The GriPhyN Project

(Grid Physics Network)

GriPhyN

Data Intensive Science

Paul Avery
University of Florida
http://www.phys.ufl.edu/~avery/
avery@phys.ufl.edu

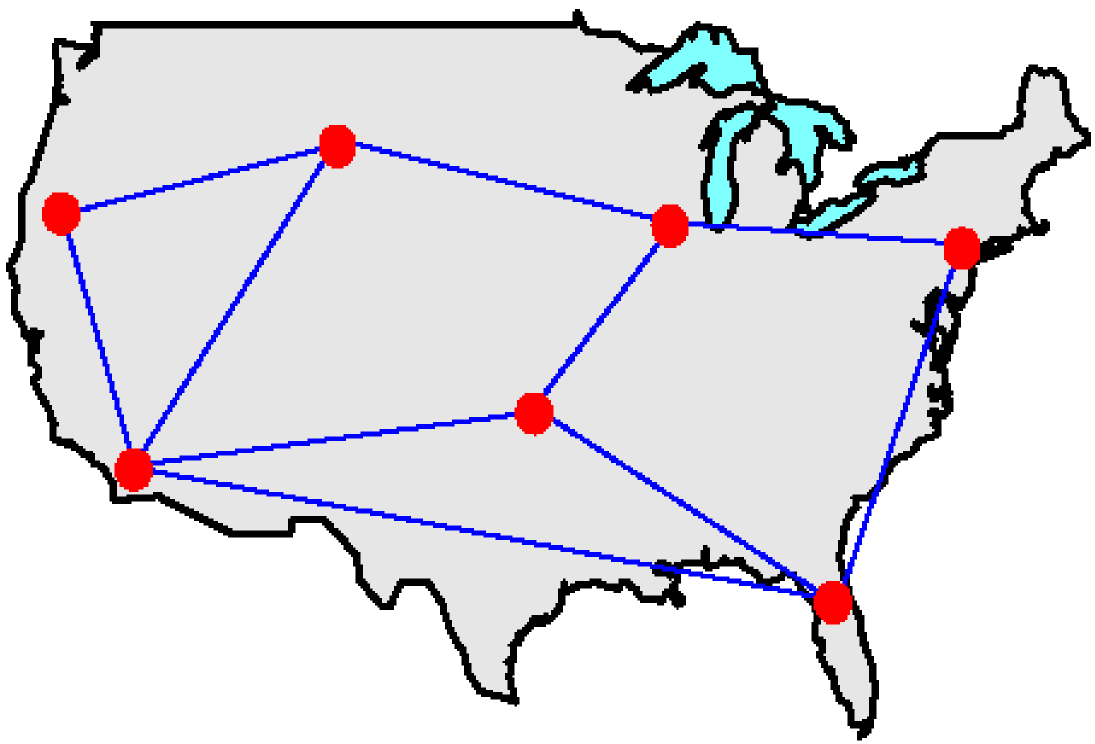
SCOP @ FNAL
Oct. 24, 2000

Motivation: Data Intensive Science

- **Scientific discovery increasingly driven by IT**
  - Computationally intensive analyses
  - Massive data collections
  - Rapid access to large subsets
  - Data distributed across networks of varying capability
- **Dominant factor: data growth**
  - ~0.5 Petabyte in 2000 (BaBar)
  - ~10 Petabytes by 2005
  - ~100 Petabytes by 2010
  - ~1000 Petabytes by 2015?
- **Robust IT infrastructure essential for science**
  - Provide rapid turnaround
  - Coordinate, manage the limited computing, data handling and network resources effectively

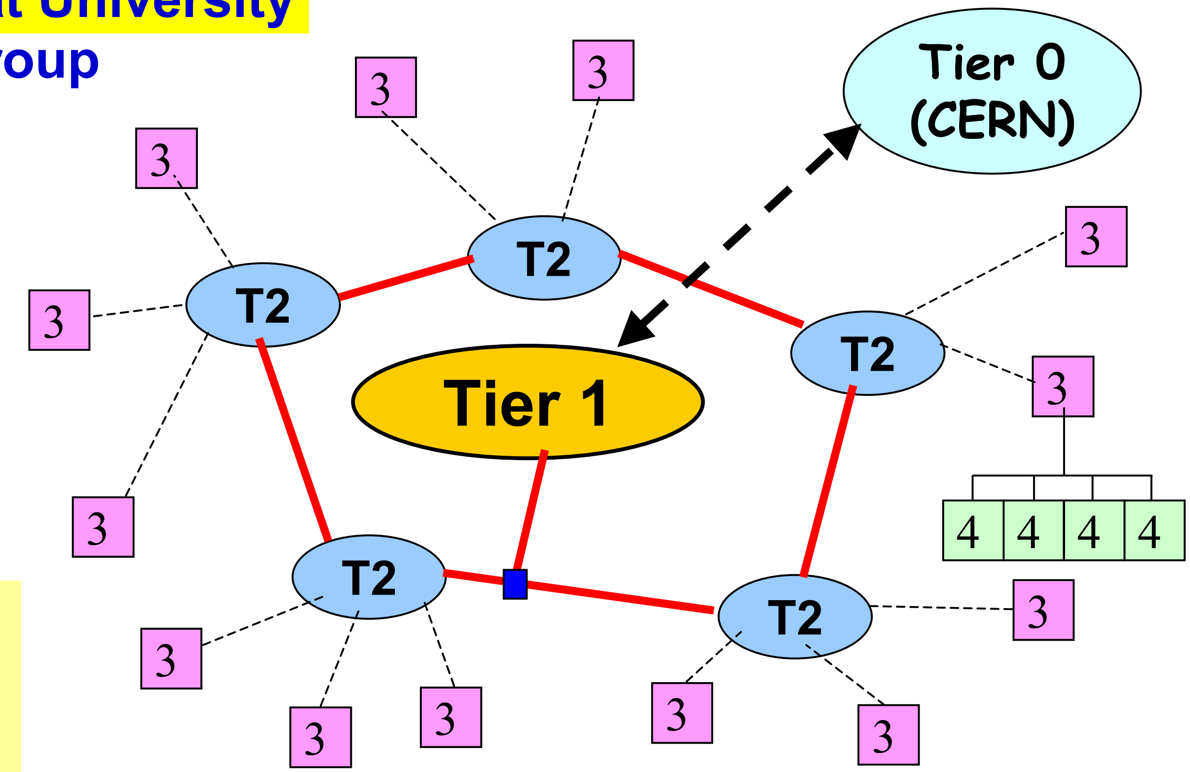
Grids as IT Infrastructure

- **Grid: Geographically distributed IT resources configured to allow coordinated use**
- **Physical resources & networks provide raw capability**
- **"Middleware" services tie it together**



Data Grid Hierarchy (LHC Example)

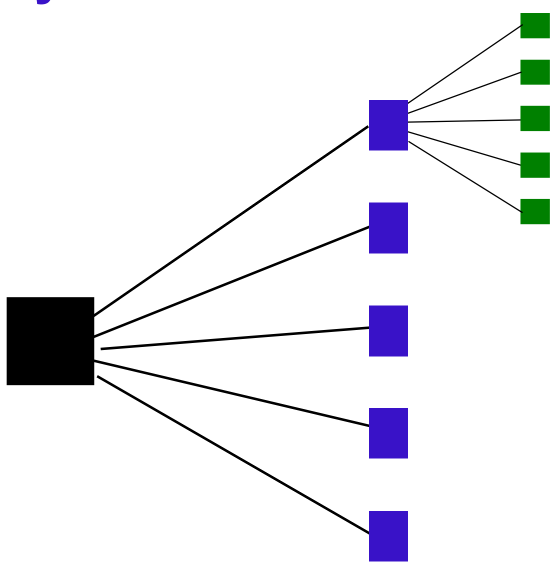
Tier0 CERN
Tier1 National Lab
Tier2 Regional Center at University
Tier3 University workgroup
Tier4 Workstation



GriPhyN:
- R&D
- Tier2 centers
- Unify all IT resources

Why a Data Grid: Physical

- Unified system: all computing resources part of grid
  - Efficient resource use (manage scarcity)
  - Resource discovery / scheduling / coordination truly possible
  - "The whole is greater than the sum of its parts"
- Optimal data distribution and proximity
  - Labs are close to the (raw) data they need
  - Users are close to the (subset) data they need
  - Minimize bottlenecks
- Efficient network use
  - local > regional > national > oceanic
  - No choke points
- Scalable growth



Why a Data Grid: Demographic

- Central lab cannot manage / help 1000s of users
  - Easier to leverage resources, maintain control, assert priorities at regional / local level
- Cleanly separates functionality
  - Different resource types in different Tiers
  - Organization vs. flexibility
  - Funding complementarity (NSF vs DOE), targeted initiatives
- New IT resources can be added “naturally”
  - Matching resources at Tier 2 universities
  - Larger institutes can join, bringing their own resources
  - Tap into new resources opened by IT “revolution”
- Broaden community of scientists and students
  - Training and education
  - Vitality of field depends on University / Lab partnership

GriPhyN = Applications + CS + Grids

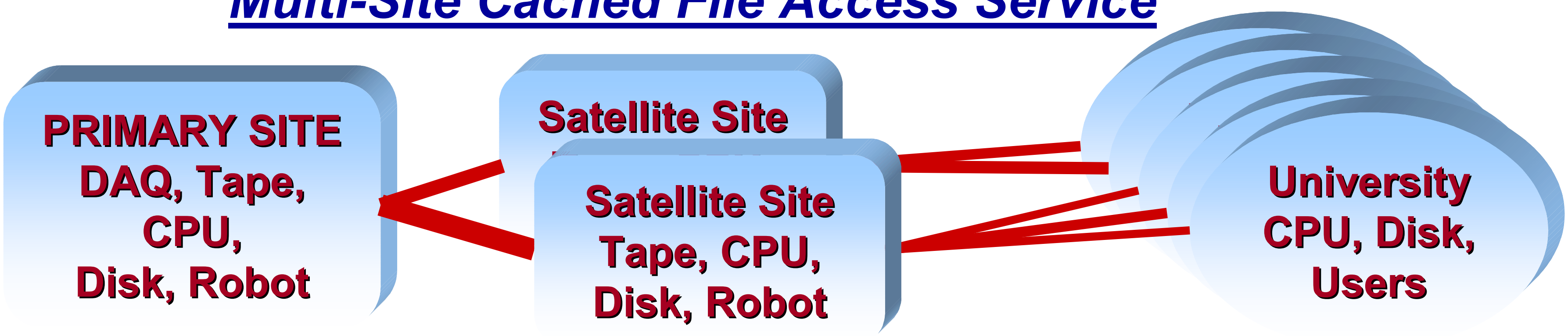
- Several scientific disciplines
  - US-CMS — High Energy Physics
  - US-ATLAS — High Energy Physics
  - LIGO/LSC — Gravity wave research
  - SDSS — Sloan Digital Sky Survey
- Strong partnership with computer scientists
- Design and implement production-scale grids
  - Maximize effectiveness of large, disparate resources
  - Develop common infrastructure, tools and services
  - Build on foundations ⇒ PPDG, MONARC, CONDOR, …
  - Integrate and extend existing facilities
- ≈ \$70M total cost ⇒ NSF(?)
  - \$12M — R&D
  - \$39M — Tier 2 center hardware, personnel, operations
  - \$19M? — Networking

Particle Physics Data Grid (PPDG)

ANL, BNL, Caltech, FNAL, JLAB, LBNL, SDSC, SLAC, U.Wisc/CS

- First Round Goal: Optimized cached read access to 10-100 Gbytes drawn from a total data set of 0.1 to ~1 Petabyte
- Matchmaking, Co-Scheduling: SRB, Condor, Globus services; HRM, NWS

Multi-Site Cached File Access Service



GriPhyN R&D Funded!

- **NSF/ITR results announced Sep. 13**
  - $11.9M from Information Technology Research Program
  - $ 1.4M in matching from universities
  - Largest of all ITR awards
  - Excellent reviews emphasizing importance of work
  - Joint NSF oversight from CISE and MPS
- **Scope of ITR funding**
  - Major costs for people, esp. students, postdocs
  - No hardware or professional staff for operations !
  - 2/3 CS + 1/3 application science
  - Industry partnerships being developed
    - Microsoft, Intel, IBM, Sun, HP, SGI, Compaq, Cisco

Still require funding for implementation and operation of Tier 2 centers

GriPhyN Institutions

- U Florida
- U Chicago
- Caltech
- U Wisconsin, Madison
- USC/ISI
- Harvard
- Indiana
- Johns Hopkins
- Northwestern
- Stanford
- Boston U
- U Illinois at Chicago
- U Penn
- U Texas, Brownsville
- U Wisconsin, Milwaukee
- UC Berkeley
- UC San Diego
- San Diego Supercomputer Center
- Lawrence Berkeley Lab
- Argonne
- Fermilab
- Brookhaven

Fundamental IT Challenge

"Scientific communities of thousands, distributed globally, and served by networks with bandwidths varying by orders of magnitude, need to extract small signals from enormous backgrounds via computationally demanding (Teraflops-Petaflops) analysis of datasets that will grow by at least 3 orders of magnitude over the next decade: from the 100 Terabyte to the 100 Petabyte scale."

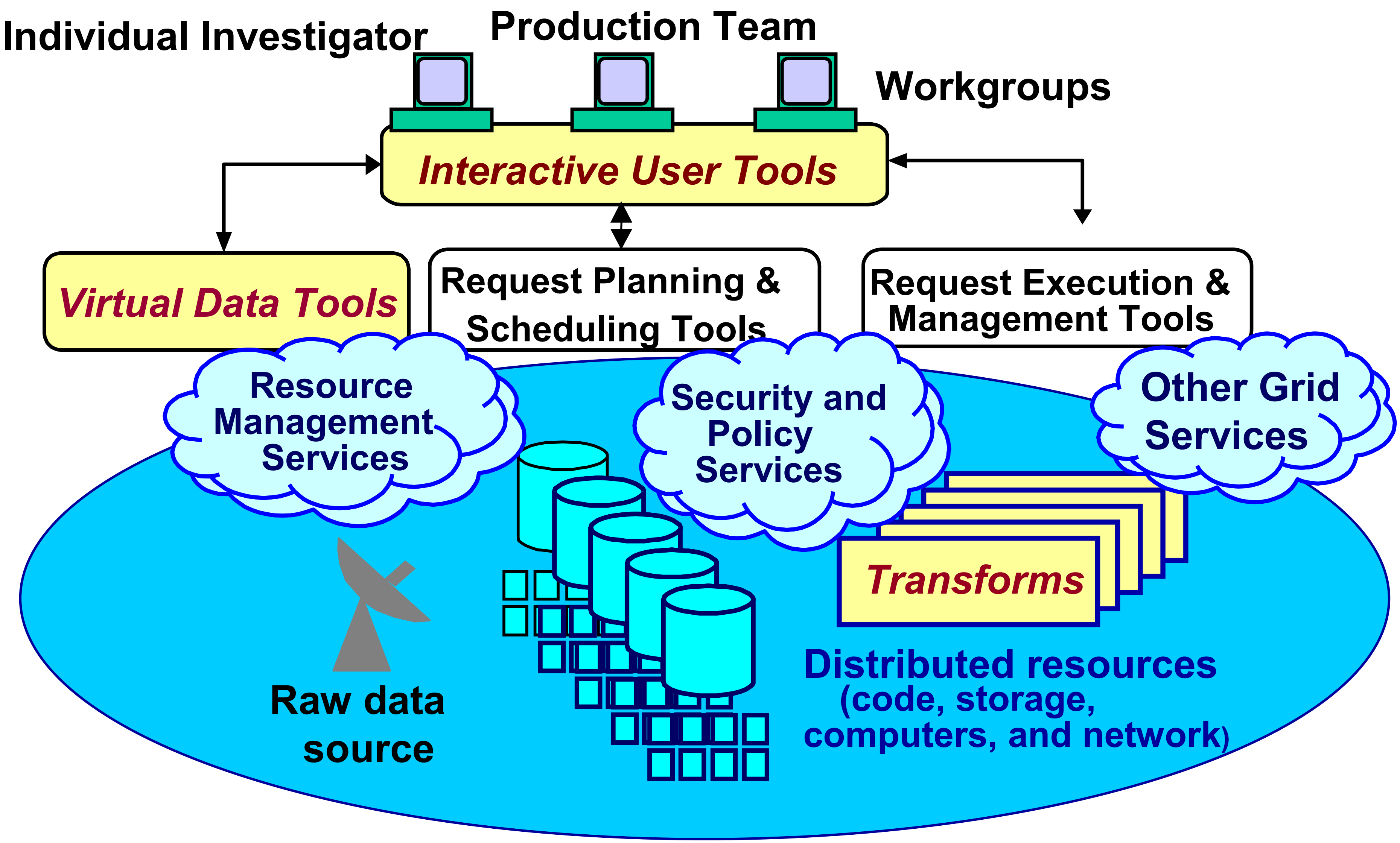
GriPhyN Research Agenda

- **Virtual Data technologies**
  - Derived data, calculable via algorithm (e.g., most HEP data)
  - Instantiated 0, 1, or many times
  - Fetch vs execute algorithm
  - Very complex (versions, consistency, cost calculation, etc)
- **Planning and scheduling**
  - User requirements (time vs cost)
  - Global and local policies + resource availability
  - Complexity of scheduling in dynamic environment (hierarchy)
  - Optimization and ordering of multiple scenarios
  - Requires simulation tools, e.g. MONARC

Research Agenda (cont.)

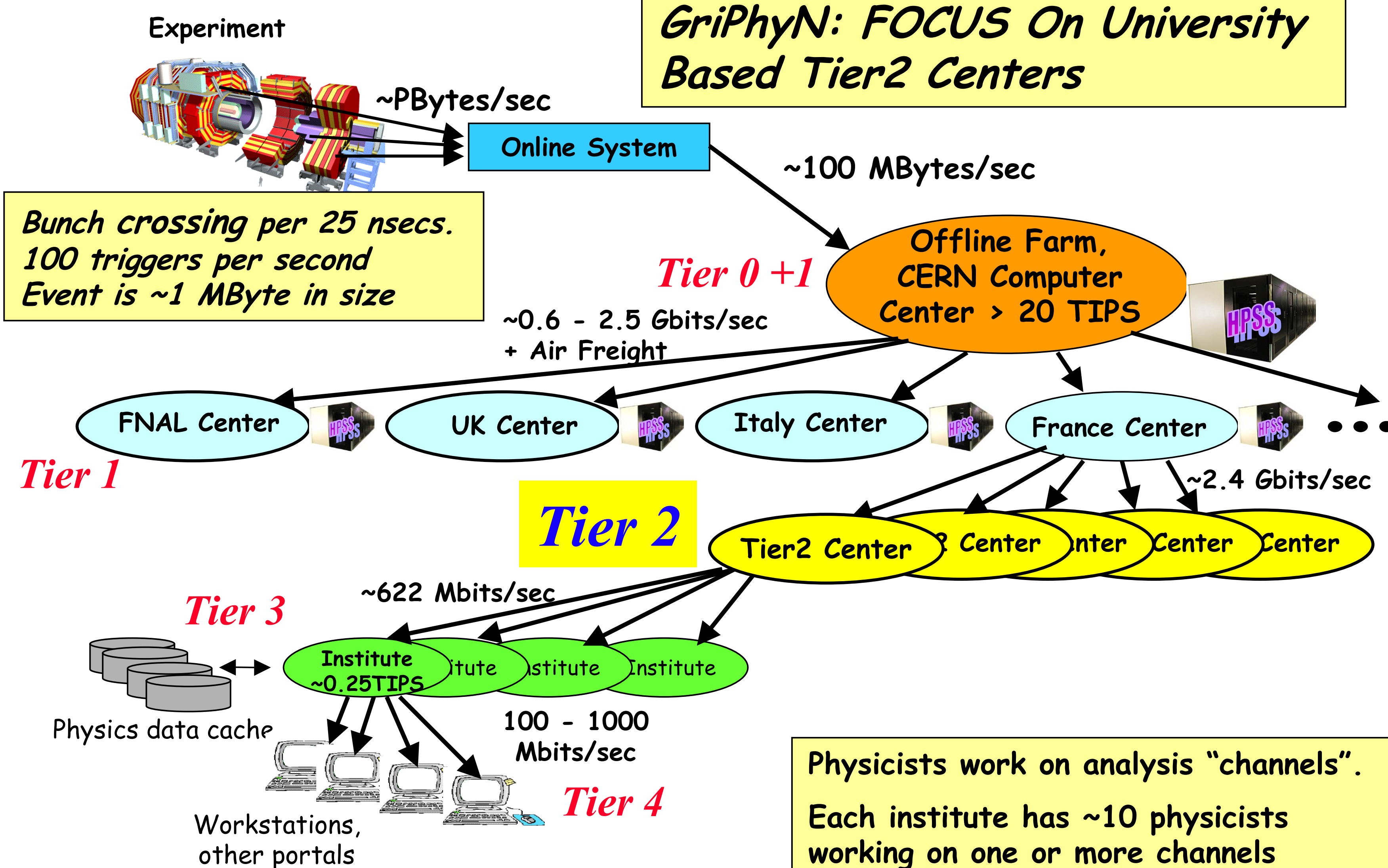
- Execution management
  - Co-allocation of resources (CPU, storage, network transfers)
  - Fault tolerance, error reporting
  - Agents (co-allocation, execution) $\Rightarrow$ H. Newman talk
  - Reliable event service across Grid
  - Interaction, feedback to planning
- Performance analysis (new)
  - Instrumentation and measurement of all grid components
  - Understand and optimize grid performance
- Virtual Data Toolkit (VDT)
  - VDT = virtual data services + virtual data tools
  - One of the primary deliverables of R&D effort
  - Ongoing activity + feedback from experiments (5 year plan)
  - Technology transfer mechanism to other scientific domains

GriPhyN: PetaScale Virtual Data Grids



LHC Vision (e.g., CMS Hierarchy)

GriPhyN: FOCUS On University Based Tier2 Centers

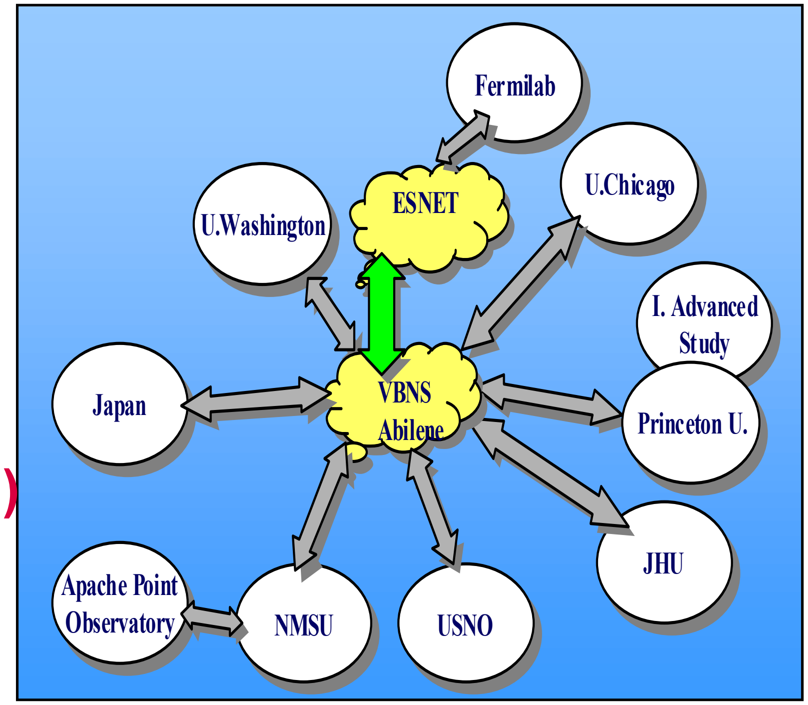


Bunch crossing per 25 nsecs.
100 triggers per second
Event is ~1 MByte in size

Physicists work on analysis "channels".

Each institute has ~10 physicists working on one or more channels

SDSS Vision



Three main functions:

Main data processing (FNAL)

- Processing of raw data on a grid
- Rapid turnaround with multiple TB data
- Accessible storage of all imaging data

Fast science analysis environment (JHU)

- Combined data access and analysis of calibrated data
- Shared by the whole collaboration
- Distributed I/O layer and processing layer
- Connected via redundant network paths

Public data access

- Provide the SDSS data for the NVO (National Virtual Observatory)
- Complex query engine for the public
- SDSS data browsing for astronomers, and outreach

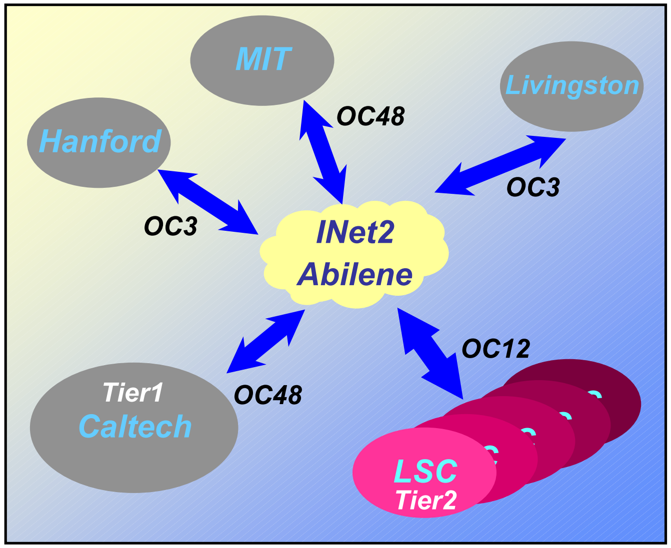
LIGO Vision

LIGO I Science Run: 2002 – 2004+
LIGO II Upgrade: 2005 – 20xx (MRE to NSF 10/2000)

Principal areas of GriPhyN applicability:

Main data processing (Caltech/CACR)

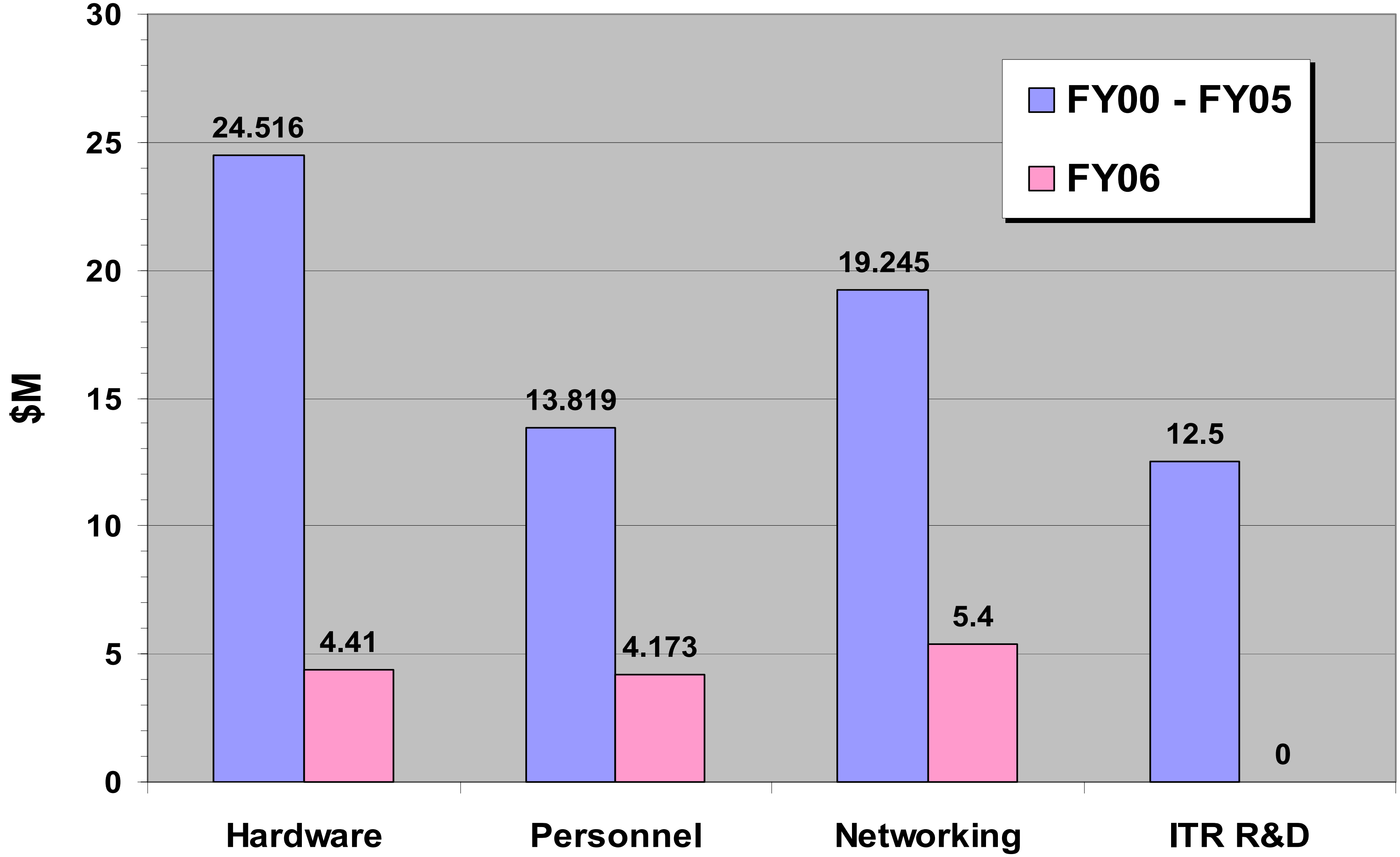
- Enable computationally limited searches
  ⇒ periodic sources)
- Access to LIGO deep archive
- Access to Observatories



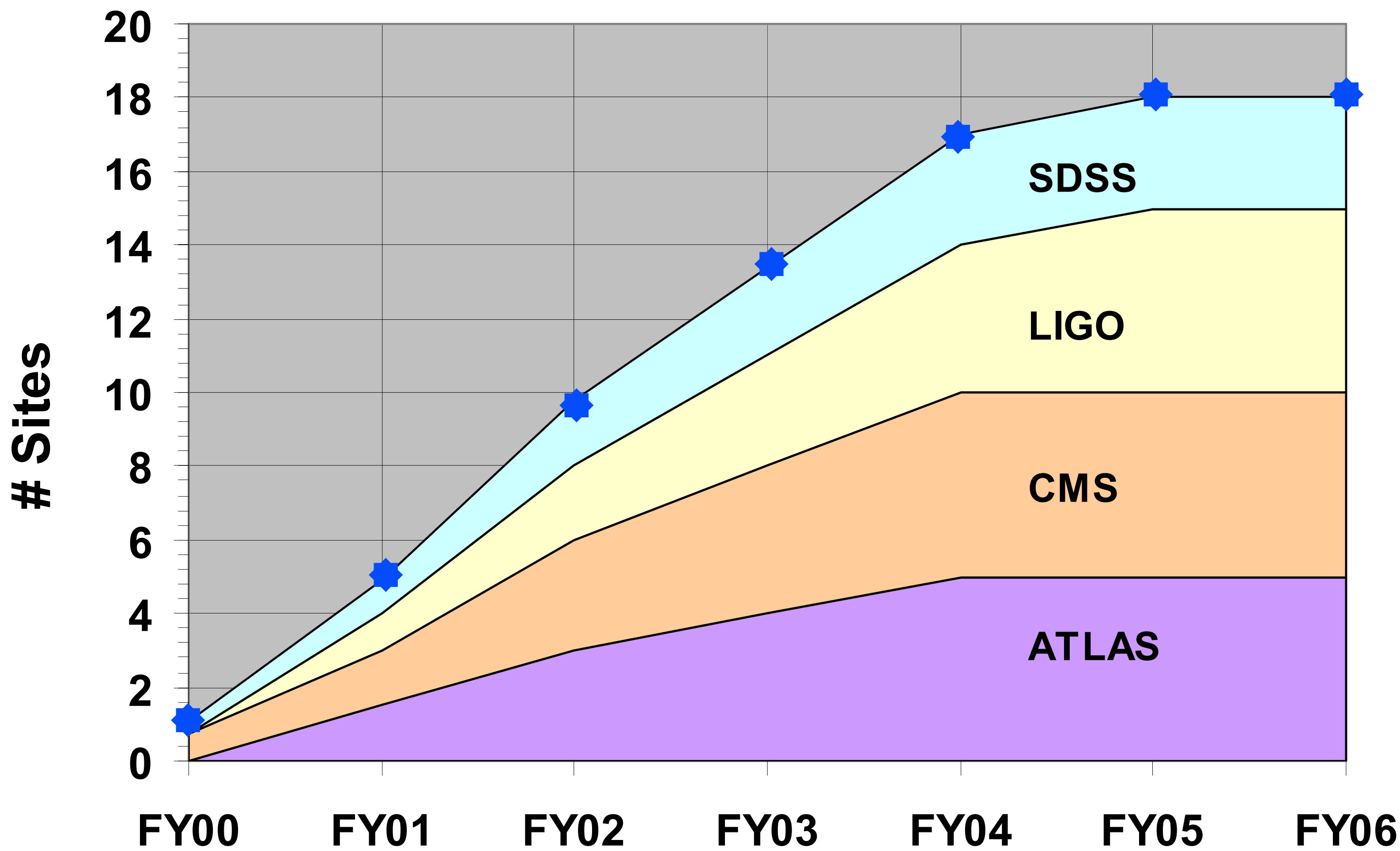
Science analysis environment for LSC (LIGO Scientific Collaboration)

- Tier2 centers: shared LSC resource
- Exploratory algorithm, astrophysics research with LIGO reduced data sets
- Distributed I/O layer and processing layer builds on existing APIs
- Data mining of LIGO (event) metadatabases
- LIGO data browsing for LSC members, outreach

GriPhyN Cost Breakdown (May 31)



Number of Tier2 Sites vs Time (May 31)



GriPhyN

Grid Program of Work (e.g., CMS)

2000
- Q3: First proto-Tier2 center; Build basic services; "4 Million" event samples
- Q4: MONARC Phase 3 simulations

2001
- Q1: Initial work with Tier1+Tier2 Simulations
- Q1: First set of Tier2 centers
- Q2: Early Grid services
  - File replication
  - Multi-site caching

2002
- Q3: Early subsystem data challenges

2003
- Q2: Tier2 partially deployed at all sites
- Q4: Production-prototype test of Grid hierarchy, with 1st elements of Tier1

2004
- Q2: Larger data challenges
- Q4: Production scale Grid hierarchy

2005
- Q1: Final production shakedown
- Q2: Full distributed system software and instrumentation

LHC Tier2 Architecture and Cost

◆ Linux Farm of 128 Nodes (256 CPUs + disk)	$ 350 K
◆ Data Server with RAID Array	$ 150 K
◆ Tape Library	$ 50 K
◆ Tape Media and Consumables	$ 40 K
◆ LAN Switch	$ 60 K
◆ Collaborative Tools & Infrastructure	$ 50 K
◆ Installation & Infrastructure	$ 50 K
◆ Net Connect to WAN (Abilene)	$ 300 K
◆ Staff (Ops and System Support)	$ 200 K♥
◆ Total Estimated Cost (First Year)	$1,250 K
◆ Average Yearly Cost including evolution, upgrade and operations✸	$ 750K

♥ 1.5 – 2 FTE support required per Tier2

✸ Assumes 3 year hardware replacement

Current Grid Developments

- **EU "DataGrid" initiative ⇒ F. Gagliardi talk**
  - Approved by EU in August (3 years, $9M)
  - Exploits GriPhyN and related (Globus, PPDG) R&D
  - Collaboration with GriPhyN (tools, Boards, interoperability, some common infrastructure)
  - http://grid.web.cern.ch/grid/
- **Rapidly increasing interest in Grids**
  - Nuclear physics
  - Advanced Photon Source (APS)
  - Earthquake simulations (http://www.neesgrid.org/)
  - Biology (genomics, proteomics, brain scans, medicine)
  - "Virtual" Observatories (NVO, GVO, …)
  - Simulations of epidemics (Global Virtual Population Lab)
- **GriPhyN continues to seek funds to implement vision**