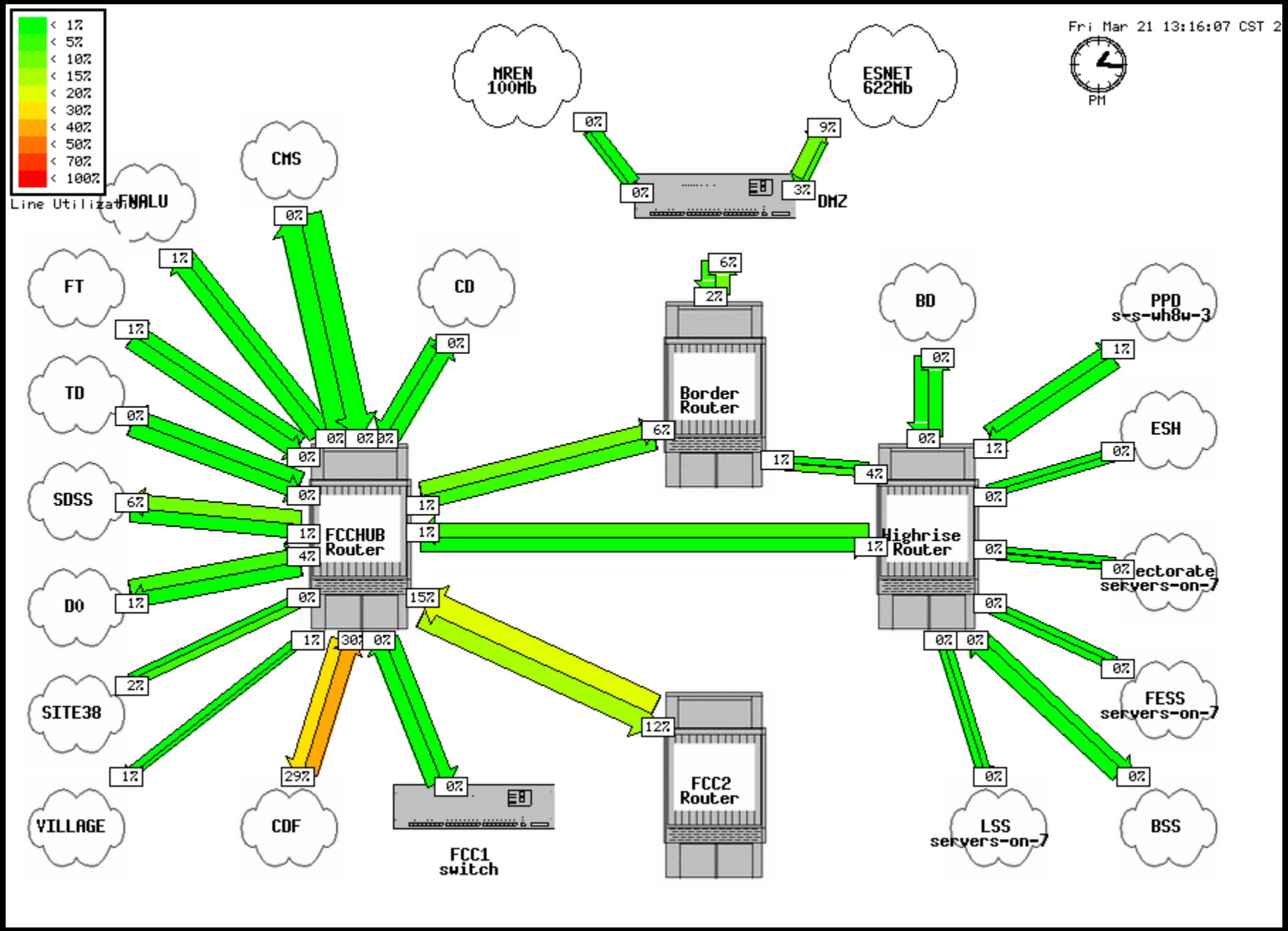
←Fixed Target →

← Run I →

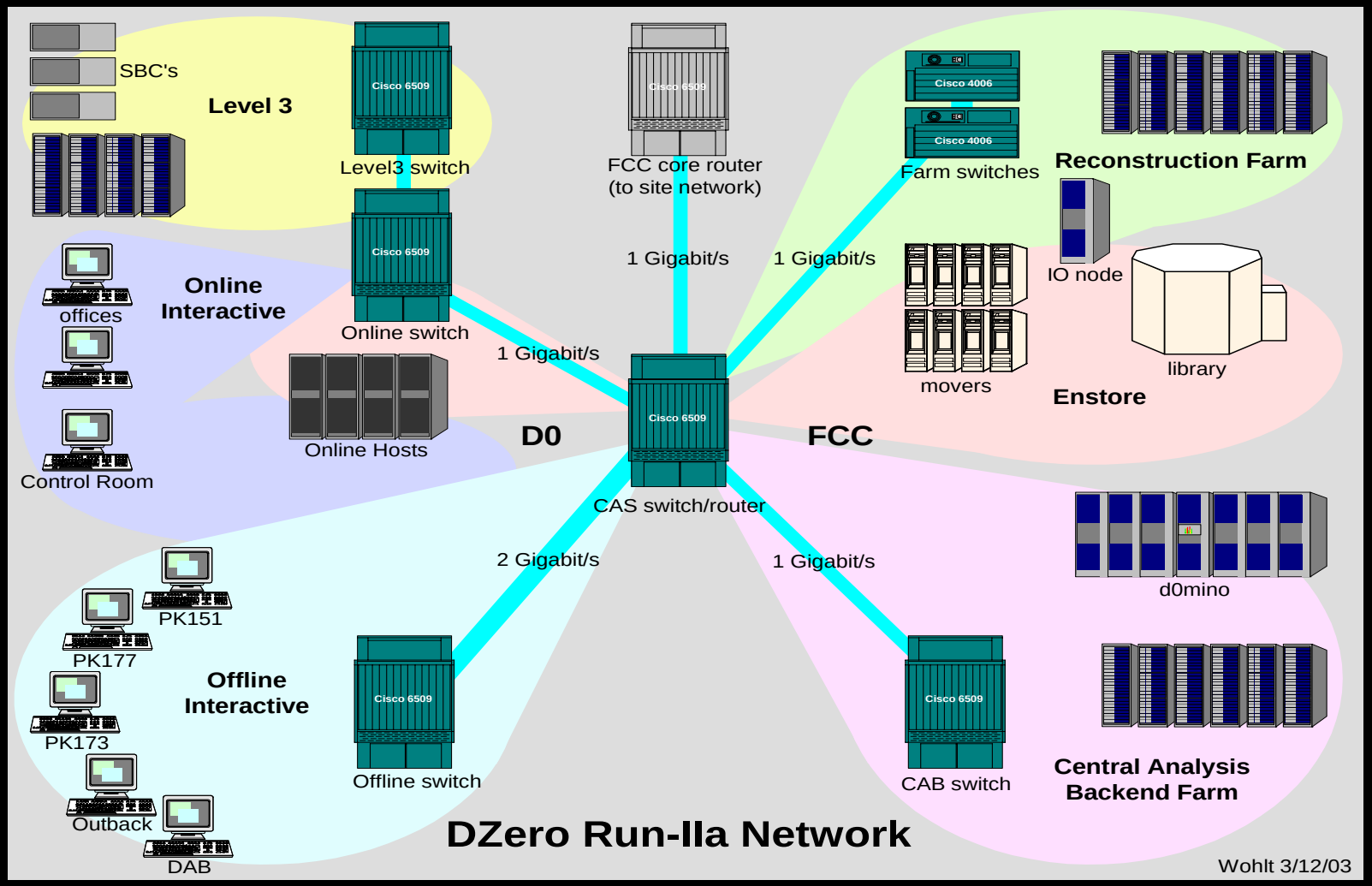
← Run II →

Analysis Computing/SMP/PC

- Analysis Computing in Run II is rapidly moving from big SMPs to PC based systems.
  - Driven by cost, performance, capabilities.
  - SMPs remain as fileservers and common shared environments.
  - Rapidly being phased out – high maintenance costs
- Original Run II design envisioned most analysis computing in SMP
  - Easier to operate, large I/O capabilities, shared environment.

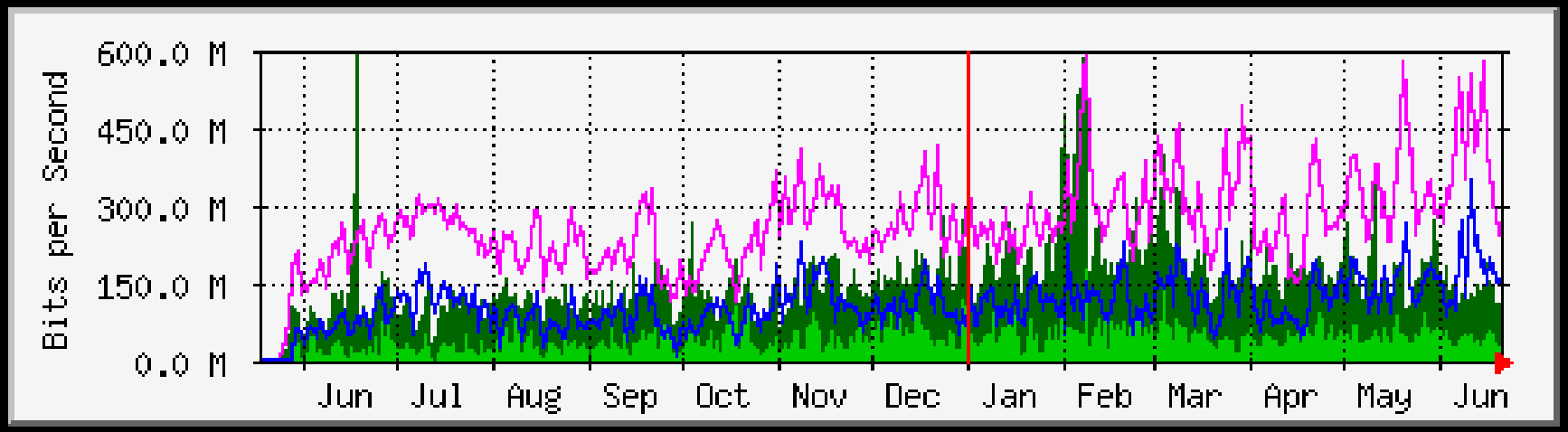

< 1%
< 5%
< 10%
< 15%
< 20%
< 30%
< 40%
< 50%
< 70%
< 100%
Line Utilization
Fri Mar 21 13:16:07 CST 2
PM
MREN 100Mb
ESNET 622Mb
DMZ
FNALU
CMS
FT
CD
TD
SDSS
DO
SITE38
VILLAGE
CDF
FCCHUB Router
FCC1 switch
Border Router
FCC2 Router
BD
PPD s-s-wh8w-3
ESH
Highrise Router
ectorate servers-on-7
FESS servers-on-7
LSS servers-on-7
BSS

D0 Run-II Network Topology



Networks are essential

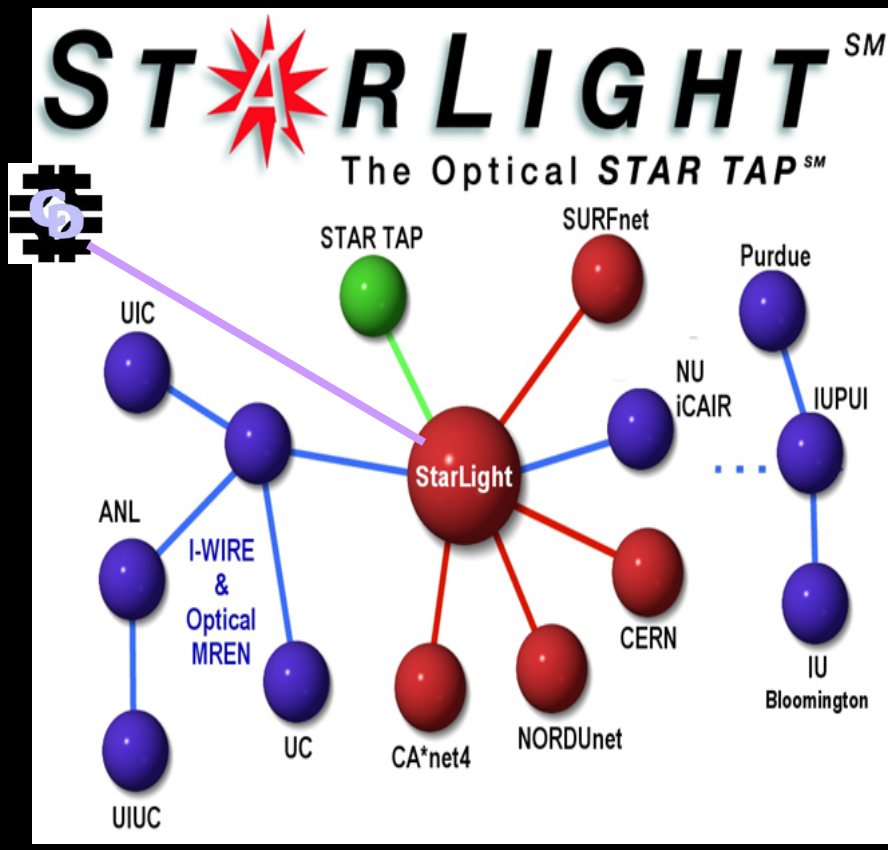
Fermilab Site production network traffic



GREEN	Incoming Traffic in Bits per Second
BLUE	Outgoing Traffic in Bits per Second
DARK GREEN	Maximal 5 Minute Incoming Traffic
MAGENTA	Maximal 5 Minute Outgoing Traffic

"FermiLight" – we are connected to Starlight with dark fiber

- News release recently
- In test mode now – looks good – 1 Gb/sec
- Will light up 3 wavelengths this year – potential for up to 10Gb/s each
- Potential for several partners to connect to FermiLight
- Potential for network R&D – 1 proposal funded



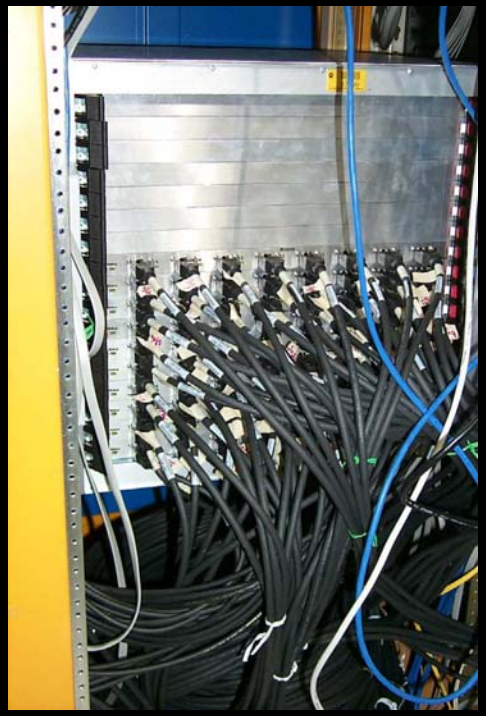
Optical networking interconnection point downtown Chicago.

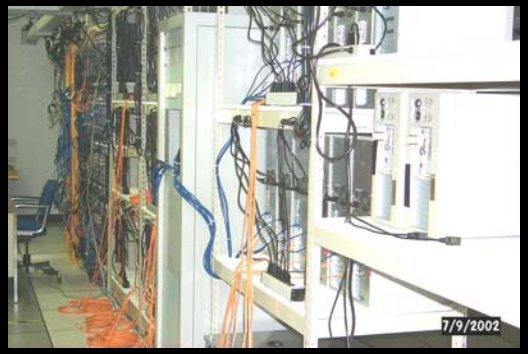
Owned by Northwestern University

Lattice QCD Computing

Muon Lab

Complex Interconnects





Lattice QCD "Commodity" Supercomputer

Satellite Computing Facilities to FCC

Wideband / HDCF on 6/17/2004 looking from red pole towards circuit breaker panels and air conditioners put in place.

Run II meets Lattice QCD – building re-use

CD New Muon Lab Facility
Lattice QCD on left & CDF/DZero on right

High Density Computing Facility Under construction – delivery in the fall. 2500 KVA facility

Grid Computing - What and Why

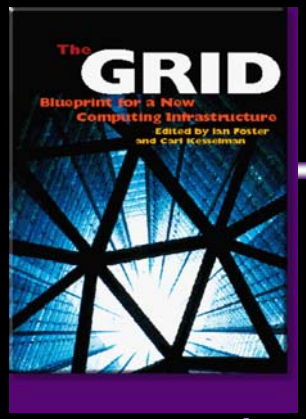
What is GRID computing :

Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. [Ian Foster]

- A virtual organization is a collection of users sharing similar needs and requirements in their access to processing, data and distributed resources and pursuing similar goals.

Key concept :

- **ability to negotiate resource-sharing arrangements among a set of participating parties (providers and consumers) and then to use the resulting resource pool for some purpose.** [Ian Foster]



Ian Foster and Carl Kesselman, editors,

"The Grid: Blueprint for a New Computing Infrastructure,"

Morgan Kaufmann, 1999,

http://www.mkp.com/grids

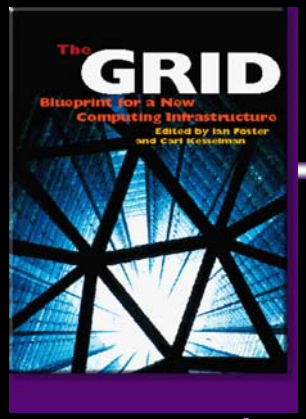
Who is this guy Ian Foster?

Computer Scientist at University of Chicago and Argonne National Lab

Leader of sequence of efforts in distributed computing ➔ Globus Project

Globus is a toolkit of software that helps do some of the things you need to do to make Grid Computing a reality

It is not the only software offering



Ian Foster and Carl Kesselman, editors,

"The Grid: Blueprint for a New Computing Infrastructure,"

Morgan Kaufmann, 1999,

http://www.mkp.com/grids

And what is he talking about?

An idea of how to work using distributed computers and storage devices and disks and networks in a global way based on emerging standards for interoperability

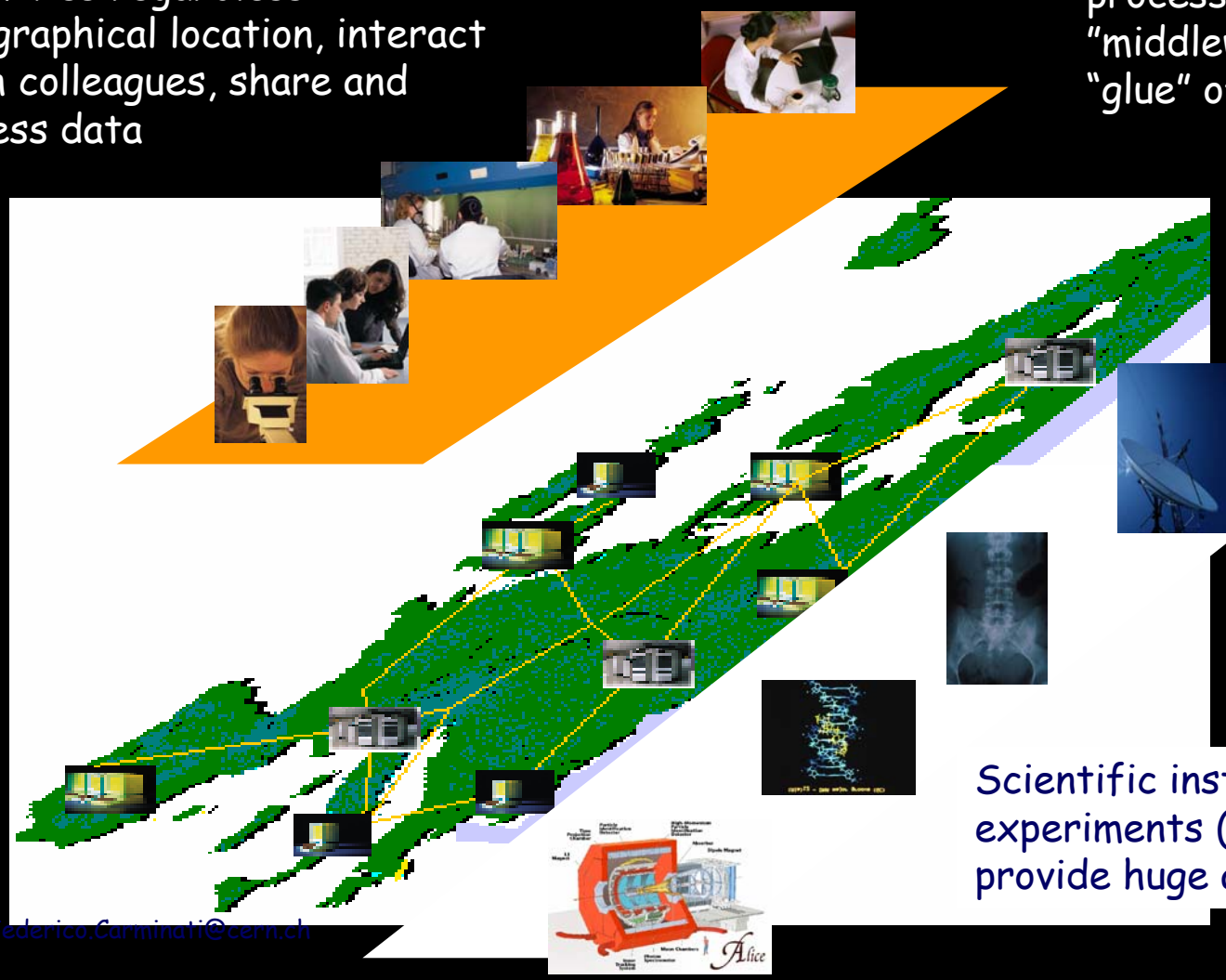
- Based on the metaphor of the Power Grid – you just plug into the wall and get power wherever you go!

(but its not quite as simple as that)

The Grid Vision

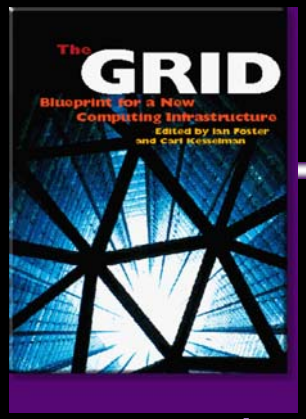
Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

The GRID: networked data processing centres and "middleware" software as the "glue" of resources.



Scientific instruments and experiments (and simulations) provide huge amount of data

Federico.Carminati@cern.ch



Ian Foster and Carl Kesselman, editors,

"The Grid: Blueprint for a New Computing Infrastructure,"

Morgan Kaufmann, 1999,

http://www.mkp.com/grids

What does Grid Computing involve?

- some specific Grid software that is evolving. Often called "Grid Middleware"
- Funding opportunities and some hype
  - Many funded grid projects all over the world
- An opportunity for Computer Scientists and Scientists with a job to do (e.g. Physicists) to get together and make something work
- Potential broad benefit to society. Vendor interest has peaked.
  - Is this what comes after the web?
- Lots and lots of buzzwords to learn

About "The Grid"

- There is no single "Grid" a la the Internet
  - Many Grids, each devote to different organizations
- Grids are (or soon will be)
  - The foundation on which to build secure, efficient, and fair sharing of computing resources
- Grids are not
  - Sources of free computing
  - The means to access and process Petabyte-scale data freely without thinking about it

What's different about Grids and are they useful now?

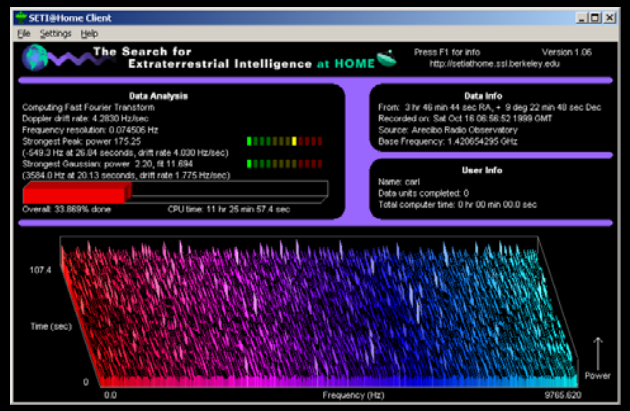
- Lots of production quality **Distributed Applications** in real life
- Physics is leading the way among sciences and making much progress on building real grids, prototypes, testbeds, even spanning continents
  - **Our Virtual Organizations are our Experiments**

Real-world example - Walmart Inventory Control

- Satellite technology used to track every item
  - Bar code information sent to remote data centers to update inventory database and cash flow estimates
  - Satellite networking used to coordinate vast operations
- Inventory adjusted in real time to avoid shortages and predict demand
  - *Data management, prediction, real-time, wide-area synchronization*

More Real World Distributed Applications

- SETI@home
  - 5M users in 226 countries
  - ~1500 CPU years/day
  - ~38 TF sustained (Japanese Earth Simulator is 40 TF peak)
  - several ZETAflop over last 3 years (10^21, beyond peta and exa …)
  - Highly heterogeneous: >150 *different* processor types



Grid for Science

- Walmart – controlled/owned distributed computing system
- SETI - other end of the spectrum
  - Same "job" run on anyone's desktop
- Science using the GRID –
  - Focus on standards and interfaces, not all resources owned and controlled
  - Sharing with other disciplines
  - Opportunistic use of resources for ever-changing jobs and investigations

Vendors and the Grid

- IBM
- Oracle
- SUN
- STK
- And many, many others
- All making GRID part of their strategic direction and marketing

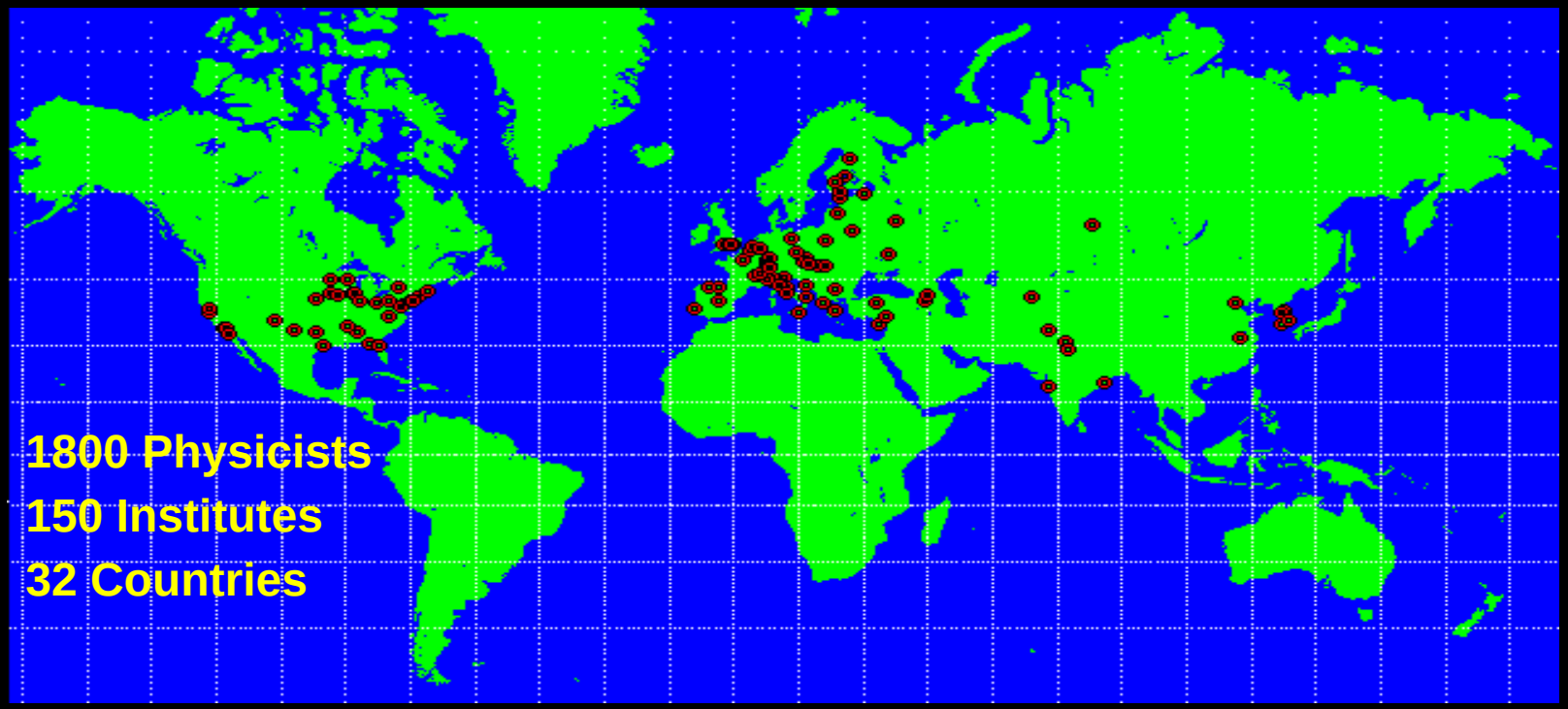
Why GRIDs for HEP and other sciences

- The **scale** of the problems *human beings* have to face to perform frontier research in many different fields **is constantly increasing**.

- Performing frontier research in these fields already today requires world-wide collaborations (i.e. multi domain access to distributed resources).

LHC Computing Overview

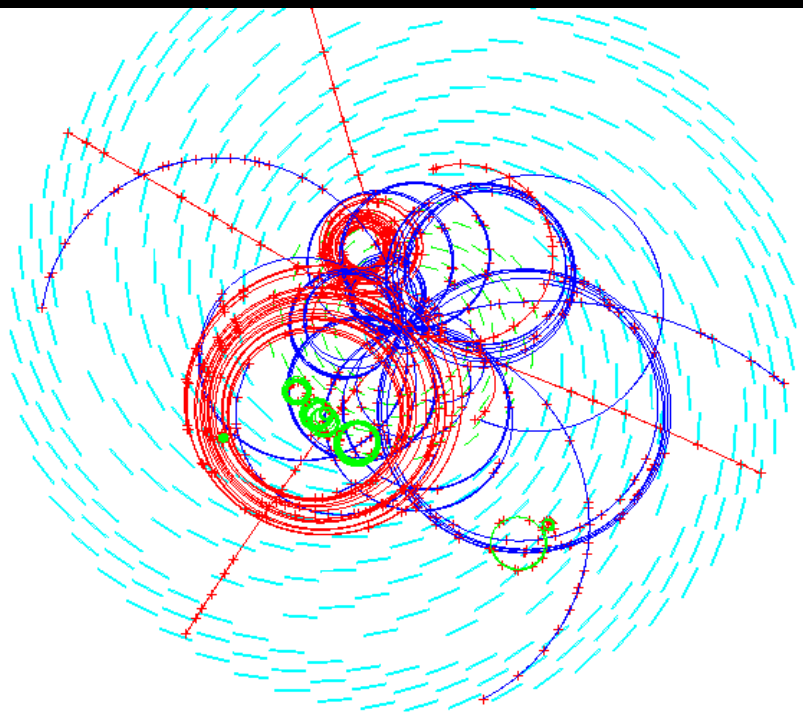
- Complexity: Millions of individual detector channels
- Scale: PetaOps (CPU), Petabytes (Data)
- Distribution: Global distribution of people & resources



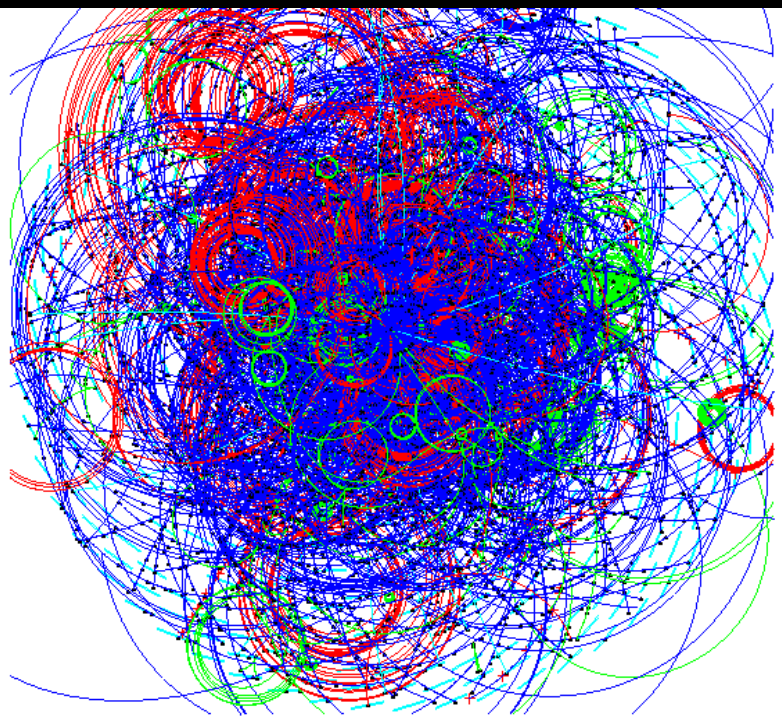
LHC Data Complexity

- "Events" resulting from beam-beam collisions:
  - Signal event is obscured by 20 overlapping uninteresting collisions in same crossing
  - CPU time does not scale from previous generations

2000

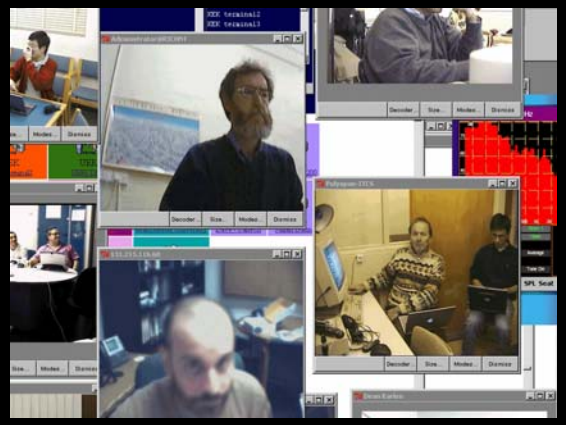


2007



LHC Physics Discoveries at Universities and Labs!

U.S. CMS is Committed to Empower the CMS Scientists at U.S. Universities and Labs to do Research on LHC Physics Data



This is why we are pushing Grids and other Enabling Technology

Lothar Bauerdick – U.S. CMS Software and Computing Project Manager

Grid Computing Tour

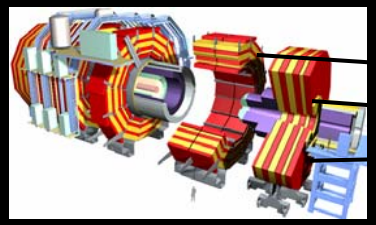
Closer to home – what are we doing at Fermilab?

- SAM-GRID for Run II Experiments
- Partipating in Grid projects and submitting proposals in the areas of
  - Grid, Cyber Security, Network Research, Collaborative Tools, Global Analysis, Lattice QCD, Sloan Digital Sky Survey
- US-CMS Software and Computing leadership
  - LHC Computing Grid
- Open Science Grid - http://www.opensciencegrid.org
- Playing a big role in coordination of Grid Efforts
  - Trillium - joining together 3 US HEP Grid Projects
  - Grid2003 http://www.ivdgl.org/grid2003
- Grid-enabled storage systems using standards like SRM (Storage Resource Manager)
- Network and Storage system research to sustain high rate data movement

Global LHC Data Grid

Experiment (e.g., CMS, ATLAS)

Tier0/(Σ Tier1)/(Σ Tier2) ~1:1:1



Online System

~100 MBytes/sec

Tier 0

CERN Computer Center > 20 TIPS

Tier 1

2.5 Gbits/sec

France

UK

Italy

USA

• • •

2.5 Gbits/sec

Tier 2

Tier2 Center

~0.6 Gbits/sec

Tier 3

Institute ~0.25TIPS

0.1 - 1 Gbits/sec

Physics data cache

Tier 4 PCs, other portals

What's SAM-GRID?

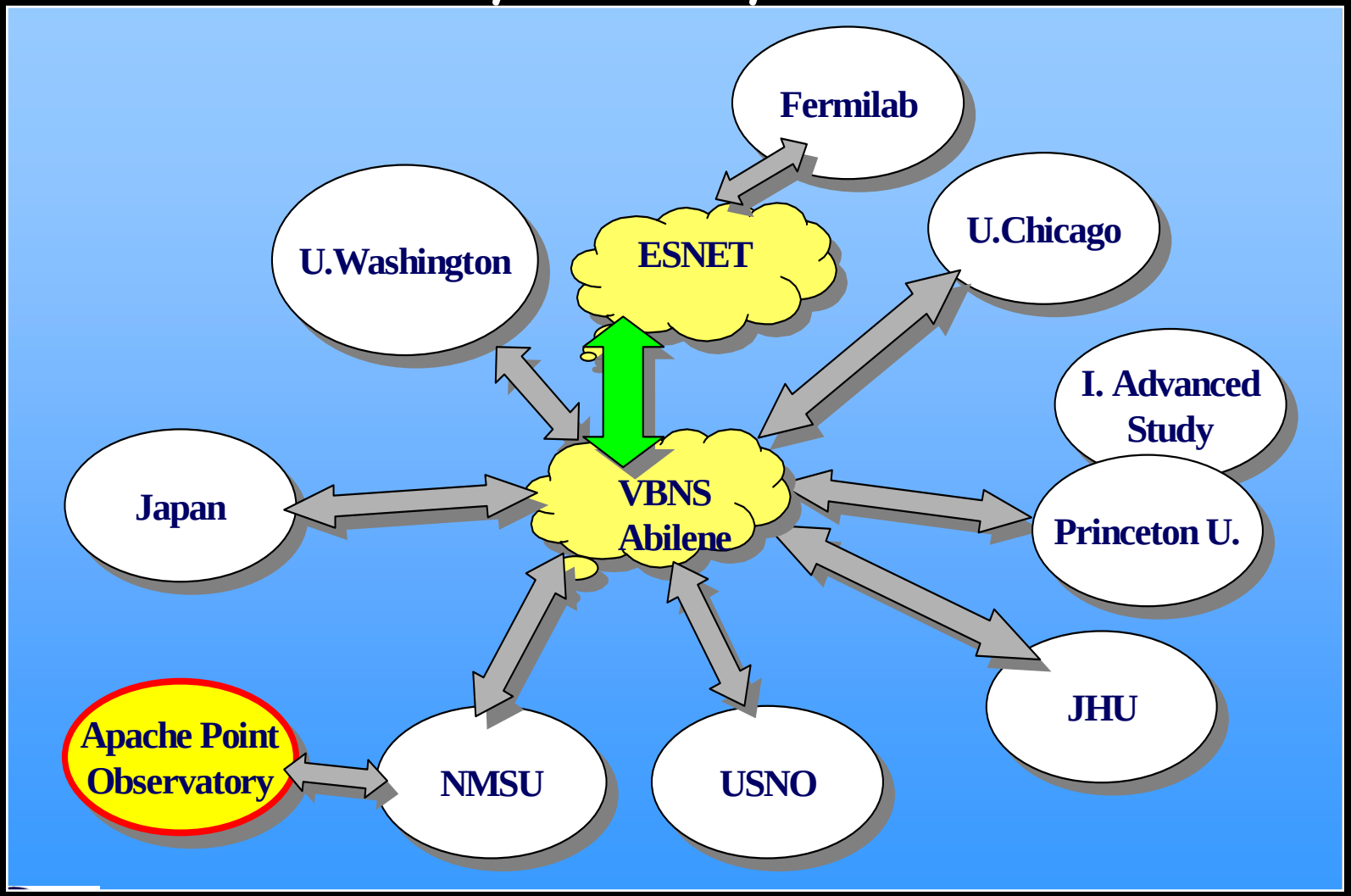
- SAM Sequential data Access via Meta-data is a data handling system with a Grid job management services (JIM) being added
- **http://d0db.fnal.gov/sam**
- Joint project between D0 and Computing Division started in 1997 to meet the FNAL Run II data handling needs
- Now used by CDF, D0 and MINOS
- Globally distributed system
- Provides high-level collective services of *reliable* data storage and replication
  - **Arguably the most functional "Grid" in HEP currently**

SAM-GRID Monitoring

- http://samgrid.fnal.gov:8080/

- http://d0db.fnal.gov/sam/

Sloan Sky Survey Data Grid

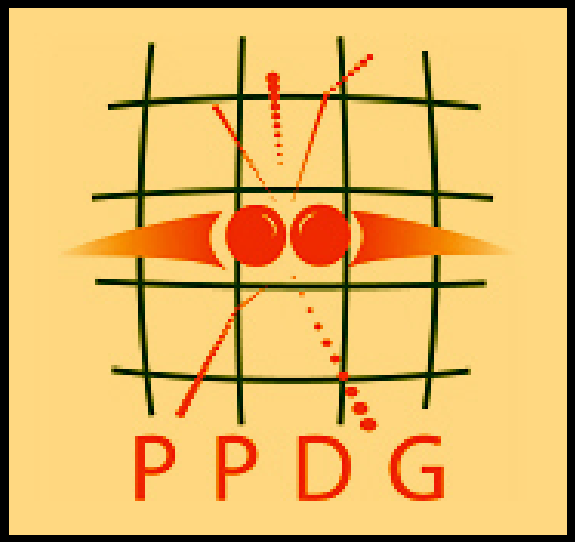


Enormous amount going on with Grid Computing

- Here are just a few of the projects most closely related to Physics – but not all of them
  - And just a flying visit so you can go back if you are interested
- General Grid stuff and Middleware at
  - http://www.globus.org/
  - http://www.cs.wisc.edu/condor/
  - http://www.globalgridforum.org/
  - http://sdm.lbl.gov/srm-wg/

Particle Physics Data Grid

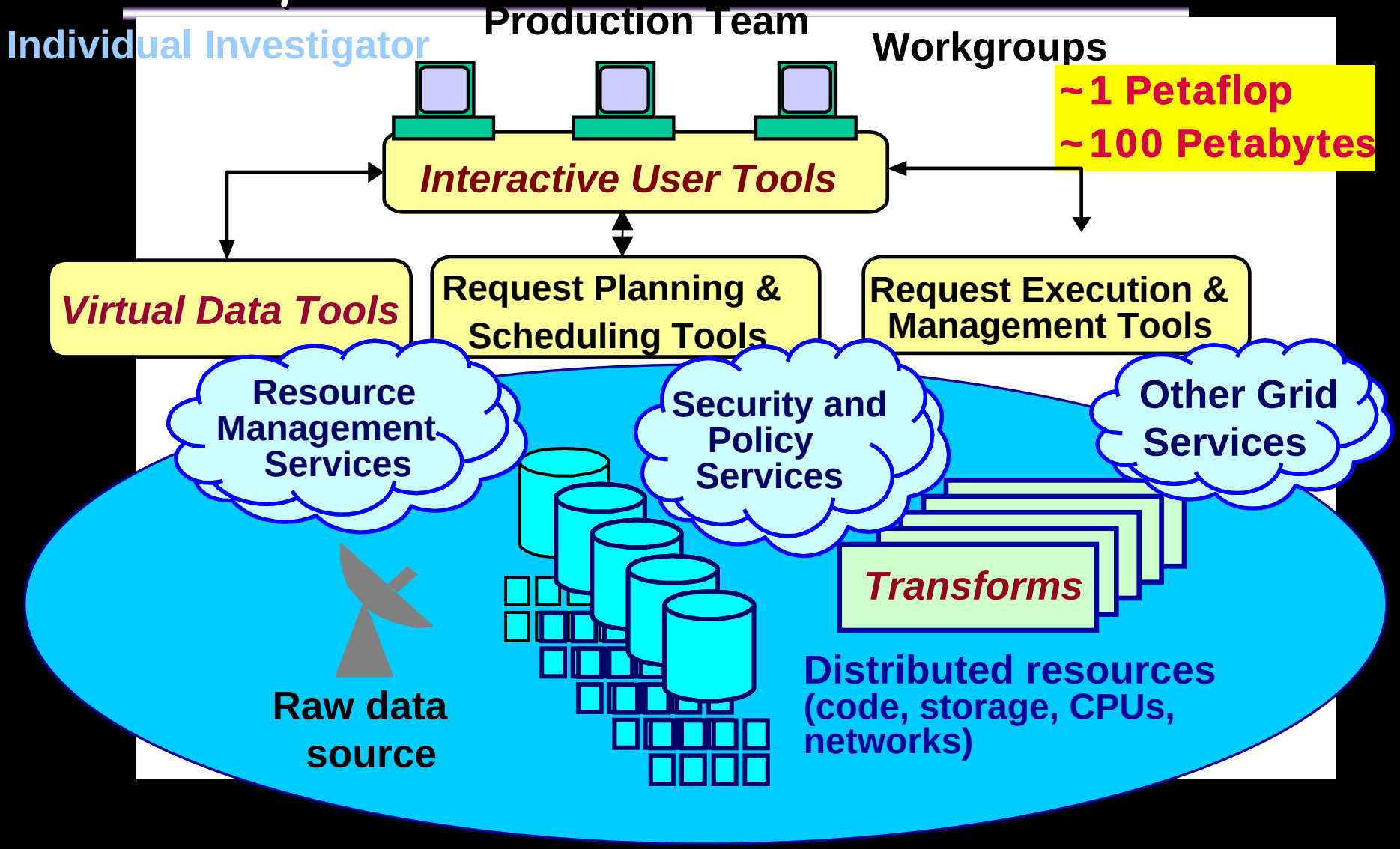
- Funded by Department of Energy
- Practical orientation: networks, instrumentation, monitoring
- D0, Babar, ATLAS, CMS, Jlab, STAR, CDF
- http://www.ppdg.net/



Computer Science Program of Work

- **CS1: Job Description Language**
- **CS2: Schedule and Manage Data Processing & Placement Activities**
- **CS3 Monitoring and Status Reporting**
- **CS4 Storage Resource Management**
- **CS5 Reliable Replication Services**
- **CS6 File Transfer Services**
- **....**
- **CS11 Grid-enabled Analysis**

GriPhyN: PetaScale Virtual-Data Grids



http://www.griphyn.org/

iVDGL: A Global Grid Laboratory

"We propose to create, operate and evaluate, over a sustained period of time, an international research laboratory for data-intensive science."

From NSF proposal, 2001

- International Virtual-Data Grid Laboratory
  - A global Grid laboratory (US, EU, Asia, South America, …)
  - A place to conduct Data Grid tests "at scale"
  - A mechanism to create common Grid infrastructure
  - A laboratory for other disciplines to perform Data Grid tests
  - A focus of outreach efforts to small institutions
- U.S. part funded by NSF (2001-2006)
  - $13.7M (NSF) + $2M (matching)
  - UF directs this project
  - International partners bring own funds

Grid2003 Project and Grid3 Grid

Grid3: Demonstrator for a U.S.-wide Grid

"Robust" (but "minimal") middleware basis for next step: Grid2003 Project

Joint Project to build a first common grid environment to:

- Provide the infrastructure and services needed for LHC production and analysis applications running **at scale** in a **common** grid environment
- Provide the next phase of the iVDGL Laboratory
- Provide a platform for computer science technology demonstrators
- Provide a common grid environment for LIGO and SDSS applications, and possibly others

Demonstrate the functionalities and capabilities of this Grid,

- according to some **well-defined metrics**
  - – thousands of processes, efficiency, throughput of TBytes/day, etc
- specific CMS (and ATLAS, SDSS, LIGO, CS) applications
  - – event simulation, data streaming,
- demonstrator milestone coincided with the SC2003 conference

Grid3 today

- http://www.ivdgl.org/grid2003/

LHC Computing Grid related

- http://lhcgrid.web.cern.ch/lcg
- European Commission funded EGEE (Enabling Grids for E-Science in Europe project)
  - http://egee-intranet.web.cern.ch/egee-intranet/gateway.html
- UK Physics Grid – GridPP
  - http://www.gridpp.ac.uk/

Need for Common Grid Infrastructure

- Grid computing sometimes compared to electric grid
  - You plug in to get a resource (CPU, storage, …)
  - You don't care where the resource is located

This analogy is more appropriate than originally intended

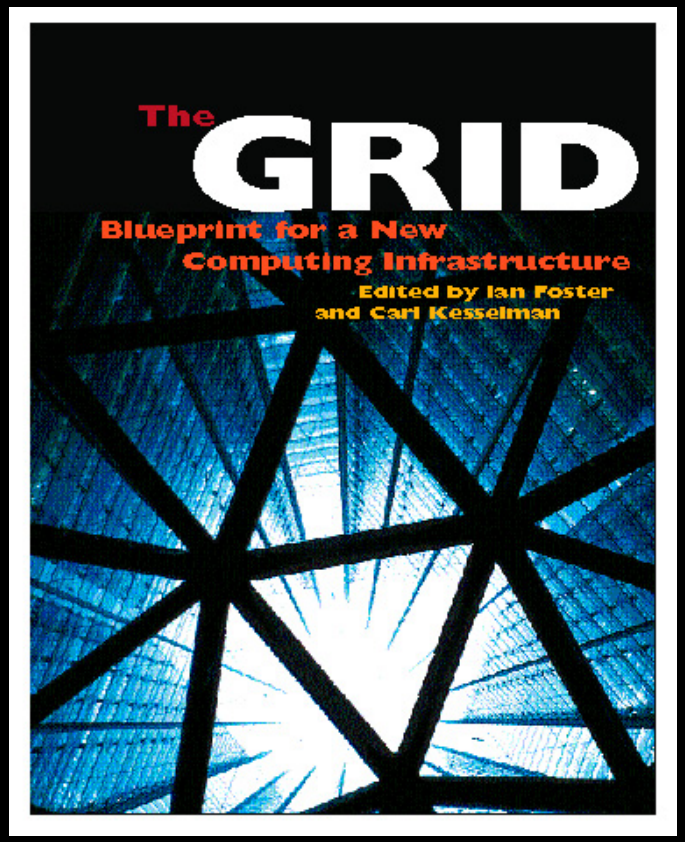
- It expresses a USA viewpoint ⇒ uniform power grid
- What happens when you travel around the world?

Different frequencies	60 Hz, 50 Hz
Different voltages	120 V, 220 V
Different sockets!	USA, 2 pin, France, UK, etc.

Want to avoid this situation in Grid computing

Other Grid References and "what is Grid" places

- Grid Book
  - www.mkp.com/grids
- Grid Café
  - http://gridcafe.web.cern.ch/gridcafe/
- http://www.gridcomputing.com/
- http://www-unix.mcs.anl.gov/%7Ewilde/summer-grid/



Will the Grid help us with Global Science projects

I think it can and it is. High Energy Physics is prime example and a real driving force for global scale science projects with massive data and huge data processing needs

The promise of the Grid has been not been oversold but the difficulty of developing the requisite Grid infrastructure has been underestimated.

It is a new way of working – a bit less focused on HEP but with a greater chance to see and be seen and have an impact more broadly than our science and more aligned with the long-term standards and developments in Information Technology

The End