NSF Elementary Particle Physics and Particle Astrophysics

LHC Tier 2 Centers

Presentation at the HEPAP Meeting

July 11, 2005

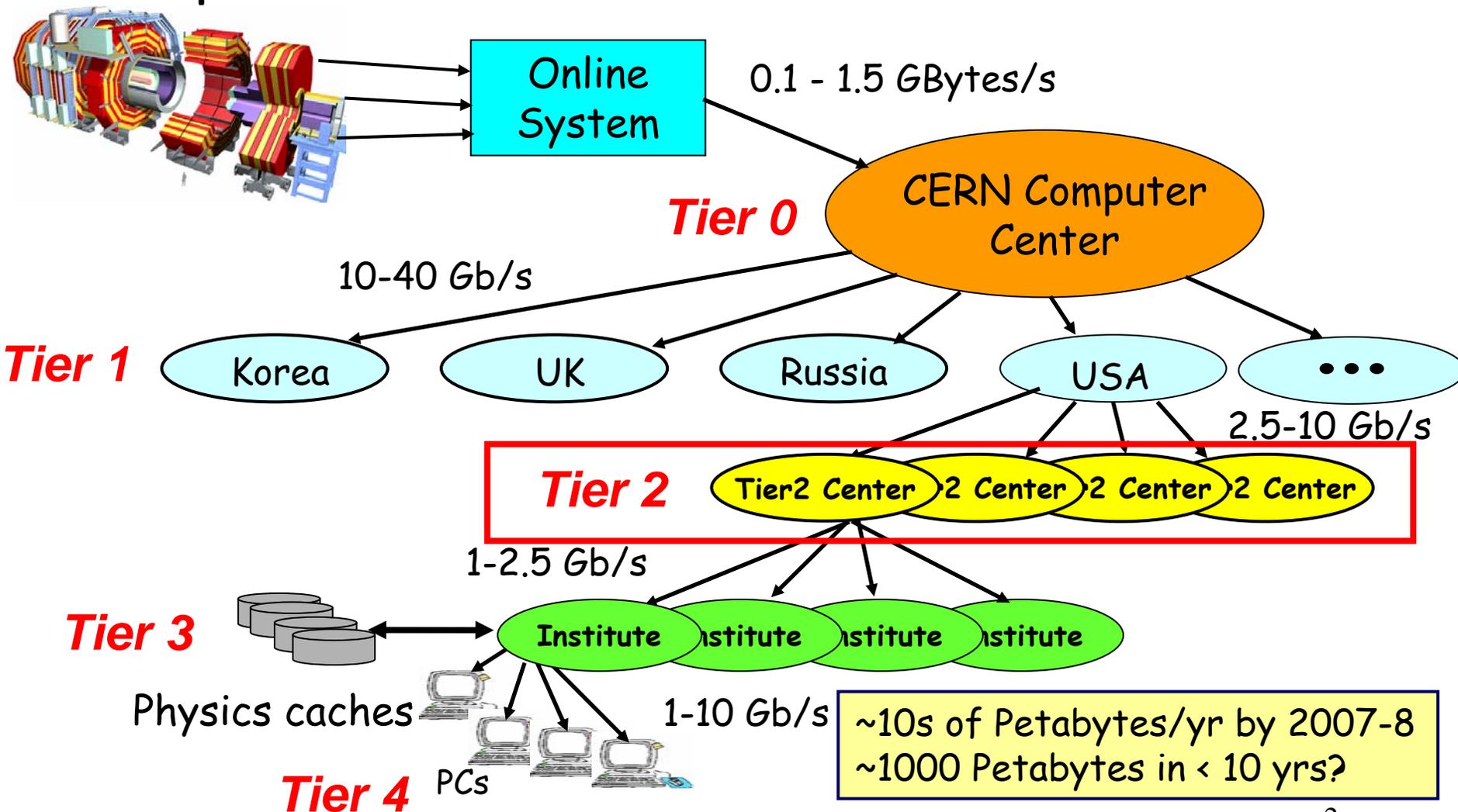Jim Whitmore   Marv Goldberg

Jim Stone   Randy Ruchti

How can the LHC Collaborators at Universities Participate?
## THE GRID BASED LHC DISCOVERY MACHINE

Experiment

Work within ITR/CISE Community ; LHC a Driver

Online System

0.1 - 1.5 GBytes/s

Tier 0

CERN Computer Center

10-40 Gb/s

Tier 1

Korea   UK   Russia   USA   •••

2.5-10 Gb/s

Tier 2   Tier2 Center   2 Center   2 Center   2 Center

1-2.5 Gb/s

Tier 3

Institute   Institute   Institute   Institute

Physics caches

1-10 Gb/s

Tier 4   PCs

~10s of Petabytes/yr by 2007-8
~1000 Petabytes in < 10 yrs?

Cyberinfrastructure and Grids

- Grid: Geographically distributed computing resources configured for coordinated use
    - Fabric: Physical resources & networks provide raw capability
    - Ownership: Resources *controlled* by owners and *shared* w/ others
    - Middleware: Software ties it all together: tools, services, etc.
- Cyberinfrastructure: "infrastructure based upon distributed computer, information and communication technology", ie "the enabling hardware, algorithms, software, communications, institutions, and personnel" (Atkins Report, Jan 2003)
- Enhancing collaboration via transparent resource sharing
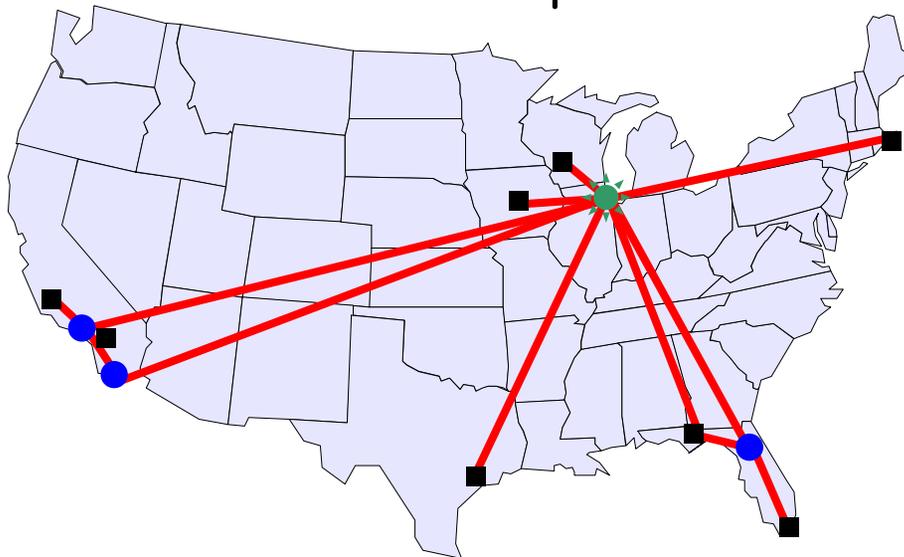
Initial Science Grid Drivers

- Experiments at Large Hadron Collider (LHC)
  - New fundamental particles and forces
  - 10s of Petabytes/yr   2007 - ?
- High Energy & Nuclear Physics expts
  - Top quark, nuclear matter at extreme density
  - ~1 Petabyte (1000 TB)   1997 – present
- LIGO (gravity wave search)
  - Search for gravitational waves
  - 100s of Terabytes   2002 – present
- Sloan Digital Sky Survey
  - Systematic survey of astronomical objects
  - 10s of Terabytes   2001 – present

2009

2007

2005

2003

2001

Data growth
Community growth

Data Distributed by Tier n Centers

Integrating Universities and Laboratories In a National Cyberinfrastructure

## With International Interoperability
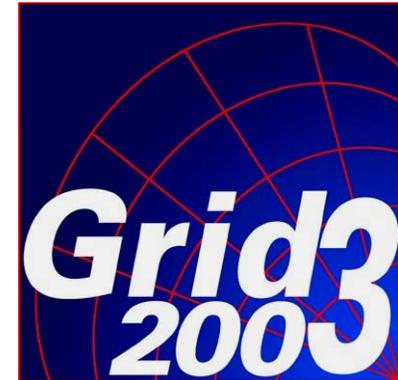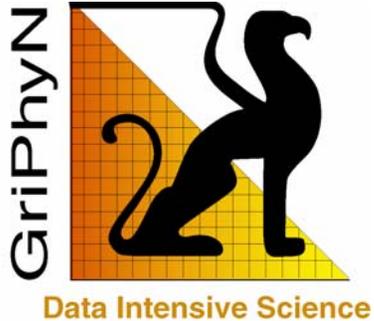
IMPLEMENTATION

GriPhyN + iVDGL + DOE Particle Physics Data Grid (PPDG) = **Trillium**

End-to-end

- Develop the **technologies** & **tools** needed to exploit a Grid-based cyberinfrastructure
- **Apply** and **evaluate** those technologies & tools in challenging scientific problems
- Develop the **technologies** & **procedures** to support a permanent Grid-based cyberinfrastructure
- **Create** and **operate** a persistent Grid-based cyberinfrastructure in support of discipline-specific research goals
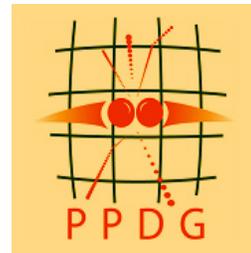
➢ (~150 people)

Grid3: A National Grid Infrastructure

- 35 sites, 3500 CPUs: Universities + 4 national labs
- Part of LHC Grid
- Running since October 2003; Grid Ops Center
- Applications in HEP, LIGO, SDSS, Genomics, CS Testbed

http://www.ivdgl.org/grid3
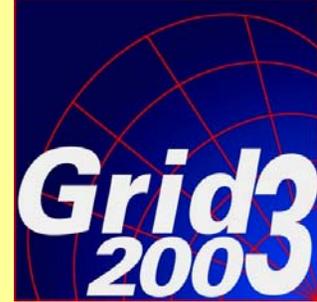
Open Science Grid -- Roadmap

- Build upon existing achievements towards a sustained US national production grid for the long term – past 2010

- US LHC will build and contribute their resources into a coherent infrastructure to provide the initial federation

- Develop the general Grid infrastructure to support other sciences

- Partnership between application scientists, technology providers and resource owners based on proven achievements as an effective strategy for success

From R. Pordes (Fermilab) June 2004

US CMS Tier-2 Activities

- US-CMS is bringing up services on 3 new Tier-2s (MIT, Nebraska and Purdue) and two Tier-2Cs (CIT/UCSD/Florida and Wisconsin). "Tier-2C" means it will also develop grid technology for other disciplines on their campus
  - Tier-2s will participate in LCG Service Challenge 3 (SC3)
    - The goal is to transfer data from CERN to Tier-1 centers and on to Tier-2 centers.   Nebraska and Purdue have both started getting data from CERN via Fermilab
  - The Tier-2s are deploying the production version of the Open Science Grid software stack (OSG 0.2)
    - The Tier-2C centers all participated in the OSG integration activities and experiment simulation and analysis activities; these sites now use OSG 0.2
  - All Tier-2 centers will be involved in simulated event production and will ramp up analysis activities

US CMS Tier-2 Goals

Processing

Goal is to deploy around 40-60 boxes (80-120) CPUs at each Tier-2 site this year.   CMS has a need for simulation and analysis resources and the medium sized farm provides a reasonable resource for exercising the storage installation

Storage

Sites will deploy 20-40TB of dCache space this year.   This puts US-CMS on a reasonable operations ramp toward 200TB in 2008.

Networking

The existing network available at each Tier-2 varies widely, but CMS would like to see usage of at least 50% of the slowest link between the Tier-2s and FNAL.

Grid Interfaces and CMS specific services installed

Central CMS simulation supported through the OSG interfaces

Tier-2s will host samples of CMS simulated data for local analysis activities through Grid and interactive access
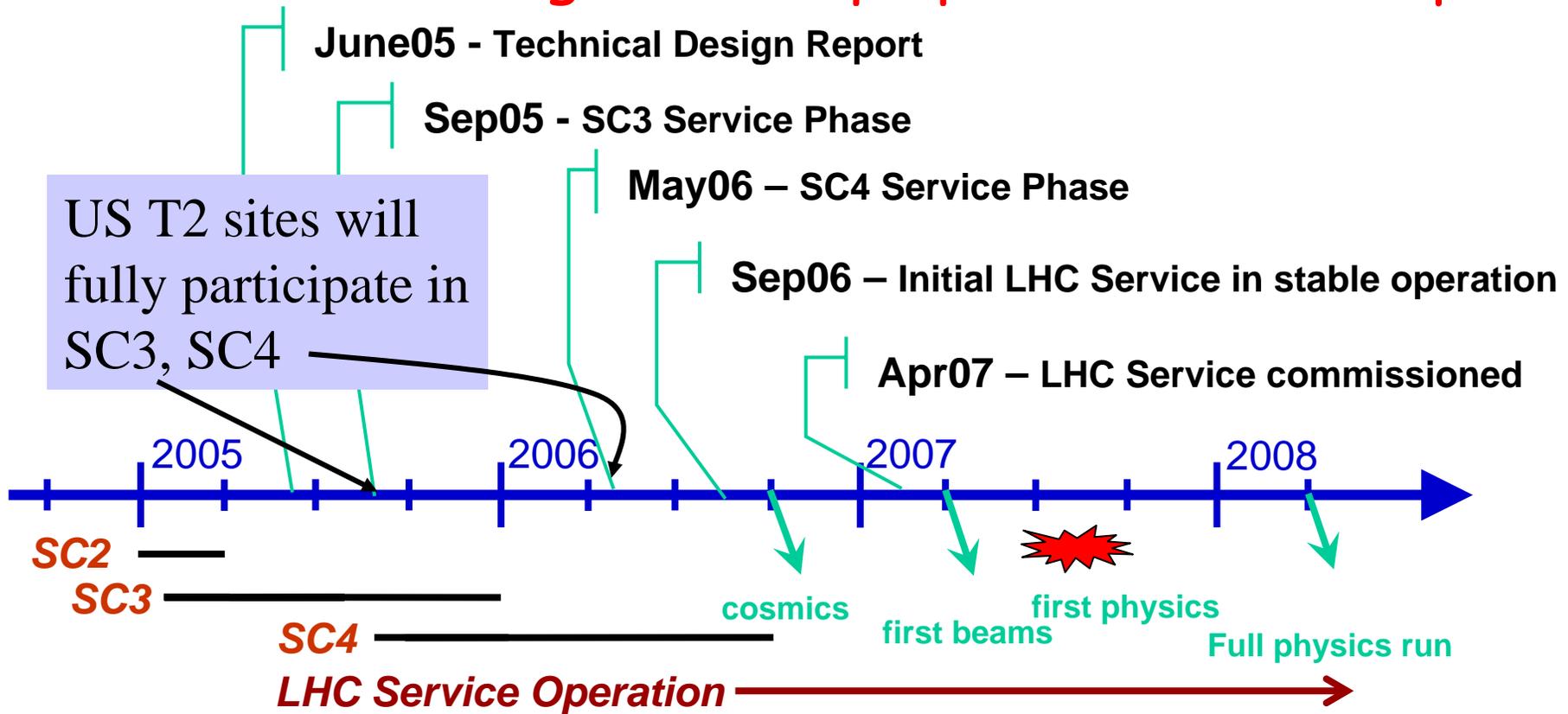
US CMS Tier-2 Progress

- US CMS recently received an NSF award for Data Intensive Science University Network (**DISUN**) for Tier-2C as a joint **EPP/SCI** venture ($2M/yr for 5 yrs).

- The DISUN Tier-2C work plan for distributed computing infrastructure is now starting

- **CISE/SCI** has a definite interest in this since the Tier-2Cs will be developing grid technology for both EPP and other disciplines in science

US ATLAS Tier-2 activity

- We are now having discussions with **US ATLAS** on how to help satisfy their needs and also to further the progress of general Grid technology for other users:

- U.S. ATLAS has recently selected three Tier-2 Centers:
  - One Tier-2: U. Chicago/Indiana U.
  - And two Tier-2Cs: Boston (Boston U./Harvard) and the Southwest (UTA/Oklahoma/Langston/New Mexico)
  - They are already working and are major producers of ATLAS data
  - They need to function as part of the ATLAS Computing System Commissioning/Service Challenges in 2005!

There are LCG Timelines.
Service Challenges – ramp up to LHC start-up

June05 - Technical Design Report

Sep05 - SC3 Service Phase

May06 – SC4 Service Phase

Sep06 – Initial LHC Service in stable operation

Apr07 – LHC Service commissioned

US T2 sites will fully participate in SC3, SC4

2005   2006   2007   2008

SC2

SC3

SC4

LHC Service Operation

cosmics

first beams

first physics

Full physics run

SC2 – Reliable data transfer (disk-network-disk) – 5 Tier-1s, aggregate 500 MB/sec sustained at CERN
SC3 – Reliable base service – most Tier-1s, **some Tier-2s** – basic experiment software chain – grid data throughput 500 MB/sec, including mass storage (~25% of the nominal final throughput for the proton period)
SC4 – All Tier-1s, major Tier-2s – capable of supporting full experiment software chain inc. analysis – sustain nominal final grid data throughput
LHC Service in Operation – September 2006 – ramp up to full operational capacity by April 2007 – capable of handling twice the nominal data throughput

The ATLAS Computing Model

- Tier 2 centers play a vital role in ATLAS
  - The only resource for the large simulations needed to fully understand the systematic errors in ATLAS
  - A prime resource for physicists to do analysis.
- Worldwide, there are approximately 30 T2s in ATLAS
  - Approximate Overall ATLAS CAPACITY in 2008 needs:
    - **21 MSi2k   CPU**
    - **9 PB   Disk**
  - 20% in US to satisfy commitments to ATLAS:
    - **4.2 MSi2k CPU**
    - **1.8 PB   Disk**
  - U.S. ATLAS physicists needs at our T2s require more resources. Current estimate (March): 6.5 MSi2k for our total T2 capacity, Resulting in our average T2 in 2008:   Current T2's (approx.):

    - **1.3 MSi2k CPU**
    - **600 TB   Disk**

    - **0.1 MSi2k CPU**
    - **10 TB   Disk**

Coordinated Tier 1, Tier2 Roles for DC2/ATLAS Rome Physics Workshop

- Tier 1
  - Hosted numerous required grid and DB services.
  - 2nd largest producer of Rome data for all of ATLAS ~6%
  - Hosted all U.S. data – disk based and tape archive
  - Support personnel for core services and data management
  - 24x7 Operations model with Indiana iGOC
- Tier 2's
  - Large scale production and user support services ~14% of ATLAS-wide Rome production was done at three U.S. Tier 2 sites
  - Boston Tier 2: **Pacman** – adopted by ATLAS, VDT + others, package for software distribution, crucial for Grid3 success
  - SW Tier 2: **Windmill** – automated supervisor used for ATLAS-wide production, unique in supporting all grids (Grid3, Nordugrid, LCG)
  - MW Tier 2: Grid3 deployment, metrics, Operations, **Capone** Workload management system for Grid3
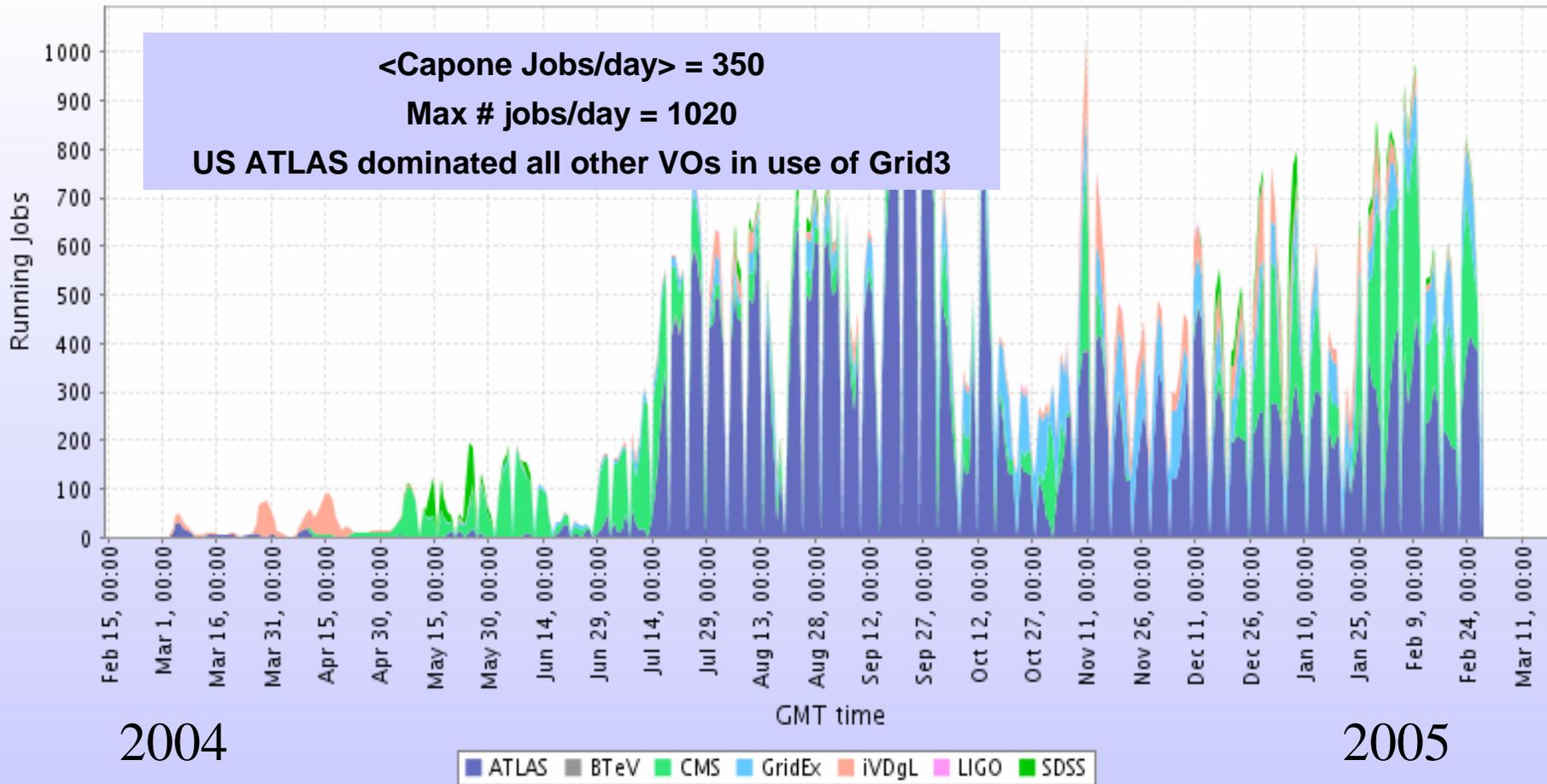
Rome Physics Workshop Grid Production Successful Job Count at 83 ATLAS sites
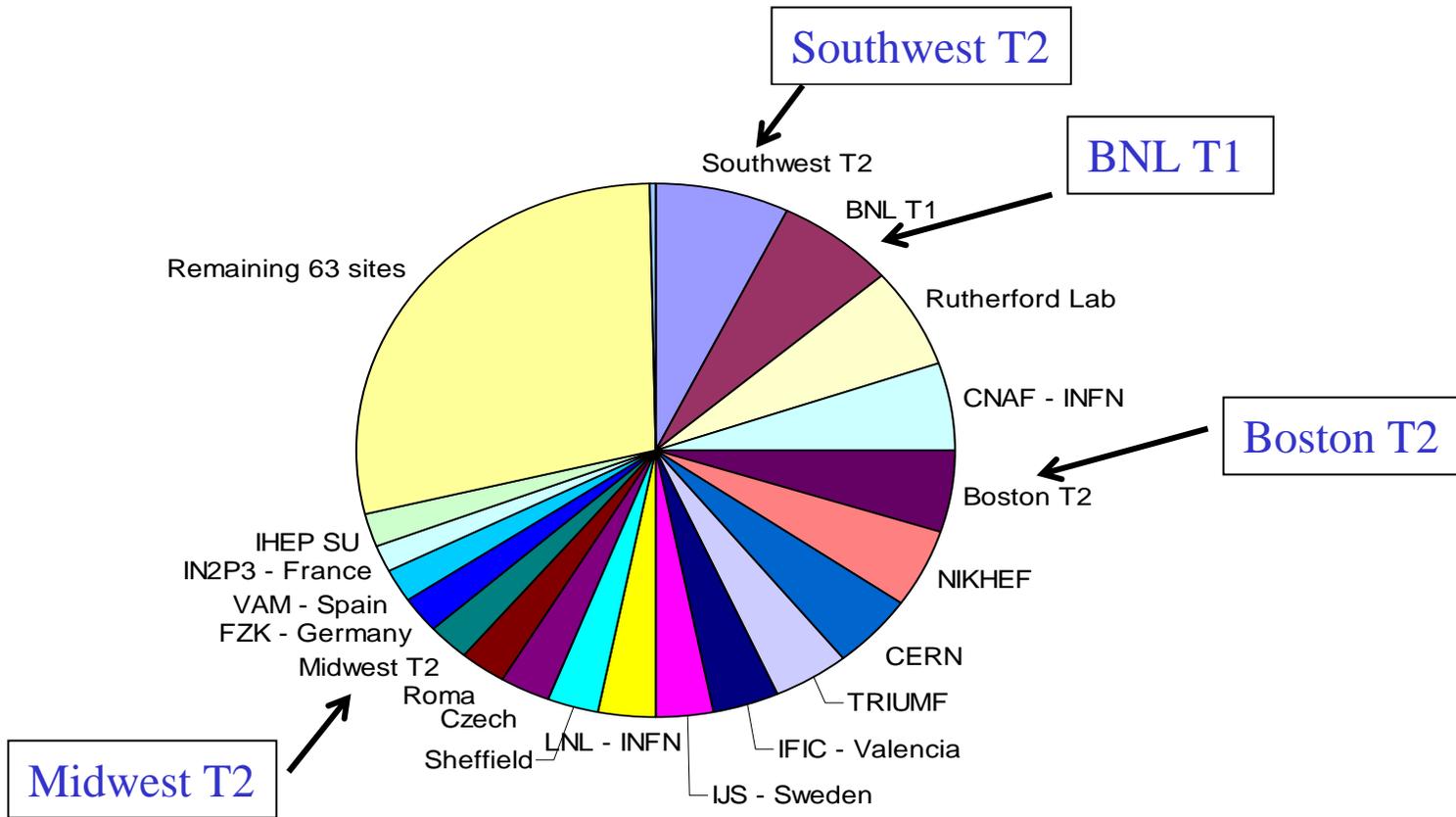
3 out of top 5 contributions in ATLAS were from U.S.
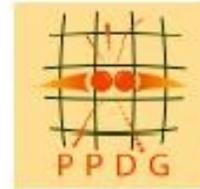
US LHC Domination on Grid3

Global jobs view

<Capone Jobs/day> = 350

Max # jobs/day = 1020

US ATLAS dominated all other VOs in use of Grid3

2004   2005

Legend: ATLAS, BTeV, CMS, GridEx, iVDgL, LIGO, SDSS

17

U.S. "Trillium" Grid Partnership

> Trillium = PPDG + GriPhyN + iVDGL
  - ◆ Particle Physics Data Grid: $12M (DOE)   (1999 – 2004+)
  - ◆ GriPhyN: $12M (NSF)   (2000 – 2005)
  - ◆ iVDGL: $14M (NSF)   (2001 – 2006)
> Basic composition   (~150 people)
  - ◆ PPDG: 4 universities, 6 labs
  - ◆ GriPhyN: 12 universities, SDSC, 3 labs
  - ◆ iVDGL: 18 universities, SDSC, 4 labs, foreign partners

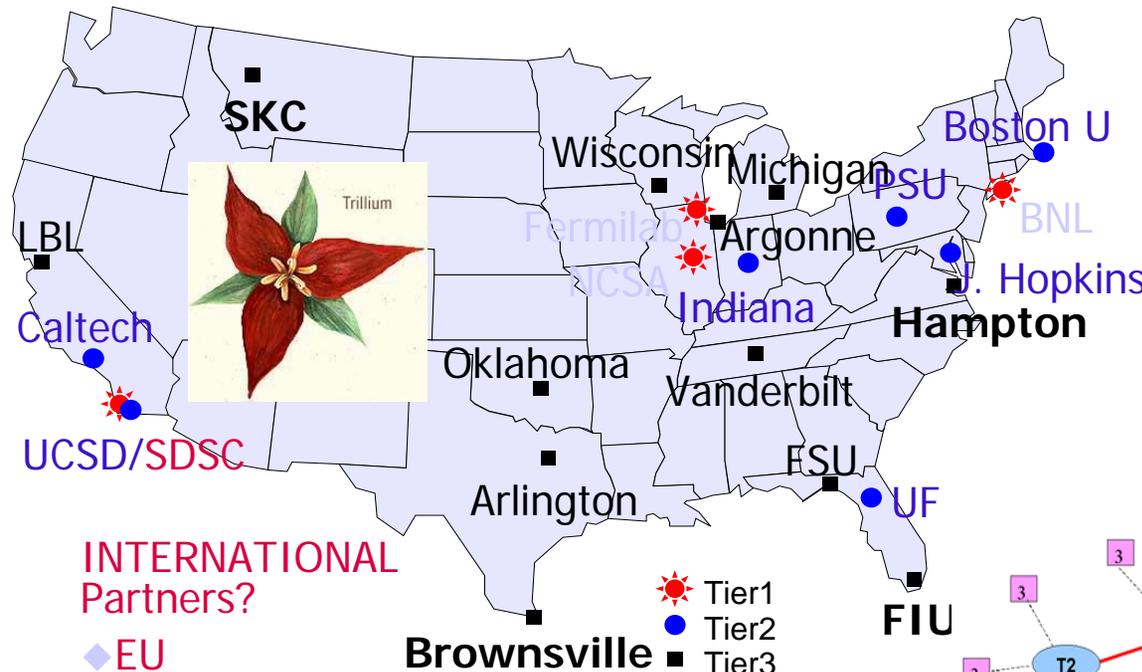**NSF FUNDING THROUGH ITR; *NOW SCI!***

LHC Research Program

- Funding supported by NSF-CISE (and PHY):

	FY03	FY04	FY05	FY06	FY07
GriPhyN	$2.5M	$2.2M			
iVDGL	$2.8M	$2.8M	$2.8M		
Ultralight		$0.5M	$0.5M	$0.5M	$0.5M
Other ITR	$1.0M	$1.0M	$0.9M	$0.4M	$0.4M

- Support for OSG-LCG-EGEE cooperation (**$0.25M**)

- Support for IT Communications PR expert (**for 2 yrs**)

- <u>Issue</u>: some of these are expiring soon

- Open Science Grid (OSG) ??

NEW PLANNING
**Trillium Sites Provide Grid Infostructure to Quarknet Sites**

SKC
Boston U
Wisconsin
Michigan
PSU
Fermilab
Argonne
BNL
LBL
NCSA
J. Hopkins
Caltech
Indiana
Hampton
Oklahoma
UCSD/SDSC
Vanderbilt
Arlington
FSU
UF
Brownsville
FIU

Trillium

Tier1
Tier2
Tier3

INTERNATIONAL Partners?
◆EU
◆CERN
◆Brazil
◆Australia
◆Korea
◆Japan

MORE NEW PLANNING
Quarknet E/O expands to teach Quarks/Cosmos and Infostructure.

Tier 0 (CERN)
T2
T2
Tier 1
T2
T2
T2
3
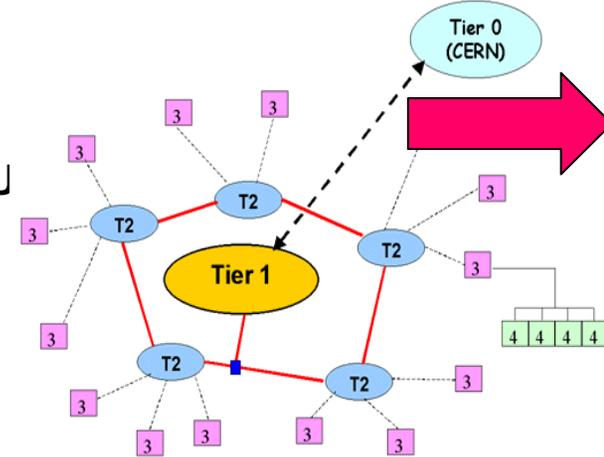
EXPT. *and Education*

0. Education Center

1. University

2. High Schools

3. Teachers

4. Students

QuarkNet

Sites

Hawaii
Puerto Rico

20

![iVD gL]

GriPhyN/iVDGL Outreach Sites

→ U Texas, Brownsville (Hispanic)
  ◆ Lead institution, Manuela Campanelli is E/O Director
  ◆ Strong LIGO physics institution

→ Hampton University (HBCU)
  ◆ Keith Baker, PI
  ◆ High Energy Physics (ATLAS), COSM (Physics Frontier Center)
  ◆ QuarkNet leadership   AFRICAN EXPERIENCES

→ Salish Kootenai College (Native American)
  ◆ Tim Olsen, PI
  ◆ LIGO physics

→ Florida International University (Hispanic) ** New **
  ◆ H. Alvarez, J. Ibarra, L. Kramer, P. Markowitz, etc.
  ▶ Strong efforts in networking (AMPATH to South America)
  ▶ CHEPREO initiative (HEP, Grids, networks, secondary schools)

Grid Meeting (Sep. 17, 2003)   Paul Avery   4

Multiple Interactions, Outsourcing

FIU: CISE; INT; EHR; EPP; OMA

21

Physics Fall Target Date

- **The target date for proposal submissions to the Division of Physics that are competing for FY 2006 funds is September 28, 2005.**

- **The above date does not apply to proposals sent to the Physics Division in response to Foundation-wide solicitations, such as the Faculty Early Career Development (CAREER – July 19, 2005) or Research Experiences for Undergraduates (REU) programs.**