HEP Data Grids, the LHC and Global Networks

Harvey B. Newman

California Institute of Technology

AMPATH Meeting, Miami

August 16, 2001

http://l3www.cern.ch/~newman/AMPATH82001.ppt

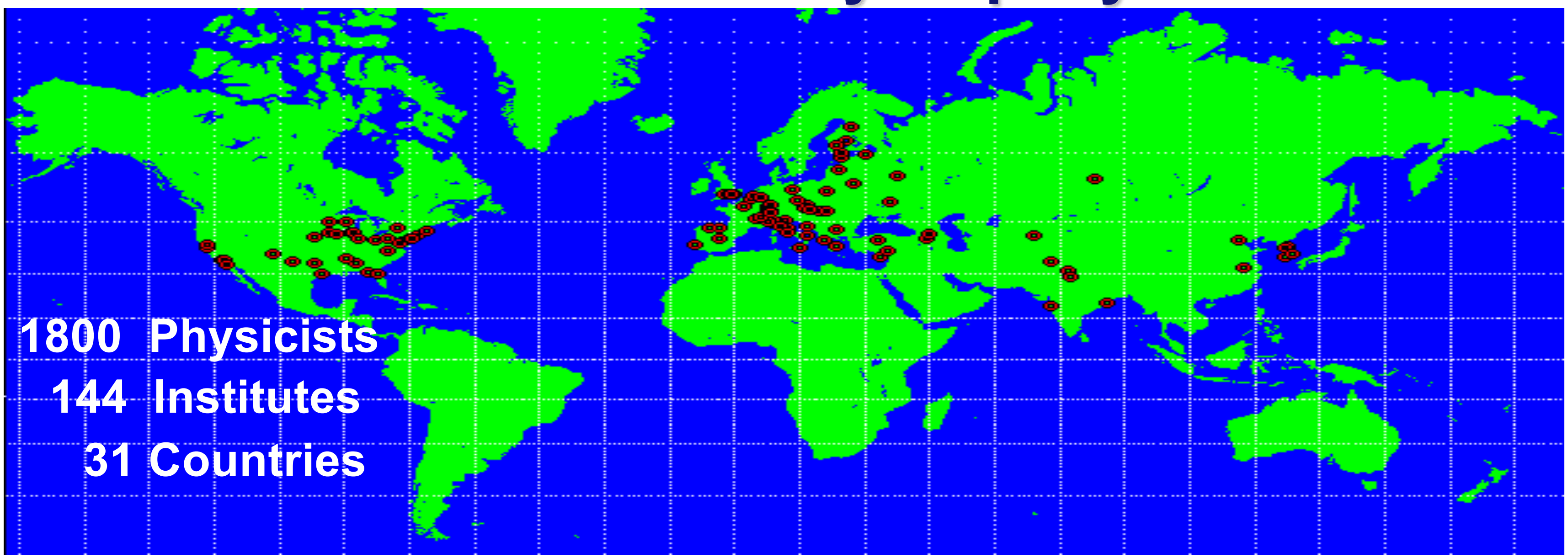
The Large Hadron Collider (2006-)

- A next-generation particle collider
  - ➡ the largest superconductor installation in the world
- A bunch-bunch collision every 25 nanoseconds: each generating ~20 interactions
  - ➡ Only one in a trillion may lead to a major physics discovery
- Real-time data filtering: Petabytes per second to Gigabytes per second
- Accumulated data of many Petabytes/Year (1 Exabyte by ~2012)

Large data samples explored and analyzed by thousands of geographically dispersed scientists, in hundreds of teams

Computing Challenges: LHC Example

- **Geographical dispersion:** of people and resources
- **Complexity:** the detector and the LHC environment
- **Scale:** Tens of Petabytes per year of data



Major challenges associated with:

Communication and collaboration at a distance

Network-distributed computing and data resources

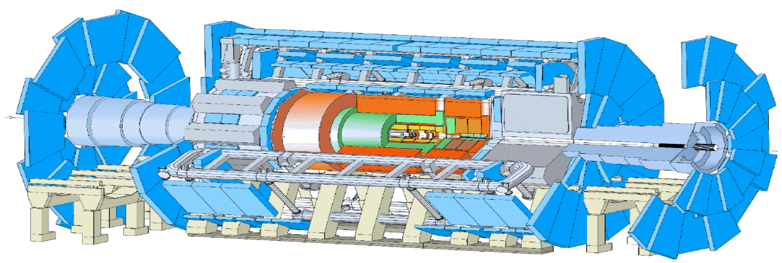
Remote software development and physics analysis

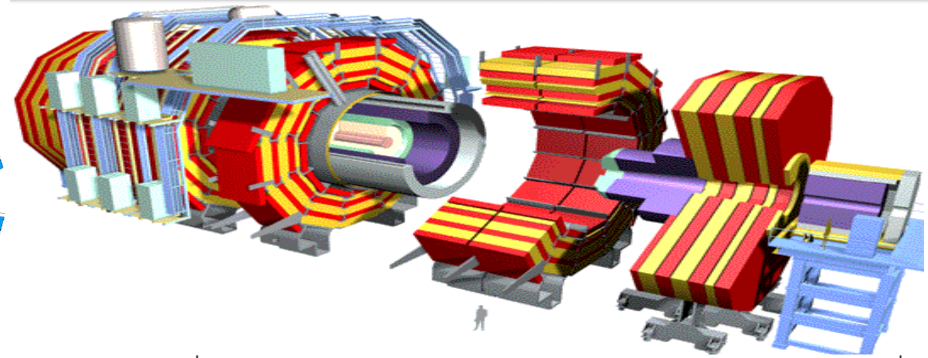
R&D: New Forms of Distributed Systems: Data Grids

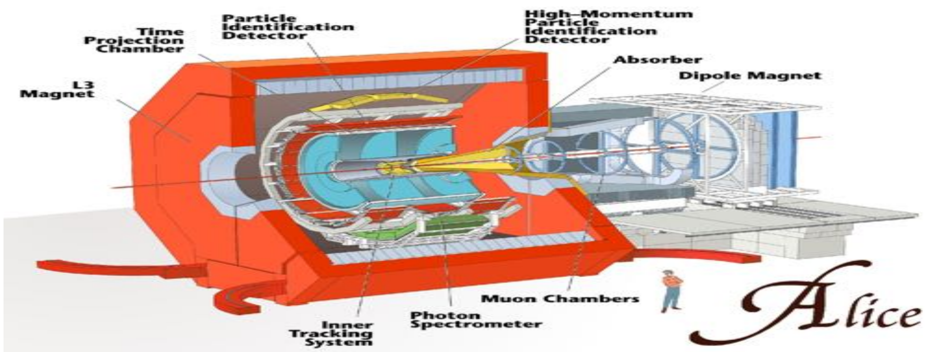
Four LHC Experiments: The Petabyte to Exabyte Challenge

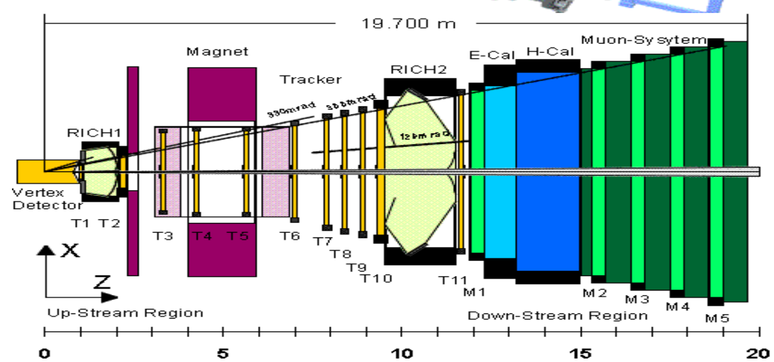
ATLAS, CMS, ALICE, LHCB

Higgs + New particles; Quark-Gluon Plasma; CP Violation



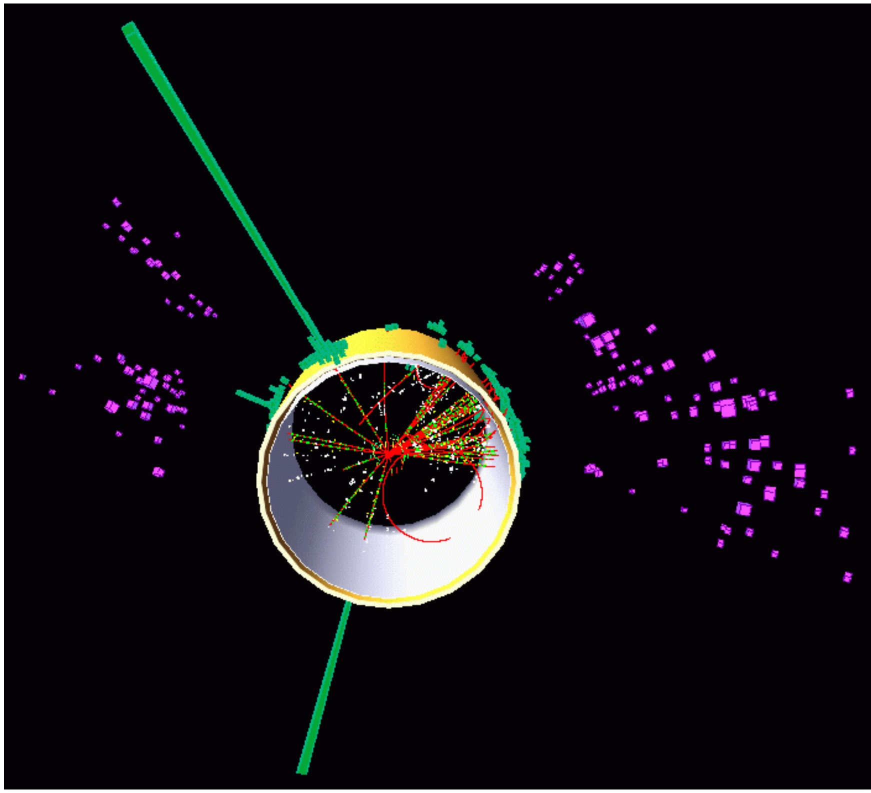




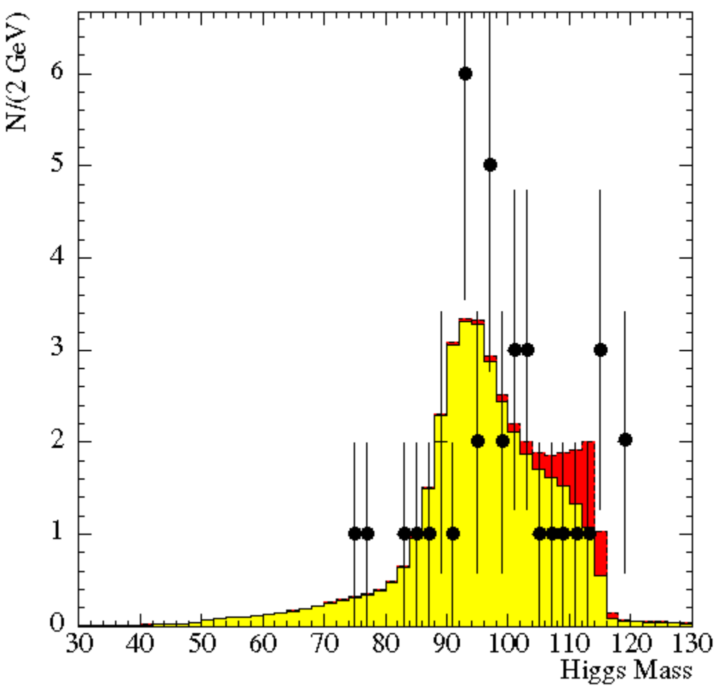


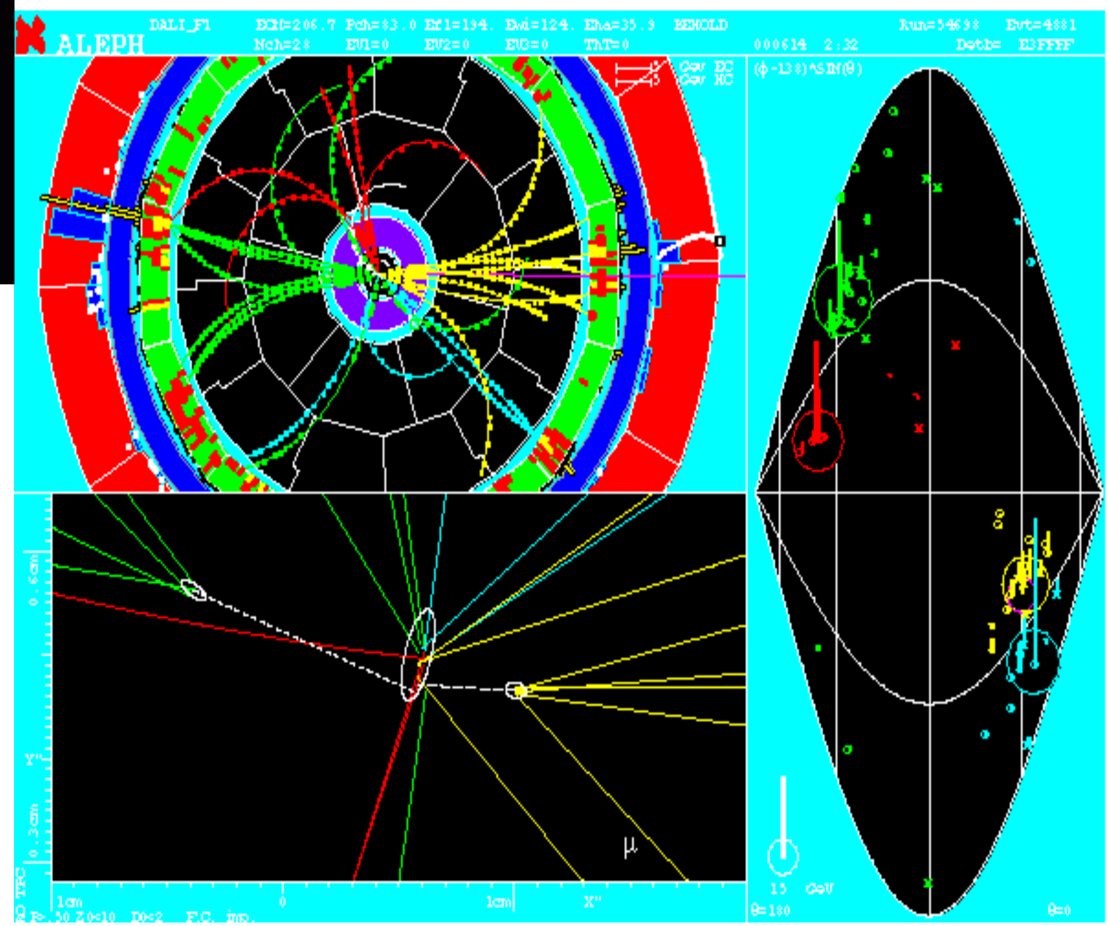
Data stored ~40 Petabytes/Year and UP;
CPU 0.30 Petaflops and UP

0.1 (2007) to 1 (~2012 ?) Exabyte (1 EB = 10^{18} Bytes) for the LHC Experiments

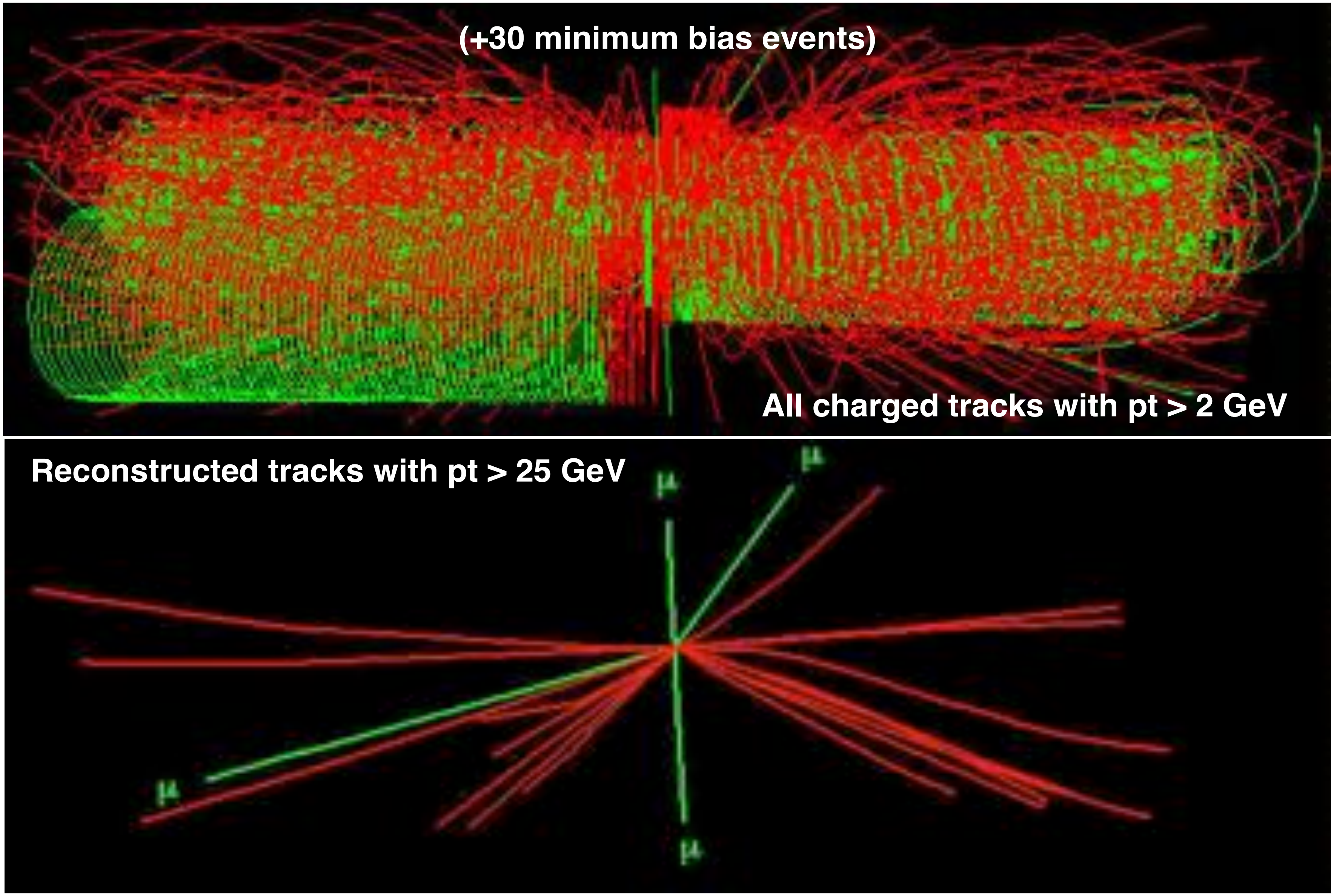


Evidence for the Higgs at LEP at M~115 GeV
The LEP Program Has Now Ended



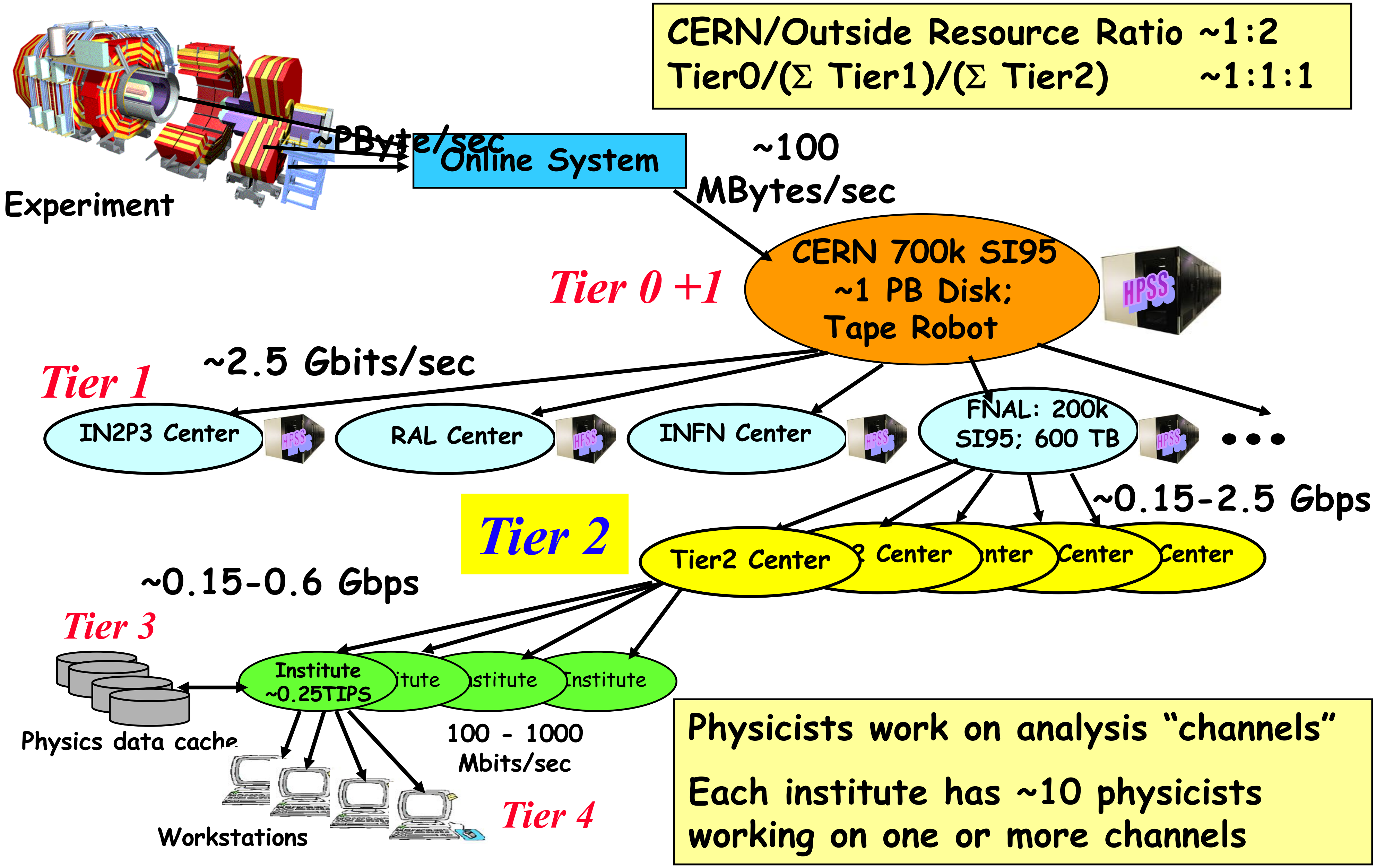


LHC: Higgs Decay into 4 muons (tracker only); 1000X LEP Data Rate



10^9 events/sec, selectivity: 1 in 10^{13} (1 person in a thousand world populations)

LHC Data Grid Hierarchy



HENP Transatlantic Net WG Bandwidth Requirements [*]

	2001	2002	2003	2004	2005	2006
CMS	100	200	300	600	800	2500
ATLAS	50	100	300	600	800	2500
BaBar	300	600	1100	1600	2300	3000
CDF	100	300	400	2000	3000	6000
D0	400	1600	2400	3200	6400	8000
BTeV	20	40	100	200	300	500
DESY	100	180	210	240	270	300
CERN BW	155-310	622	1250	2500	5000	10000

[*] Installed BW. Maximum Link Occupancy 50% Assumed

The Network Challenge is Shared by Both Next- and Present Generation Experiments

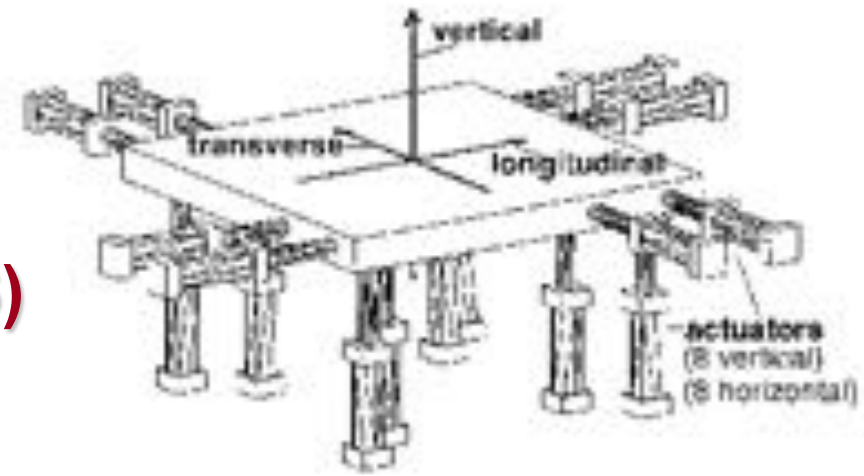
Networking Requirements: Beyond Bandwidth

Beyond the simple requirement of adequate bandwidth, physicists in DoE/DHEP's (and NSF/EPP's) major programs require:

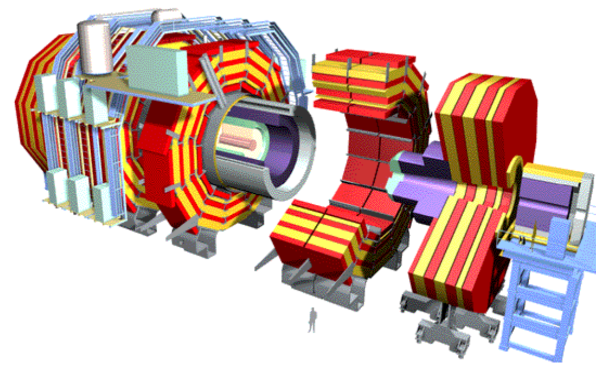
- An integrated set of local, regional, national and international networks able to interoperate seamlessly, without bottlenecks
- Network and user software that will work together to provide high throughput and manage bandwidth effectively
- A suite of videoconference and high-level tools for remote collaboration that will make data analysis from the US and from other remote sites effective

The effectiveness of participation in the LHC and other major HENP experimental programs depends on the high performance of our national and international networks

Emerging *Data Grid* User Communities



- NSF Network for Earthquake Engineering Simulation (NEES)
  - Integrated instrumentation, collaboration, simulation
- *Grid Physics Network (GriPhyN)*
  - *ATLAS, CMS, LIGO, SDSS*
- Access Grid; VRVS: supporting group-based collaboration



And

- Genomics, Proteomics, ...
- The Earth System Grid and EOSDIS
- Federating Brain Data
- Computed MicroTomography …
- Virtual Observatories

Grid Projects Timeline

Q2 00
- Submit GriPhyN proposal, $12.5M

Q3 00
- Outline of US-CMS Tier plan
- GriPhyN approved, $12M
- Caltech-UCSD install Proto-T2

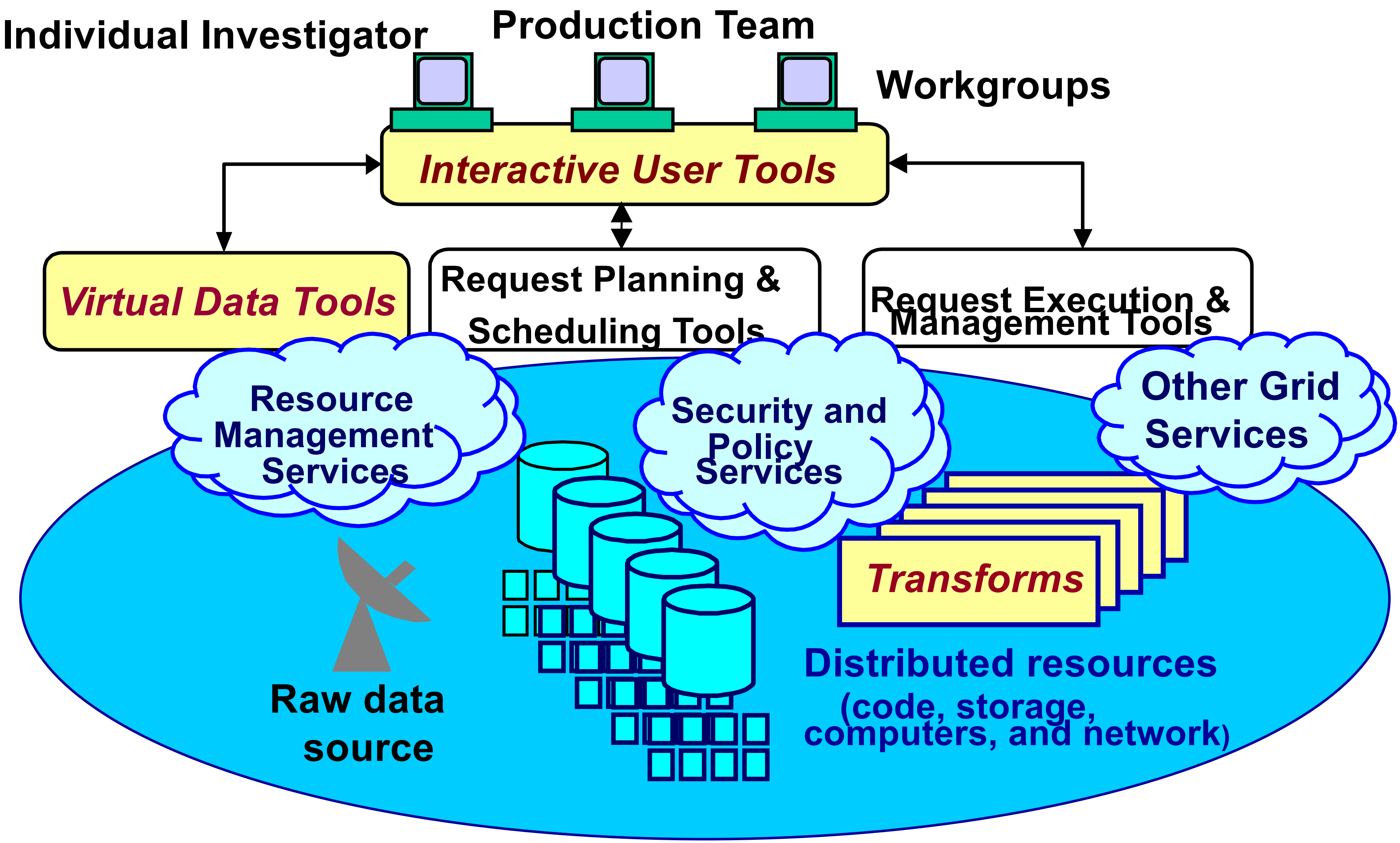
Q4 00
- Submit iVDGL preproposal
- EU DataGrid approved, $9.3M

Q1 01
- 1st Grid coordination meeting
- Submit PPDG proposal, $12M
- Submit iVDGL proposal

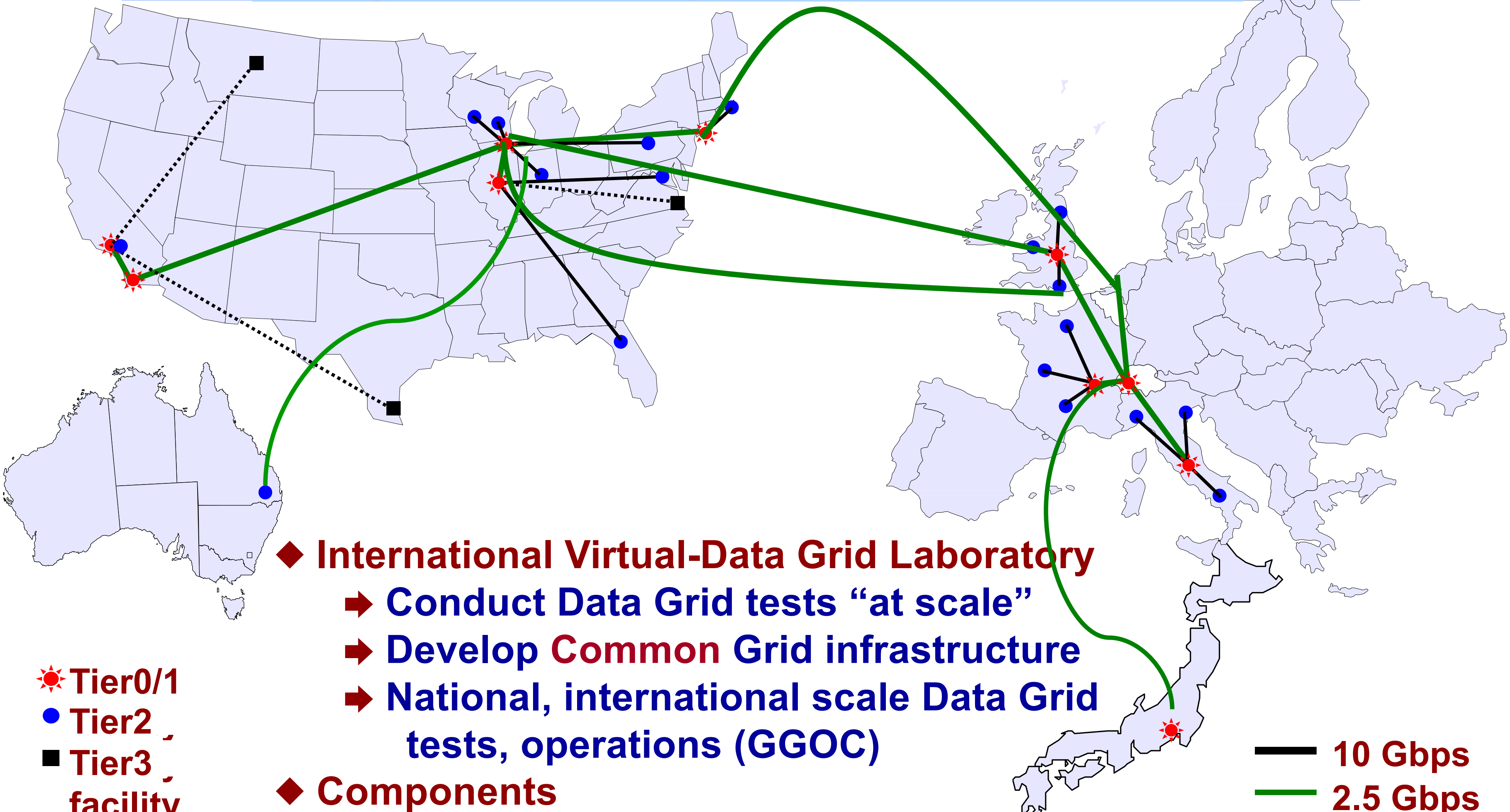
Q2 01
- Submit DTF proposal, $45M
- PPDG approved, $9.5M
- Install initial Florida proto-T2
- 2nd Grid coordination meeting

Q3 01
- iVDGL approved, $ 14 M
- DTF approved, $ 53 M

GriPhyN: PetaScale Virtual Data Grids: LHC, LIGO, SDSS



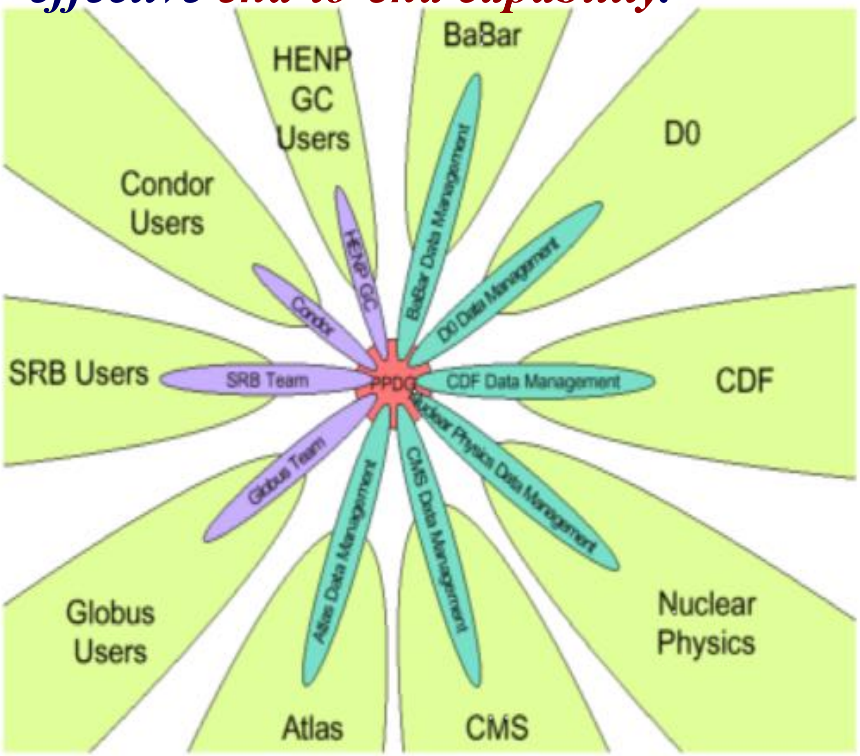
GriPhyN iVDGL 2002- US, UK, Italy, France, Japan, Australia, Brazil



- International Virtual-Data Grid Laboratory
  - Conduct Data Grid tests "at scale"
  - Develop Common Grid infrastructure
  - National, international scale Data Grid tests, operations (GGOC)
- Components
  - Tier1, Selected Tier2 and Tier3 Sites
  - Distributed Terascale Facility (DTF)
  - 0.6 - 10 Gbps networks: US, Europe, transoceanic

Particle Physics Data Grid Collaboratory Pilot (2001-2003)

"The Particle Physics Data Grid Collaboratory Pilot will develop, evaluate and deliver vitally needed Grid-enabled tools for data-intensive collaboration in particle and nuclear physics. Novel mechanisms and policies will be vertically integrated with Grid Middleware, experiment specific applications and computing resources to provide effective end-to-end capability."



Computer Science Program of Work

- **CS1: Job Description Language**
- **CS2: Schedule and Manage Data Processing and Placement Activities**
- **CS3 Monitoring and Status Reporting**
- **CS4 Storage Resource Management**
- **CS5 Reliable Replication Services**
- **CS6 File Transfer Services**
- **CS7 Collect/Document Current Exp. Practices and Potential Generalizations**

DB file/object-collection replication, caching, catalogs, End-to-end

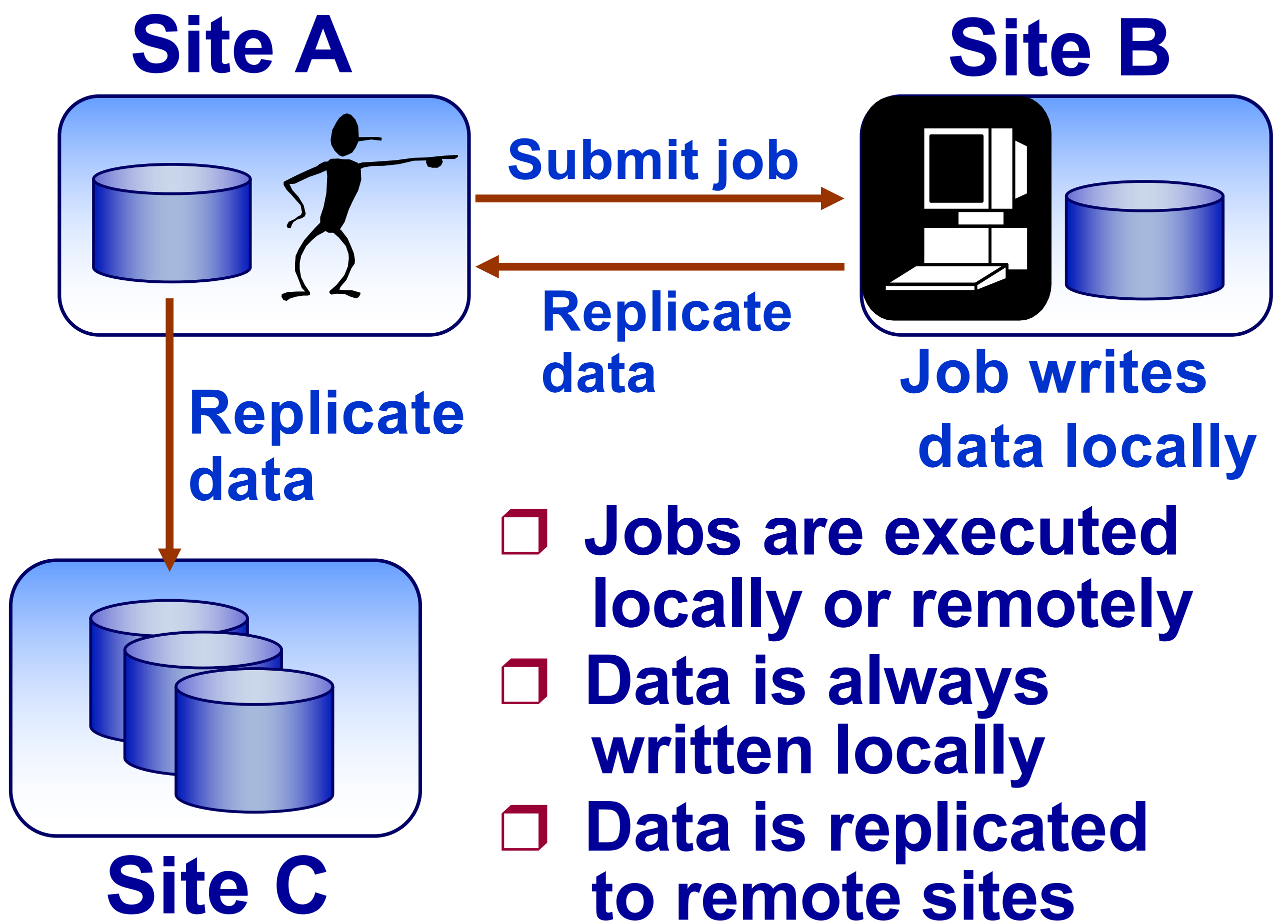
Distributed Terascale Facility (DTF)

- Alliance and NPACI Proposal to NSF; Approved August 2001: NCSA (UIUC), SDSC, Argonne, Caltech: *Linked at OC192*
  - To Deploy a DTF based on Linux clusters, large-scale data archives and high bandwidth national networks
  - Atop the DTF hardware, deploy a "TeraGrid": a new unified model of distributed data analysis, computing and communication for science
- Integration Partners: IBM, Intel, Qwest
- Four Complementary Foci
  - Computing intensive appications (NCSA)
    - 6 TF IA-64, Myrinet, > 100 TB disk, 1+ PB Tertiary
  - Data intensive applications (SDSC)
    - 4+ TF Linux Cluster, > 100 TB Disk, Multi-PB Tertiary
  - Remote rendering and visualization (Argonne)
    - Linux clusters and graphics cards serving remote imagery
  - Applications Consortia (Caltech)
- Software: Linux and vendor (IBM) cluster software; Globus, Condor and other Grid tools

Grid Data Management Prototype (GDMP)

Distributed Job Execution and Data Handling:

- Transparency
- Performance
- Security
- Fault Tolerance
- Automation



- Jobs are executed locally or remotely
- Data is always written locally
- Data is replicated to remote sites

GDMP Used In CMS Simulated Event Production Among ~10 Sites in the US, Europe and Asia 2000-2001

DTF "Seamless MicroGrid" Prototype

- Caltech/Wisconsin Condor/NCSA Production
- Simple Job Launch from Caltech
  - Authentication Using Globus Security Infrastructure (GSI)
  - Resources identified Using Globus Information Infrastructure (GIS)
- CMSIM Jobs Sent to Wisconsin Condor Flock Using Condor-G
  - Output Files Automatically Stored in NCSA Unitree (Gridftp)
- ORCA Phase: Read-in and Process Jobs at NCSA
  - Output Files Automatically Stored in NCSA Unitree
- Future: Multiple CMS Sites; Storage in Caltech HPSS Also, Using GDMP (With LBNL's HRM).

US CMS Remote Control Room For LHC


The CDF Remote Control Room

Getting in on the Action, from Afar

by Leila Belkora, Office of Public Affairs

An invisible hand draws a white circle against a black background, almost filling the screen of a computer monitor. The circle represents the outer edge of the CDF detector at Fermilab. A spray of green, blue, and red lines blooms rapidly from the center of the circle, revealing the trajectories of particles streaming from the latest collision of a proton and an antiproton in the Tevatron accelerator. One green line arcs toward the upper right of the screen, two more curl around to the lower left. Faster than you can say data visualization, the invisible hand draws a yellow box around the green line on the right, the path of the particle with the highest calculated momentum. The screen goes black again. The image of another proton-antiproton collision at CDF—an ordinary event, as these collisions go, or a rare one that will send up a flag to physicists on the experiment—is due on the screen in less than ten seconds.

continued on page 8

Inside

US CMS is using the CDF/KEK remote control room for Fermilab Run II as a starting point. However, we want to (1) expand the scope to encompass a US based physics group and US LHC accelerator tasks, and (2) extend the concept to a Global Collaboratory for realtime data acquisition + analysis

<u>VRVS</u>

6000 Hosts

In 52 Countries

Annual Growth 250%

HENP Networking WG [*] Mission

- To help ensure that the required
  - National and international network infrastructures
  - Standardized tools and facilities for high performance and end-to-end monitoring and tracking, and
  - Collaborative systems
- are developed and deployed in a timely manner, and used effectively to meet the needs of the US LHC and other major HENP Programs, as well as the general needs of our scientific community.
- To carry out these developments in a way that is broadly applicable across many fields, within and beyond the scientific community

[*] Co-Chairs: S. McKee (Michigan), H. Newman

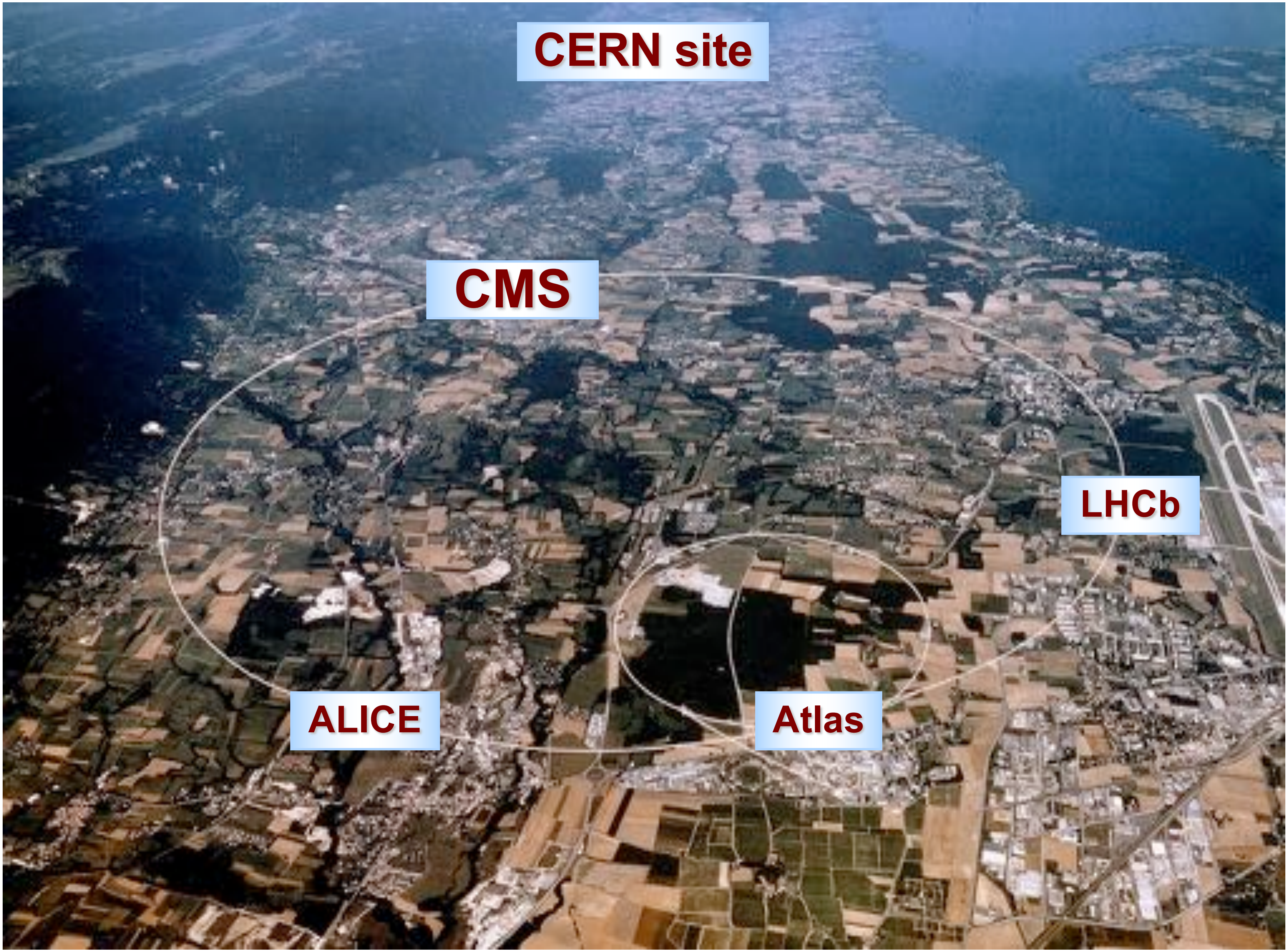
HENP Network WG Current and Next Steps

- **Internet2 HENP Network Working Group**
  - WG Draft Charter Written and sent to Internet2 Management (HN; August 4); see http://l3www.cern.ch/~newman/Internet2/ HENPWGCharter_V0.5hbn040801.doc
  - WG welcomed by the Internet2 Applications Director (T. Hanss) and the Applications Strategy Council Chair (T. DeFanti/UIC)
  - WG BOF meeting at the Internet2 Fall Meeting in Austin [*], October 1
  - WG Charter will be presented, reviewed by the Applications Strategy Council
- **Develop Liaisons**
  - AMPATH
  - ESNet Committees (ESCC and ESSC)
  - Global Grid Forum: A Networking BOF ?
  - Joint Techs Workshop presentations
- **Coordinate with I2 E2E Initiative; Network activities in PPDG, iVDGL, DataTAG in Europe, and Start Work**
  - [*] See http://www.internet2.edu/activities/html/fall01.html

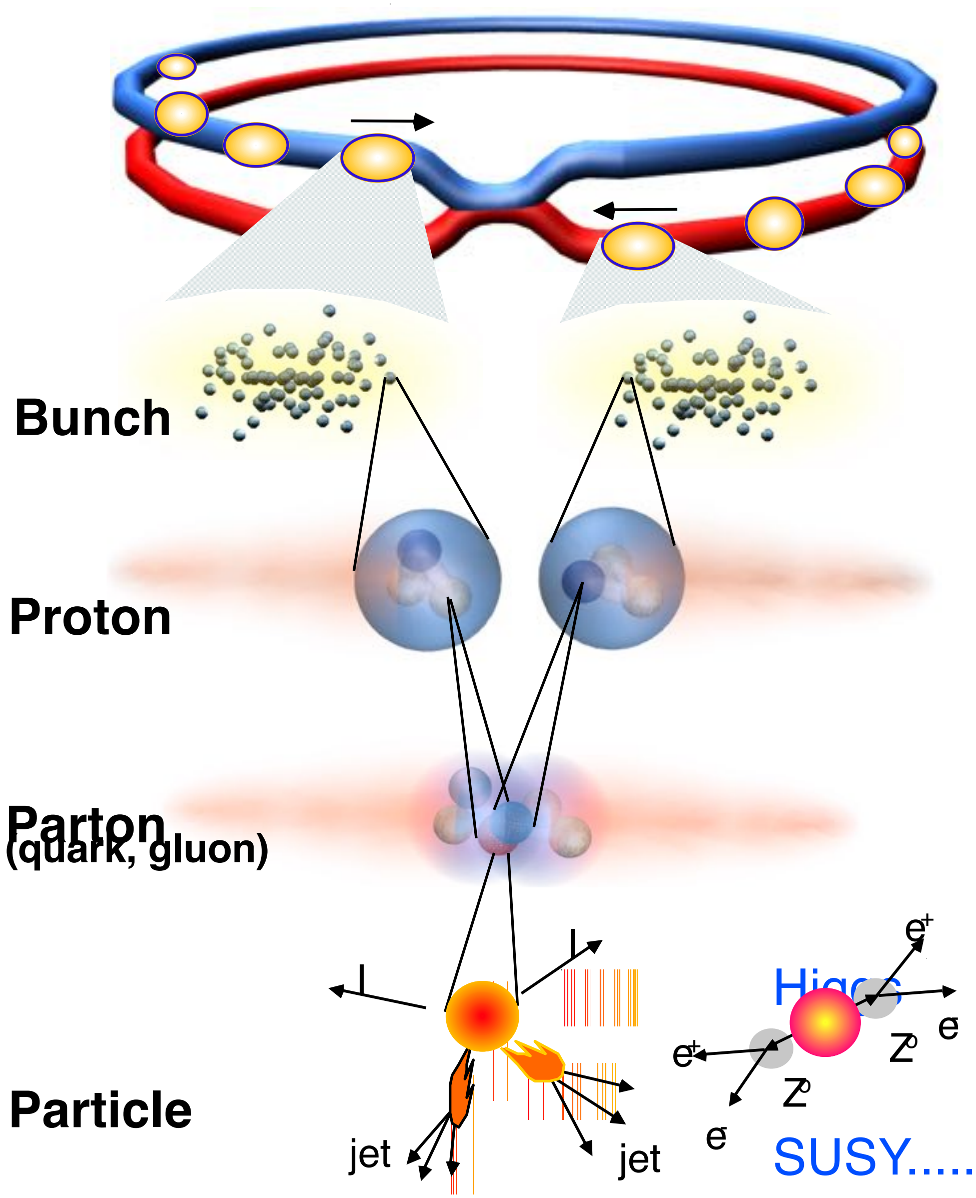
HENP DataGrids and Networks Summary

- In an era of worldwide collaboration, seamless high-performance networks are crucial to the major HENP and other programs
  - And for Grid Developments
- We hope that advances in technology will bring the Gbps networks we require financially within our reach
  - Key technical developments for scalable, high throughput, responsiveness and reliability also are required
- Since Grids are beginning to transform the way we do science
  - Future network requirements may evolve as fast as network technology
- HENP has recently realized that networking must be planned as a large-scale high priority task of major collaborations
  - We must work to ensure that the necessary tools, as well as the global network infrastructures themselves, will be there in time
- The HENP Network WG could have a key role in these developments; in coordination with PPDG and the iVDGL GGOC

Some Extra Slides Follow


CERN site
CMS
LHCb
ALICE
Atlas

Collisions at LHC



Proton- Proton 2835 bunch/beam
Protons/bunch 10^{11}
Beam energy 7 TeV ($7x10^{12}$eV)
Luminosity $10^{34} cm^{-2}s^{-1}$

Crossing rate 40 MHz

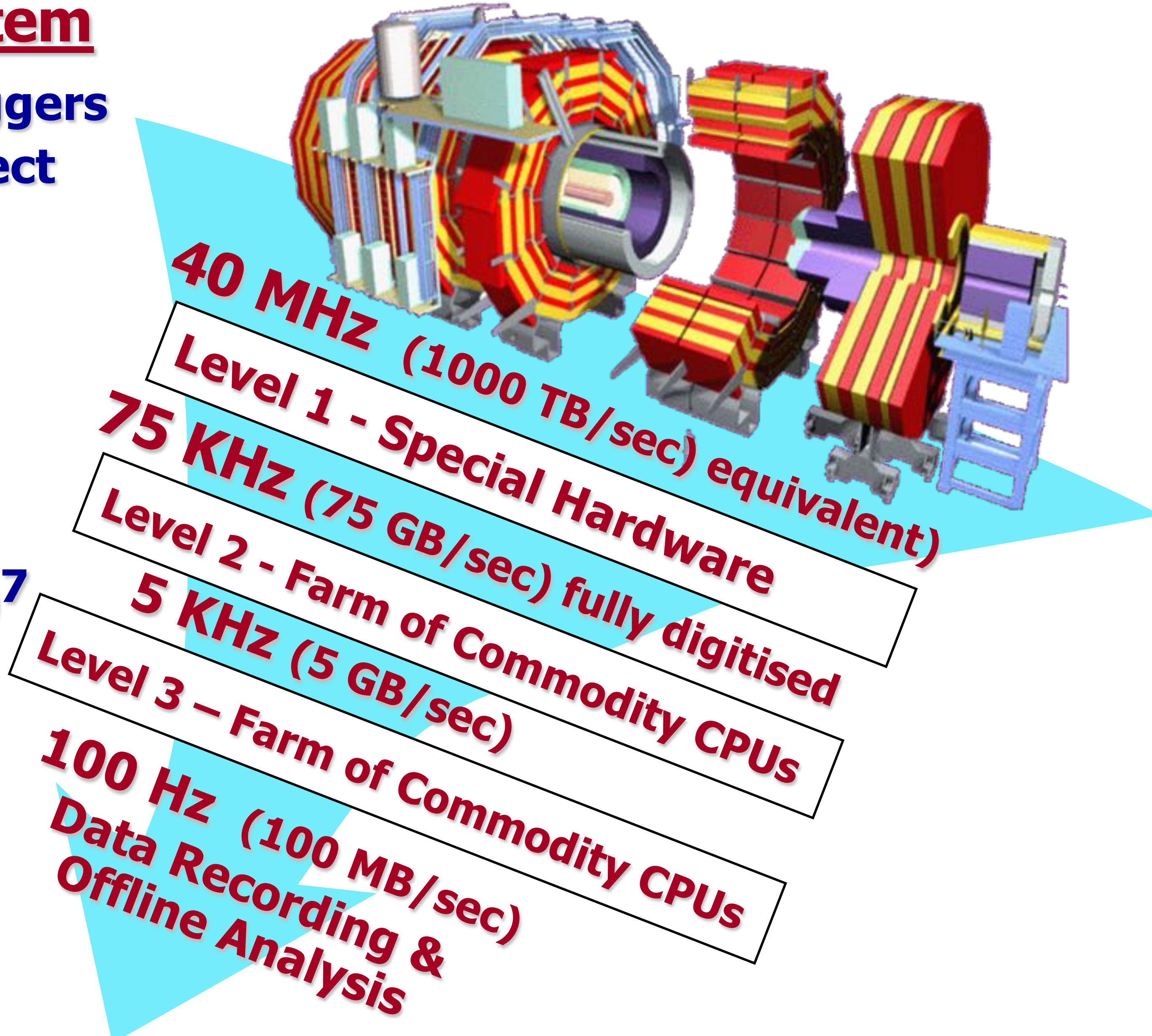
Collision rate $\sim10^9$ Hz
(Average ~20 Collisions/Crossing)

New physics rate ~ 0.00001 Hz

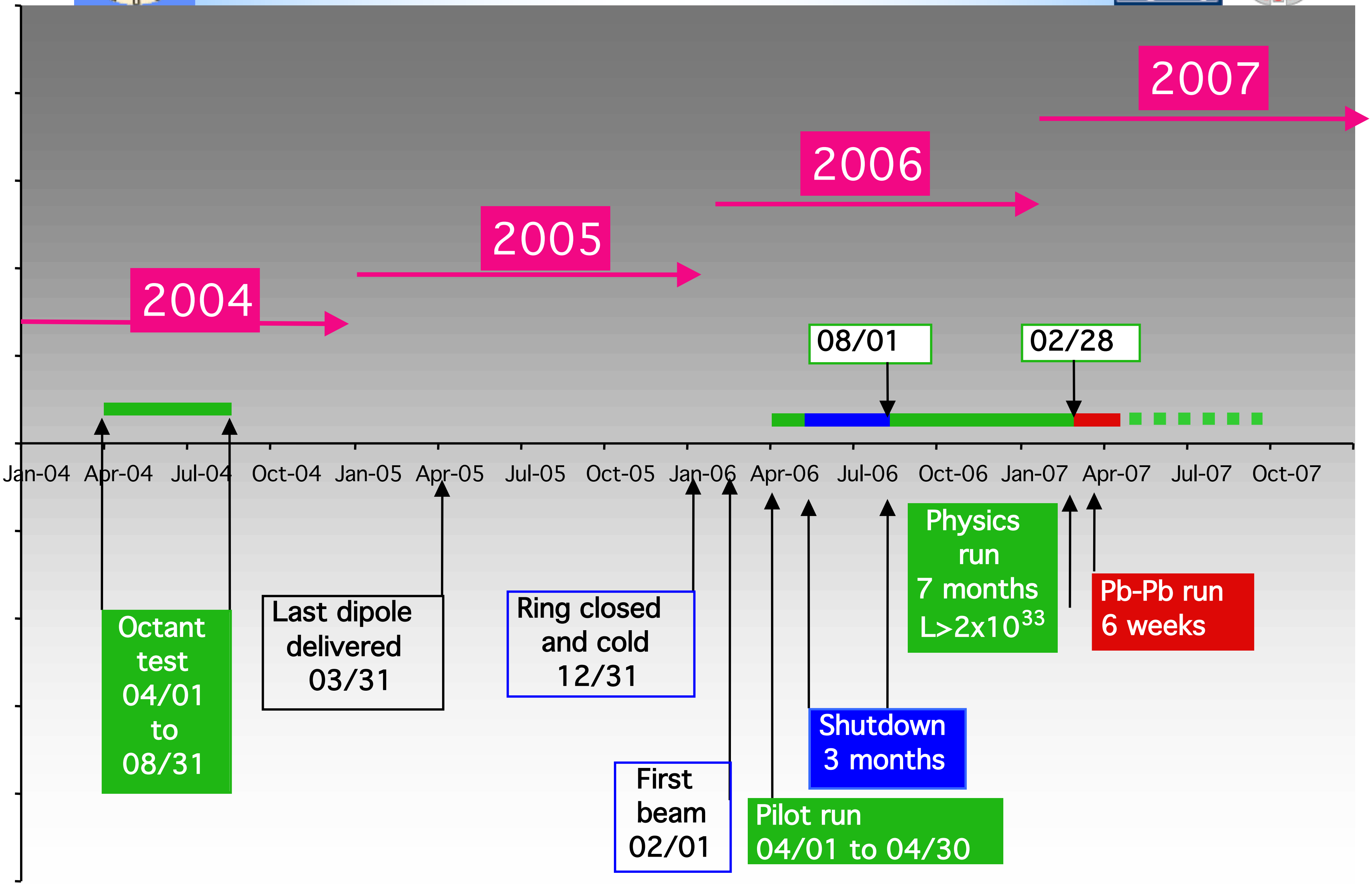
Event selection:
1 in 10,000,000,000,000

On-line Filter System

- Large variety of triggers and thresholds: select physics à la carte
- Multi-level trigger
- Filter out less interesting events
- Online reduction 10^7
- Keep highly selected events
- Result: Petabytes of Binary Compact Data Per Year



LHC Commissioning Schedule



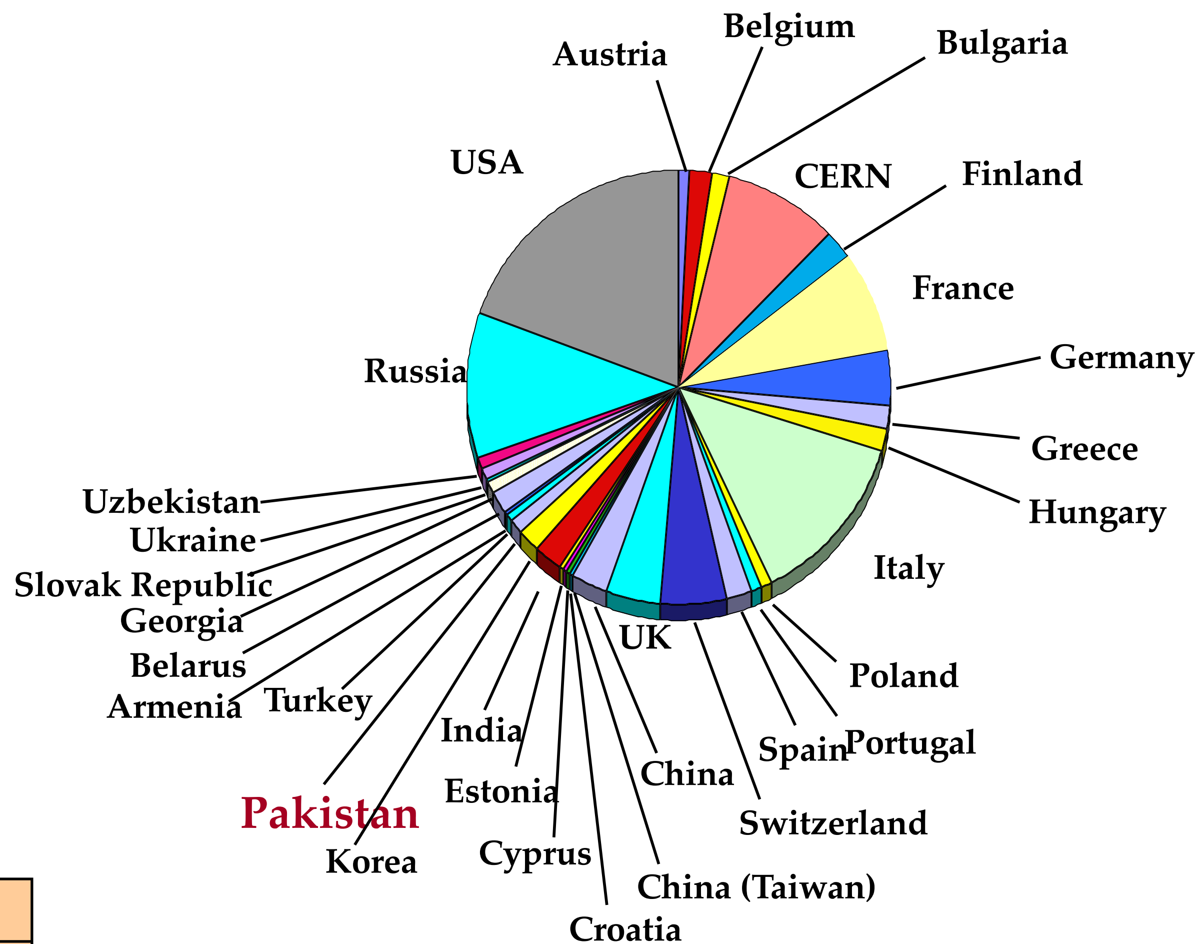
The CMS Collaboration

	Number of Laboratories
Member States	58
Non-Member States	50
USA	36
Total	144

	Number of Scientists
Member States	1010
Non-Member States	448
USA	351
Total	1809

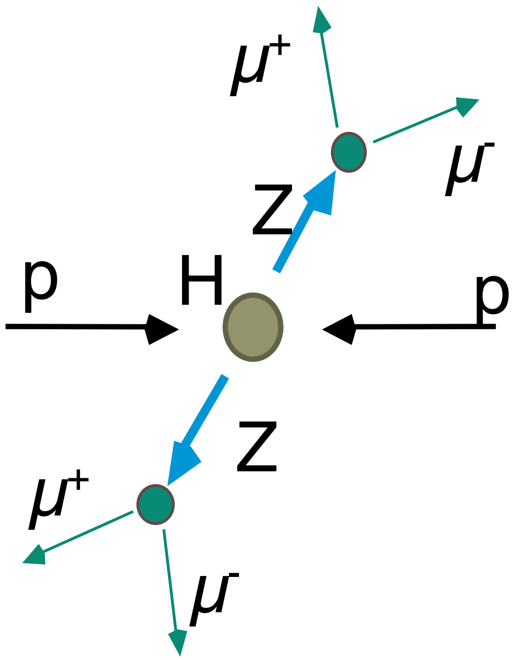
Associated Institutes	
Number of Scientists	36
Number of Laboratories	5

Oct, 17th, 2000/sm
http://cmsdoc.cern.ch/pictures/cmsorg/overview.html



1809 Physicists and Engineers
31 Countries
144 Institutions

Higgs Particles: The Next Step At the High Energy Frontier



Higgs Sector

- Clarify the Origin of Spontaneous Symmetry Breaking
- The Standard Model, *OR* Supersymmetry

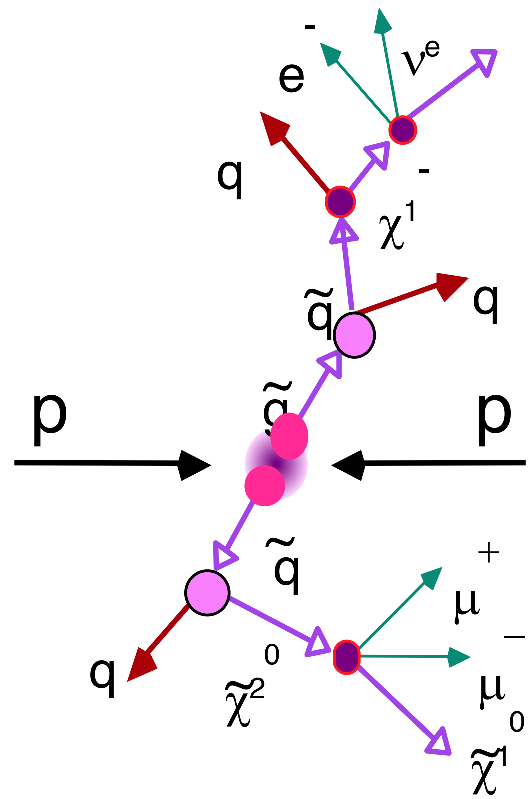
➡ New forces (symmetries)

➡ New particles

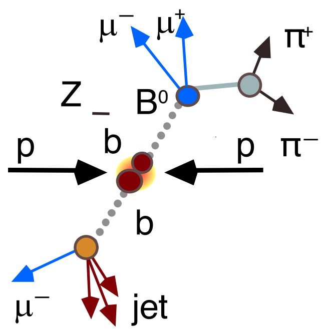
➡ Supersymmetries

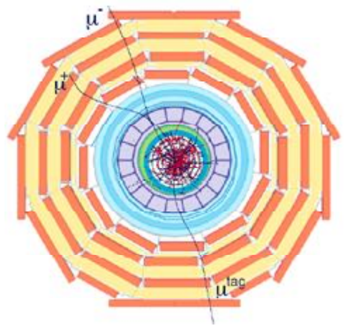
➡ Substructure

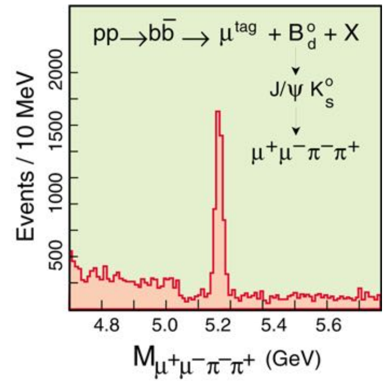
➡ ....



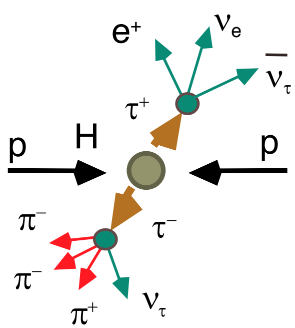
CMS physics: B & Supersymmetry

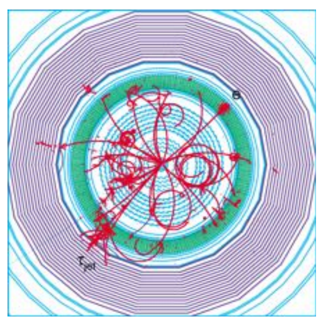




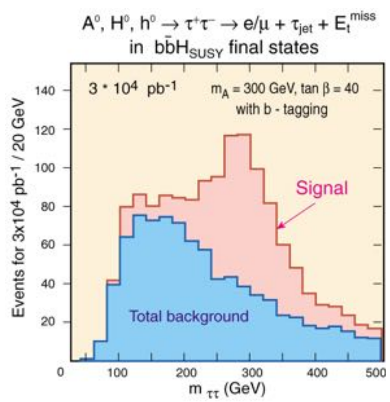


CP Violation: B Decays

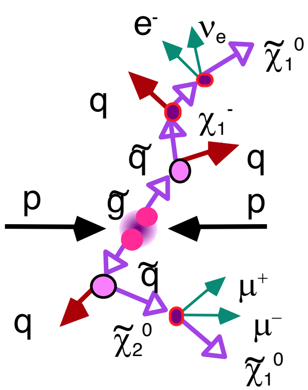


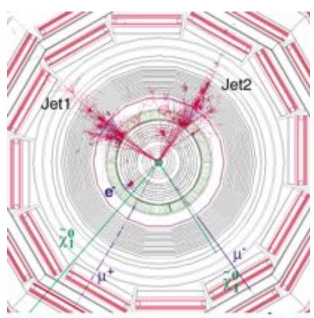


H → τ → e + τ_jet("3-prong")

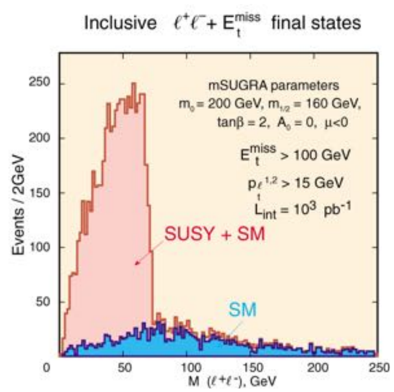


Supersymmetric Higgs Bosons





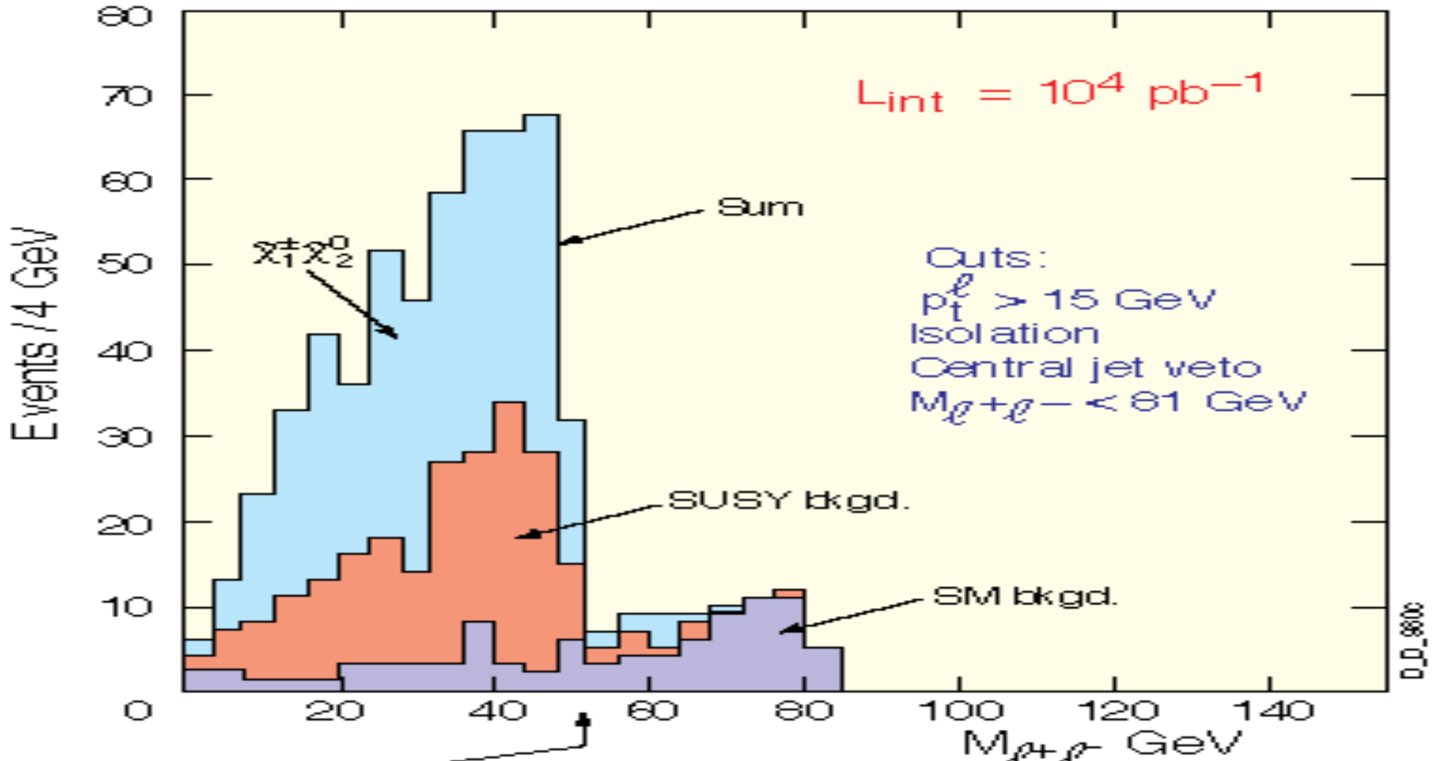
SUSY event with
3 leptons + 2 Jets signature



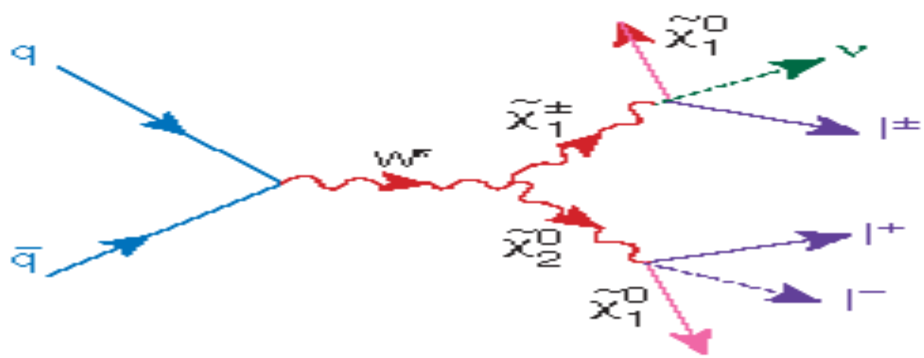
Supersymmetic Particles: Squarks, Gluinos Charginos, Neutralinos

$\tilde{\chi}_1^0$ mass determination in $3\ell^\pm$ + no jets + E_t^{miss} final state from $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production

$m_0 = 200$ GeV, $m_{1/2} = 100$ GeV, $\tan\beta = 2$, $A_0 = 0$, $\mu < 0$
$M(\tilde{\chi}_2^0) - M(\tilde{\chi}_1^0) \approx M(\tilde{\chi}_1^0) \approx 52$ GeV

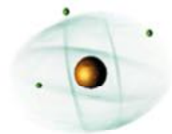
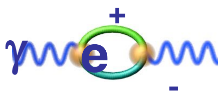
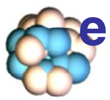
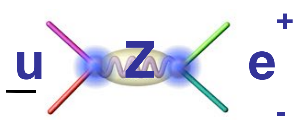


$M_{\ell\ell}^{max} = M(\tilde{\chi}_2^0) - M(\tilde{\chi}_1^0) \approx M(\tilde{\chi}_1^0) \rightarrow m_{1/2} \approx 2.5\, M(\tilde{\chi}_1^0)$
in m SUGRA



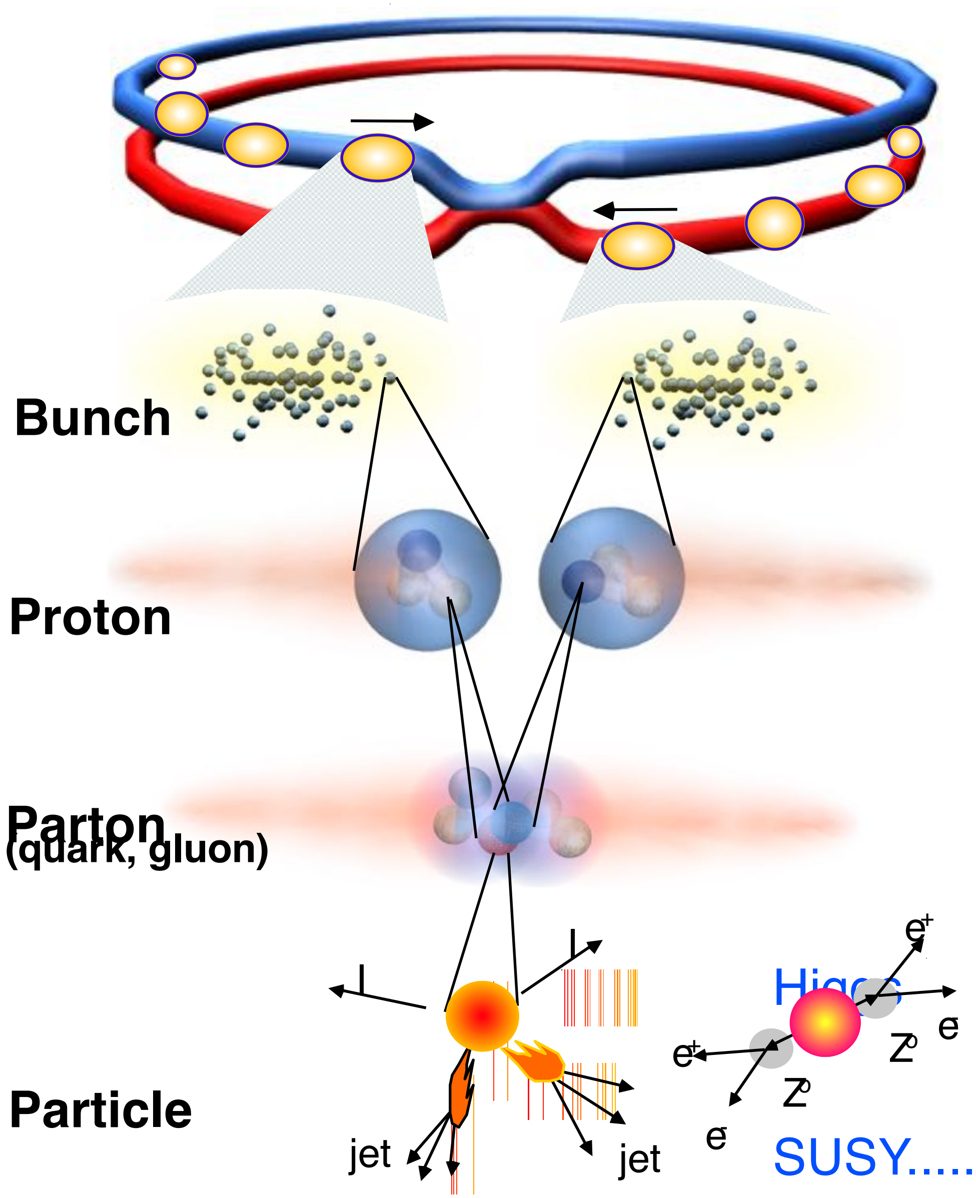
Discovery of Supersymmetry could occur in the 2006 Pilot Run. In depth study could then begin during the first Physics Run

HEP Short History and New Frontiers

	$\lambda = h/p$		Temp ~ 1/Sqrt(t)		
	10^{-10} m	~ 10 eV	>300000 Y	1900....	Quantum Mechanics Atomic Physics
				1940-50	Quantum Electro Dynamics
	10^{-15} m	MeV - GeV	~ 3 min	1950-65	Nuclei, Hadrons Symmetries, Field theories
	10^{-16} m	>> GeV	~10^{-6} sec	1965-75	Quarks. Gauge theories
	10^{-18} m	~ 100 GeV	~10^{-10} sec	1970-83 SPS	ElectroWeak Unification, QCD
				1990 LEP	3 families, Precision Electroweak
				1994 Tevatron	Top quark
Origin of masses	10^{-19} m	~10^{2} GeV	~10^{-12} sec	2006 LHC	Higgs ? Supersymmetry ?
The next step... Proton Decay ?	10^{-32} m	~10^{16} GeV	~10^{-32} sec	Underground Labs	GRAND Unified Theories ?
The Origin of the Universe	10^{-35} m	~10^{19} GeV (Planck scale)	~10^{-43} sec	??	Quantum Gravity? Superstrings ?

6 Leptons	ν_e e	ν_μ μ	ν_τ τ
6 Quarks	u d	c s	t b
3 "Colors" each quark	R	G	B

Collisions at LHC



Proton- Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34}\, cm^{-2} s^{-1}$
Crossing rate	40 MHz
Collision rate	$\sim10^{9}$ Hz

(Average ~20 Collisions/Crossing)

New physics rate ~ 0.00001 Hz

Event selection:
1 in 10,000,000,000,000

Tier0-Tier1 Link Requirements Estimate: for Hoffmann Report 2001

1) Tier1 ⇆ Tier0 Data Flow for Analysis	0.5 - 1.0 Gbps
2) Tier2 ⇆ Tier0 Data Flow for Analysis	0.2 - 0.5 Gbps
3) Interactive Collaborative Sessions (30 Peak)	0.1 - 0.3 Gbps
4) Remote Interactive Sessions (30 Flows Peak)	0.1 - 0.2 Gbps
5) Individual (Tier3 or Tier4) data transfers Limit to 10 Flows of 5 Mbytes/sec each	0.8 Gbps
TOTAL Per Tier0 - Tier1 Link	1.7 - 2.8 Gbps

NOTE:

- Adopted by the LHC Experiments; given in the Steering Committee Report on LHC Computing: "1.5 - 3 Gbps per experiment"
- Corresponds to ~10 Gbps Baseline BW Installed on US-CERN Link
- Report also discussed the effects of higher bandwidths
  - For example all-optical 10 Gbps Ethernet + WAN by 2002-3

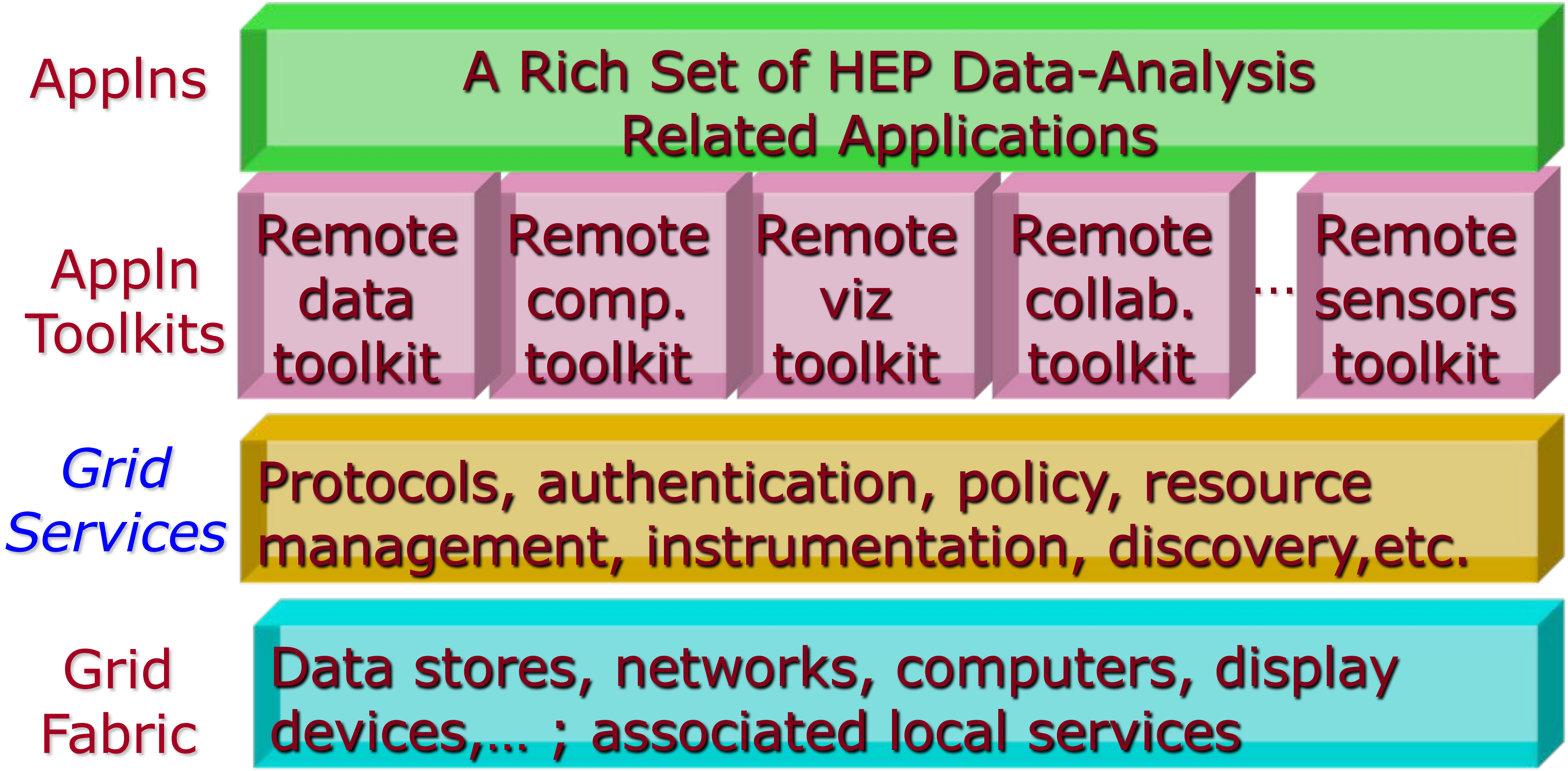
Key Network Issues & Challenges

Requirements for High Throughput

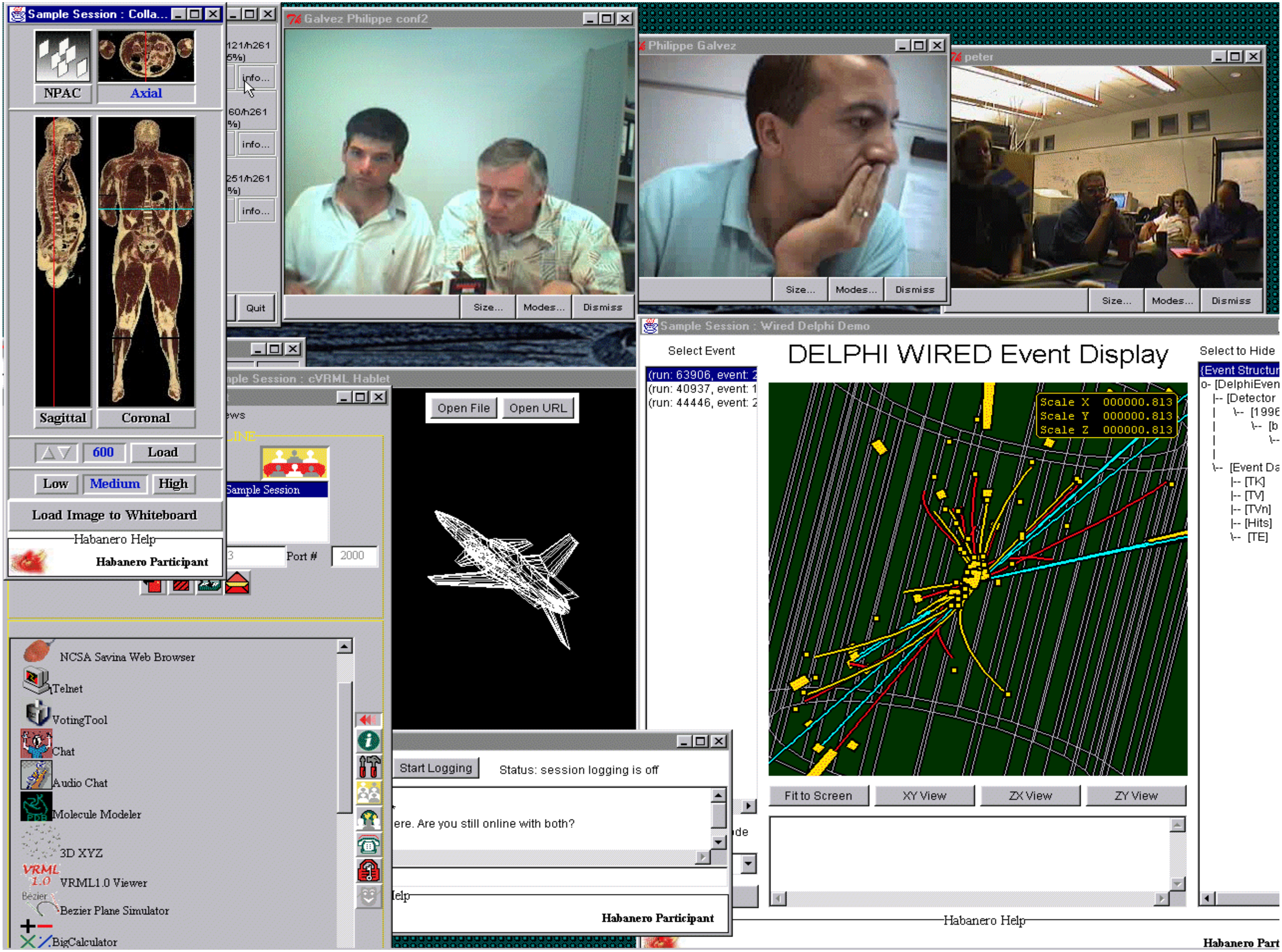
- Careful Router configuration; route stability
- Enough Router "Horsepower" (CPU, Buffer handling)
- Server and PC CPU, I/O and NIC throughput must be sufficient
- Packet Loss must be ~Zero (below 0.1%); I.e. no "commodity" networks
- TCP/IP tuning (e.g. large windows) is Absolutely Required
- End-to-end monitoring and tracking is Required
- No Local infrastructure bottlenecks can be tolerated
  - Gigabit Ethernet "clear path" between selected host pairs
  - To 10 Gbps Ethernet by ~2003
- None of this scales from 0.1 to 10 Gbps
  - New (expensive) hardware
  - New TCP/IP developments required for multiuser Gbps links
- Close collaboration with local and "regional" network engineering staffs (e.g. router and switch configuration).

✑ US National Network Infrastructure Beyond 2003

Grid Services Architecture [*]



[*] There are Computational Grids, Data Grids, Access Grids, ...


Sample Session : Colla...
NPAC
Axial
Sagittal
Coronal
600
Load
Low
Medium
High
Load Image to Whiteboard
Habanero Help
Habanero Participant
Galvez Philippe conf2
Philippe Galvez
peter
Size...
Modes...
Dismiss
Quit
Sample Session : Wired Delphi Demo
Select Event
DELPHI WIRED Event Display
Select to Hide
(run: 63906, event:
(run: 40937, event:
(run: 44446, event:
Scale X 000000.813
Scale Y 000000.813
Scale Z 000000.813
Fit to Screen
XY View
ZX View
ZY View
Open File
Open URL
Sample Session
Port #
2000
Start Logging
Status: session logging is off
ere. Are you still online with both?
NCSA Savina Web Browser
Telnet
VotingTool
Chat
Audio Chat
Molecule Modeler
3D XYZ
VRML1.0 Viewer
Bezier Plane Simulator
BigCalculator

SPARC: Real Time Data