Grid Technologies & Applications: Architecture & Achievements

Ian Foster

Mathematics and Computer Science Division

Argonne National Laboratory

and

Department of Computer Science

The University of Chicago

http://www.mcs.anl.gov/~foster

Computing in High Energy and Nuclear Physics, Beijing, P.R. China, 2001

Acknowledgments

- Globus R&D is joint with numerous people
  - Carl Kesselman, Co-PI
  - Steve Tuecke, principal architect at ANL
  - Others to be acknowledged
- GriPhyN R&D is joint with numerous people
  - Paul Avery, Co-PI; Newman, Lazzarini, Szalay
  - Mike Wilde, project coordinator
  - Carl Kesselman, Miron Livny CS leads
  - ATLAS, CMS, LIGO, SDSS participants; others
- Support: DOE, DARPA, NSF, NASA, Microsoft

Issues I Will Address

- Grids in a nutshell
  - Problem statement
  - Major Grid projects
- Progress and plans in two specific projects
  - Globus Project and Toolkit
  - GriPhyN project
- Enabling international cooperation
- Grids and commodity technologies

The Grid Problem

Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations
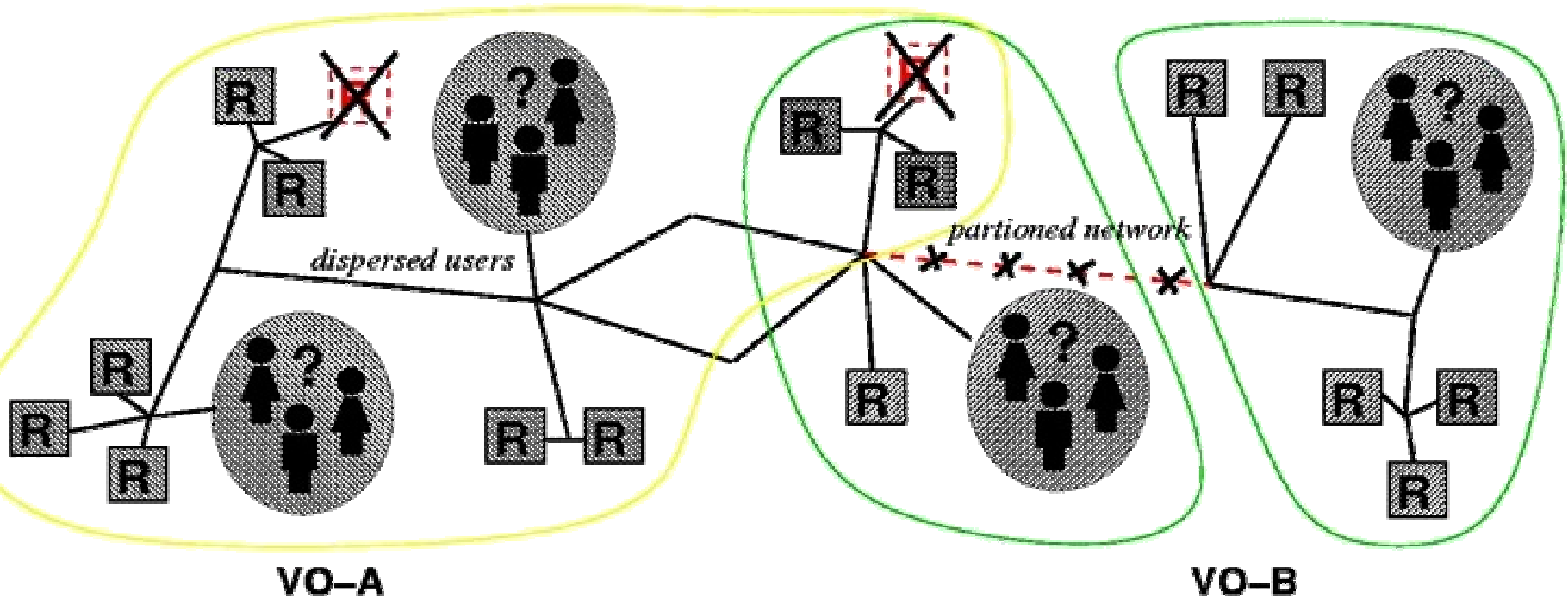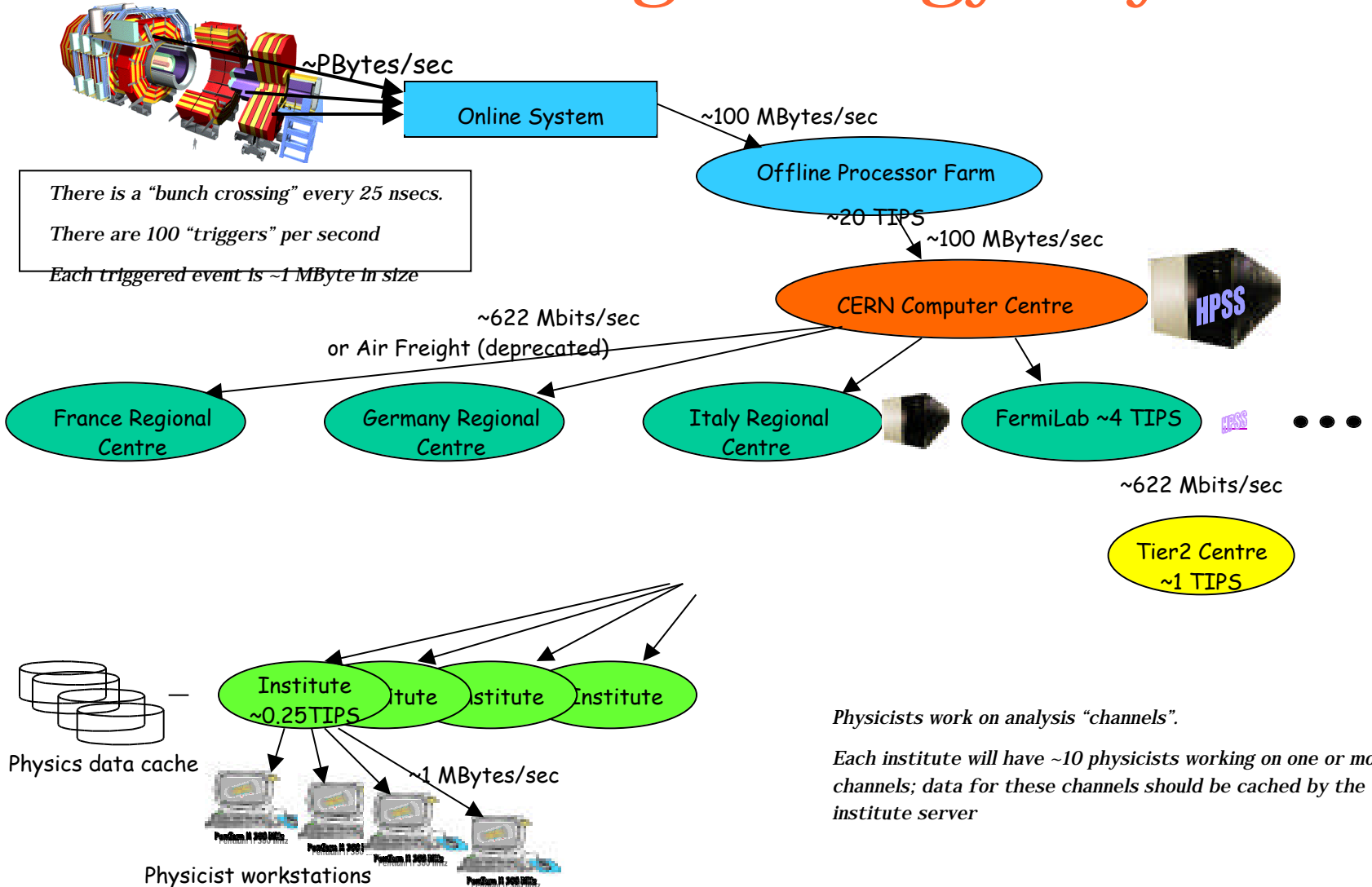
Elements of the Problem

- **Resource sharing**
  - Computers, storage, sensors, networks, ...
  - Sharing always conditional: issues of trust, policy, negotiation, payment, ...
- **Coordinated problem solving**
  - Beyond client-server: distributed data analysis, computation, collaboration, ...
- **Dynamic, multi-institutional virtual orgs**
  - Community overlays on classic org structures
 - Large or small, static or dynamic

Grid Communities & Applications: Data Grids for High Energy Physics

~PBytes/sec

Online System

~100 MBytes/sec

Offline Processor Farm

~20 TIPS

~100 MBytes/sec

There is a "bunch crossing" every 25 nsecs.

There are 100 "triggers" per second

Each triggered event is ~1 MByte in size

CERN Computer Centre

HPSS

~622 Mbits/sec
or Air Freight (deprecated)

France Regional Centre

Germany Regional Centre

Italy Regional Centre

FermiLab ~4 TIPS

• • •

~622 Mbits/sec

Tier2 Centre ~1 TIPS

Institute ~0.25TIPS **Itute** **stitute** **Institute**

Physics data cache

Physicists work on analysis "channels".

Each institute will have ~10 physicists working on one or more channels; data for these channels should be cached by the institute server

~1 MBytes/sec

Physicist workstations
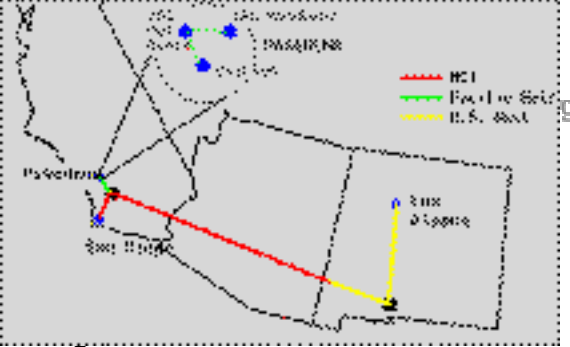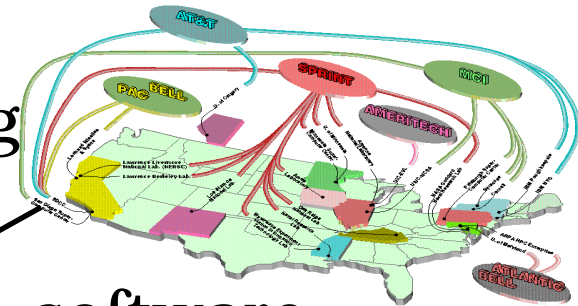
A Little History

- **Early 90s**
  - Gigabit testbeds, metacomputing
- **Mid to late 90s**
  - Early experiments (e.g., I-WAY), software projects (e.g., Globus), application experiments
- **2001**
  - Major application communities emerging
  - Major infrastructure deployments are underway
  - Rich technology base has been constructed
  - Global Grid Forum: >1000 people on mailing lists, 192 orgs at last meeting, 28 countries

Selected Major Grid Projects

Name	URL & Sponsors	Focus
Access Grid	www.mcs.anl.gov/FL/ accessgrid; DOE, NSF	Create & deploy group collaboration systems using commodity technologies
BlueGrid *New*	IBM	Grid testbed linking IBM laboratories
DISCOM	www.cs.sandia.gov/ discom DOE Defense Programs	Create operational Grid providing access to resources at three U.S. DOE weapons laboratories
DOE Science Grid *New*	sciencegrid.org DOE Office of Science	Create operational Grid providing access to resources & applications at U.S. DOE science laboratories & partner universities
Earth System Grid (ESG)	earthsystemgrid.org DOE Office of Science	Delivery and analysis of large climate model datasets for the climate research community
European Union (EU) DataGrid	eu-datagrid.org European Union	Create & apply an operational grid for applications in high energy physics, environmental science, bioinformatics

Selected Major Grid Projects

Name	URL/Sponso	Focus
EuroGrid, Grid Interoperability (GRIP) *New*	eurogrid.org European Union	Create technologies for remote access to supercomputer resources & simulation codes; in GRIP, integrate with Globus
Fusion Collaboratory *New*	fusiongrid.org DOE Off. Science	Create a national computational collaboratory for fusion research
Globus Project	globus.org DARPA, DOE, NSF, NASA, Msoft	Research on Grid technologies; development and support of Globus Toolkit; application and deployment
GridLab *New*	gridlab.org European Union	Grid technologies and applications
GridPP *New*	gridpp.ac.uk U.K. eScience	Create & apply an operational grid within the U.K. for particle physics research
Grid Research Integration Dev. & Support Center *New*	grids-center.org NSF	Integration, deployment, support of the NSF Middleware Infrastructure for research & education

Selected Major Grid Projects

Name	URL/Sponsor	Focus
Grid Application Dev. Software	hipersoft.rice.edu/ grads; NSF	Research into program development technologies for Grid applications
Grid Physics Network	griphyn.org NSF	Technology R&D for data analysis in physics expts: ATLAS, CMS, LIGO, SDSS
Information Power Grid	ipg.nasa.gov NASA	Create and apply a production Grid for aerosciences and other NASA missions
International Virtual Data Grid Laboratory *New*	ivdgl.org NSF	Create international Data Grid to enable large-scale experimentation on Grid technologies & applications
Network for Earthquake Eng. Simulation Grid *New*	neesgrid.org NSF	Create and apply a production Grid for earthquake engineering
Particle Physics Data Grid	ppdg.net DOE Science	Create and apply production Grids for data analysis in high energy and nuclear physics experiments

Selected Major Grid Projects

Name	URL/Sponsor	Focus
TeraGrid g *New*	teragrid.org NSF	U.S. science infrastructure linking four major resource sites at 40 Gb/s
UK Grid Support g Center *New*	grid-support.ac.uk U.K. eScience	Support center for Grid projects within the U.K.
Unicore	BMBFT	Technologies for remote access to supercomputers

Also many technology R&D projects:
e.g., Condor, NetSolve, Ninf, NWS

See also www.gridforum.org

Grid Communities and Applications: Network for Earthquake Eng. Simulation

- NEESgrid: national infrastructure to couple earthquake engineers with experimental facilities, databases, computers, & each other

- On-demand access to experiments, data streams, computing, archives, collaboration
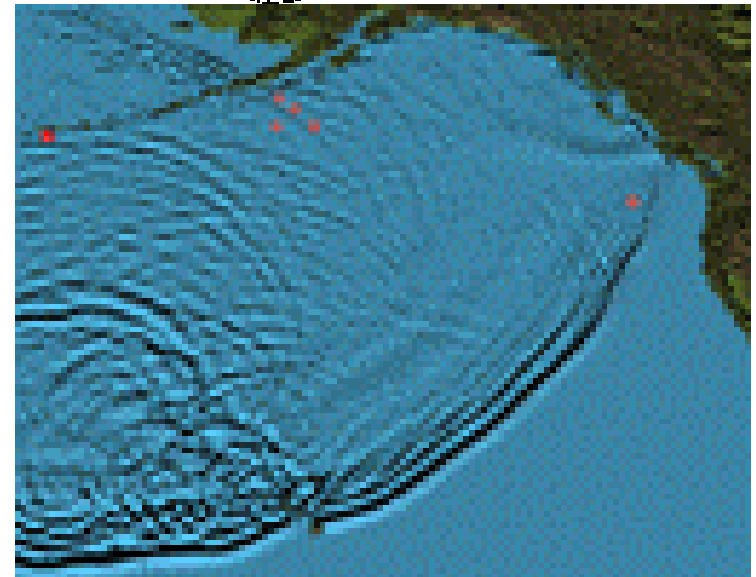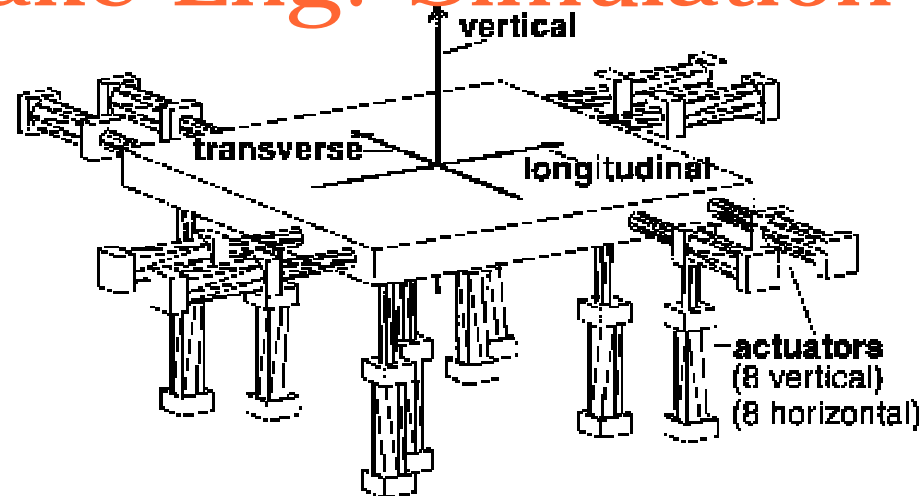
The 13.6 TF TeraGrid: Computing at 40 Gb/s

Site Resources

26

4 HPSS

24

8

External
Networks

Caltech

Site Resources

HPSS

External
Networks

Argonne

5

External
Networks

Site Resources

HPSS

**SDSC
4.1 TF
225 TB**

**NCSA/PACI
8 TF
240 TB**

External
Networks

Site Resources

UniTree

TeraGrid/DTF: NCSA, SDSC, Caltech, Argonne   www.teragrid.org
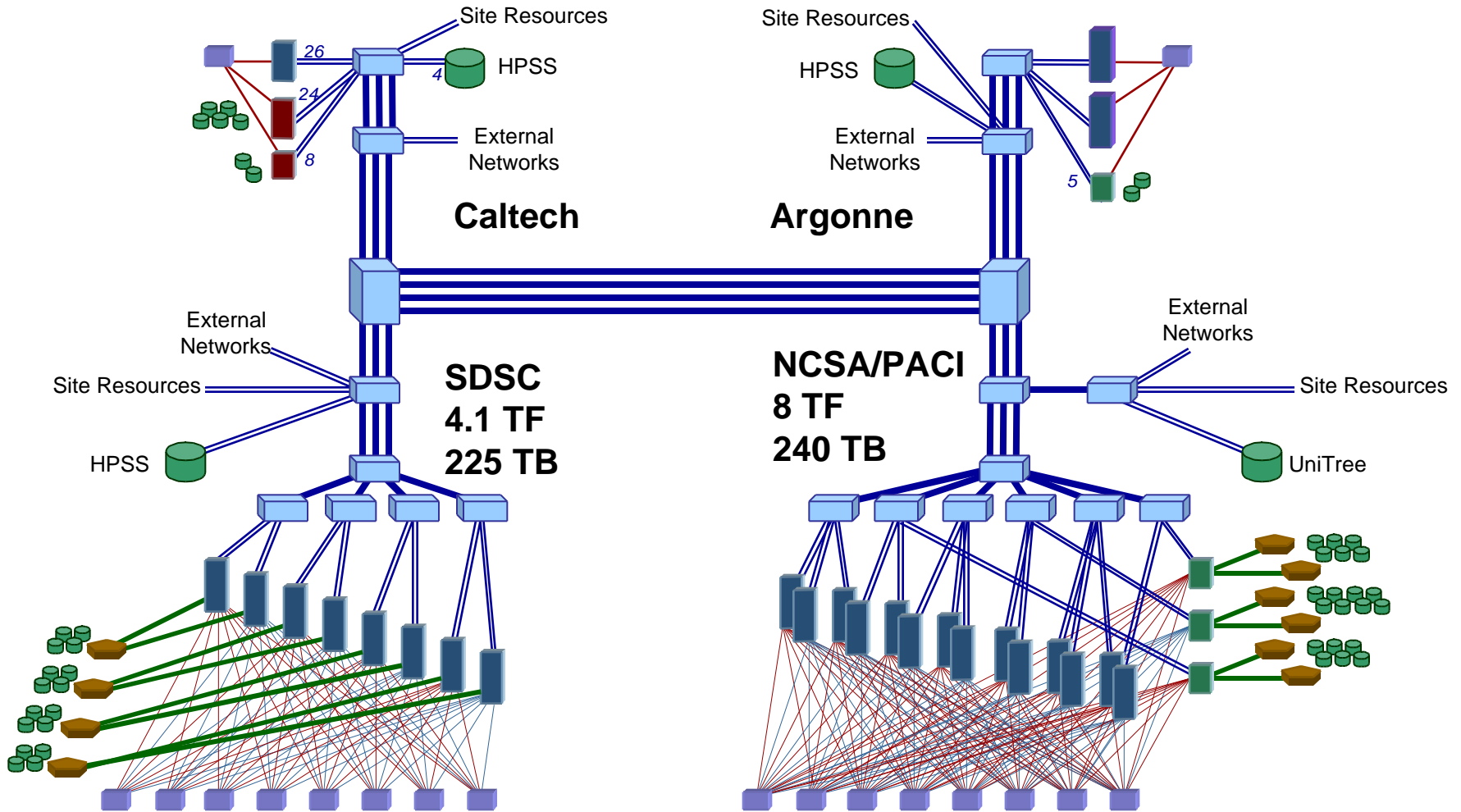
Access Grid

- Collaborative work among large groups
- ~50 sites worldwide
- Use Grid services for discovery, security
- See also www.scglobal.org

Presenter mic

Presenter camera

Ambient mic (tabletop)

Audience camera

Access Grid: Argonne, others   www.mcs.anl.gov/FL/accessgrid
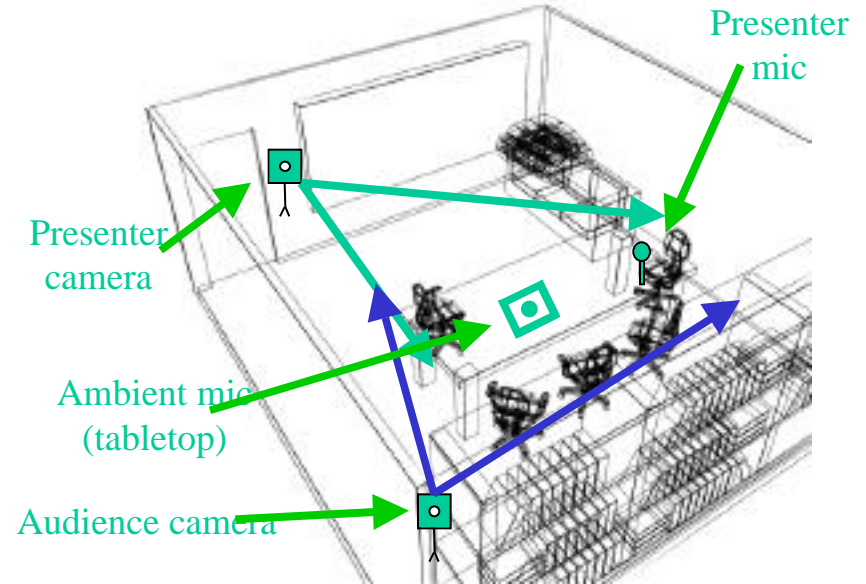
Layered Grid Architecture
(By Analogy to Internet Architecture)

"Coordinating multiple resources": ubiquitous infrastructure services, app-specific distributed services

"Sharing single resources": negotiating access, controlling use

"Talking to things": communication (Internet protocols) & security

"Controlling things locally": Access to, & control of, resources

Application

Collective

Resource

Connectivity

Fabric

Internet Protocol Architecture

Application

Transport

Internet

Link

For more info: www.globus.org/research/papers/anatomy.pdf
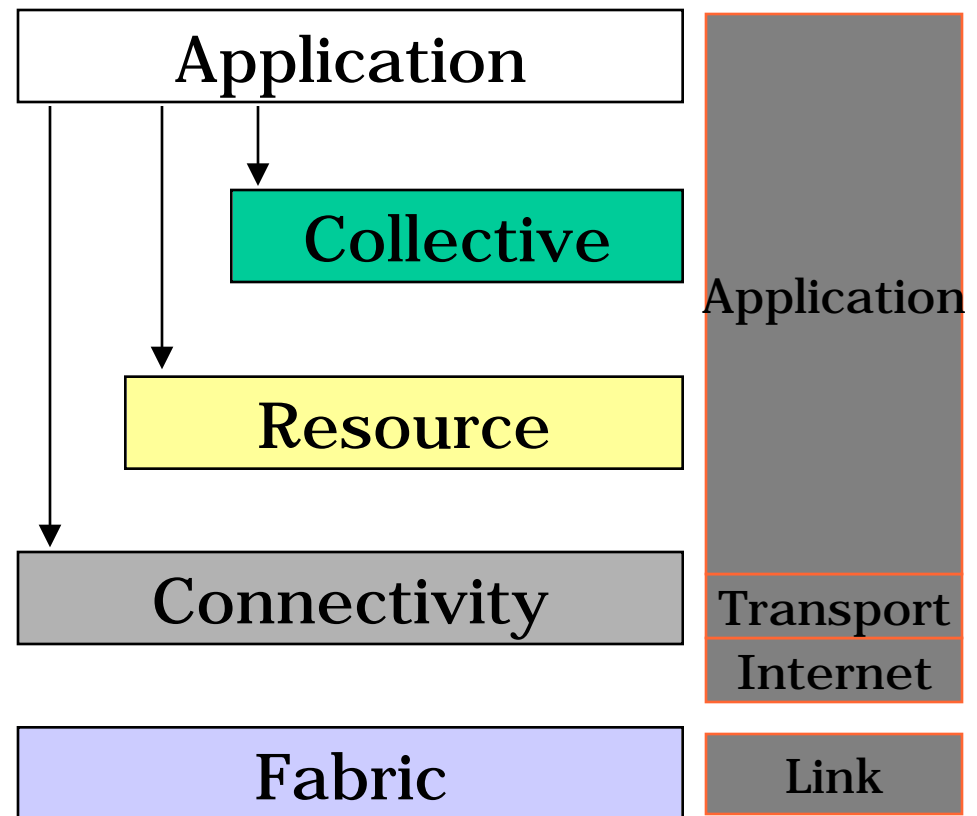
Where Are We With Architecture?

- No "official" standards exist
- But:
  - Globus Toolkit has emerged as the de facto standard for several important Connectivity, Resource, and Collective protocols
  - GGF has an architecture working group
  - Technical specifications are being developed for architecture elements: e.g., security, data, resource management, information

Globus and GriPhyN: The Focus of this Talk

- **Globus Project and Toolkit**
  - R&D project at ANL, UChicago, USC/ISI
  - Open source software and community
  - Emphasis on core protocols and services
  - Adopted by essentially all major Grid efforts
- **Grid Physics Network (GriPhyN)**
  - Data Grid R&D (ATLAS, CMS, LIGO, SDSS)
  - Defines Data Grid Reference Architecture in partnership with Particle Physics Data Grid
  - Emphasis on higher-level protocols/services

Grid Architecture and the Globus Toolkit

Grid Services Architecture (1): Fabric Layer

- Just what you would expect: the diverse mix of resources that may be shared
  - Individual computers, Condor pools, file systems, archives, metadata catalogs, networks, sensors, etc., etc.

- Few constraints on low-level technology: connectivity and resource level protocols form the "neck in the hourglass"

- Globus toolkit provides a few selected components (e.g., bandwidth broker)

Grid Services Architecture (2): Connectivity Layer Protocols & Services

- **Communication**
  - Internet protocols: IP, DNS, routing, etc.
- **Security: Grid Security Infrastructure (GSI)**
  - Uniform authentication & authorization mechanisms in multi-institutional setting
  - Single sign-on, delegation, identity mapping
  - Public key technology, SSL, X.509, GSS-API (several Internet drafts document extensions)
  - Supporting infrastructure: Certificate Authorities, key management, etc.

GSI in Action: "Create Processes at A and B that Communicate & Access Files at C"

User

Single sign-on via "grid-id" & generation of proxy cred.

Or: retrieval of proxy cred. from online repository

User Proxy

Proxy credential

Remote process creation requests*

GSI-enabled GRAM server

Authorize
Map to local id
Create process
Generate credentials

Ditto

GSI-enabled GRAM server

Site A (Kerberos)

Computer

Process

Local id

Kerberos ticket

Restricted proxy

Site B (Unix)

Computer

Process

Local id

Restricted proxy

Communication*

Remote file access request*

Site C (Kerberos)

Storage system

GSI-enabled FTP server

Authorize
Map to local id
Access file

* With mutual authentication

GSI Futures

- Scalability in numbers of users & resources
  - Credential management
  - Online credential repositories ("MyProxy")
  - Account management
- Authorization
  - Policy languages
  - Community authorization
- Protection against compromised resources
  - Restricted delegation, smartcards

Community Authorization
(Prototype shown August 2001)

User

1. CAS request, with resource names and operations

2. CAS reply, with [capability] and resource CA info

CAS

Does the collective policy authorize this request for this user?

user/group membership

resource/collective membership

collective policy information

Resource

3. Resource request, authenticated with [capability]

Is this request authorized by the capability?

local policy information

4. Resource reply

Is this request authorized for the CAS?

Laura Pearlman, Steve Tuecke, Von Welch, others

Grid Services Architecture (3): Resource Layer Protocols & Services

- Resource management: GRAM
  - Remote allocation, reservation, monitoring, control of [compute] resources
- Data access: GridFTP
  - High-performance data access & transport
- Information: MDS (GRRP, GRIP)
  - Access to structure & state information
- & others emerging: catalog access, code repository access, accounting, ...
- All integrated with GSI

GRAM Resource Management Protocol

- **Grid Resource Allocation & Management**
  - Allocation, monitoring, control of computations
- **Simple HTTP-based RPC**
  - Job request:
    - > Returns a "job contact": Opaque string that can be passed between clients, for access to job
  - Job cancel, Job status, Job signal
  - Event notification (callbacks) for state changes
    - > Pending, active, done, failed, suspended
- **Servers for most schedulers; C and Java APIs**

Resource Management Futures: GRAM-2 (planned for late 2001)

- Advance reservations
  - As prototyped in GARA in previous 2 years
- Multiple resource types
  - Manage anything: storage, networks, etc., etc.
- Recoverable requests, timeout, etc.
  - Build on early work with Condor group
- Use of SOAP (RPC using HTTP + XML)
  - First step towards Web Services
- Policy evaluation points for restricted proxies

Karl Czajkowski, Steve Tuecke, others

Data Access & Transfer

- GridFTP: extended version of popular FTP protocol for Grid data access and transfer
- Secure, efficient, reliable, flexible, extensible, parallel, concurrent, e.g.:
  - Third-party data transfers, partial file transfers
  - Parallelism, striping (e.g., on PVFS)
  - Reliable, recoverable data transfers
- Reference implementations
  - Existing clients and servers: wuftpd, nicftp
  - Flexible, extensible libraries

Bill Allcock, Joe Bester, John Bresnahan, Steve Tuecke, others

Grid Services Architecture (4): Collective Layer Protocols & Services

- **Index servers aka metadirectory services**
  - Custom views on dynamic resource collections assembled by a community
- **Resource brokers (e.g., Condor Matchmaker)**
  - Resource discovery and allocation
- **Replica management and replica selection**
  - Optimize aggregate data access performance
- **Co-reservation and co-allocation services**
  - End-to-end performance
- **Etc., etc.**

VO–A

VO–B
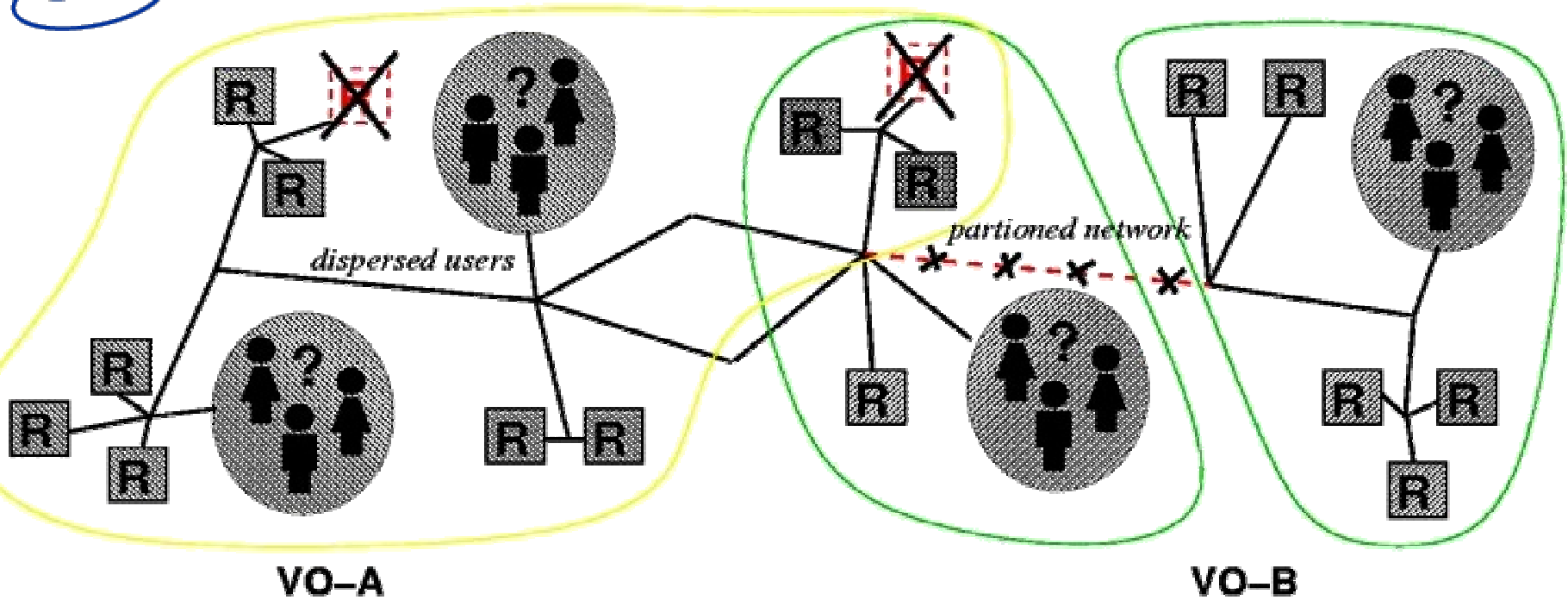
- Large numbers of distributed "sensors" with different properties

- Need for different "views" of this information, depending on community membership, security constraints, intended purpose, sensor type

registration messages

fault-partition

replicated directories

divergent directories

VO-A

VO-B
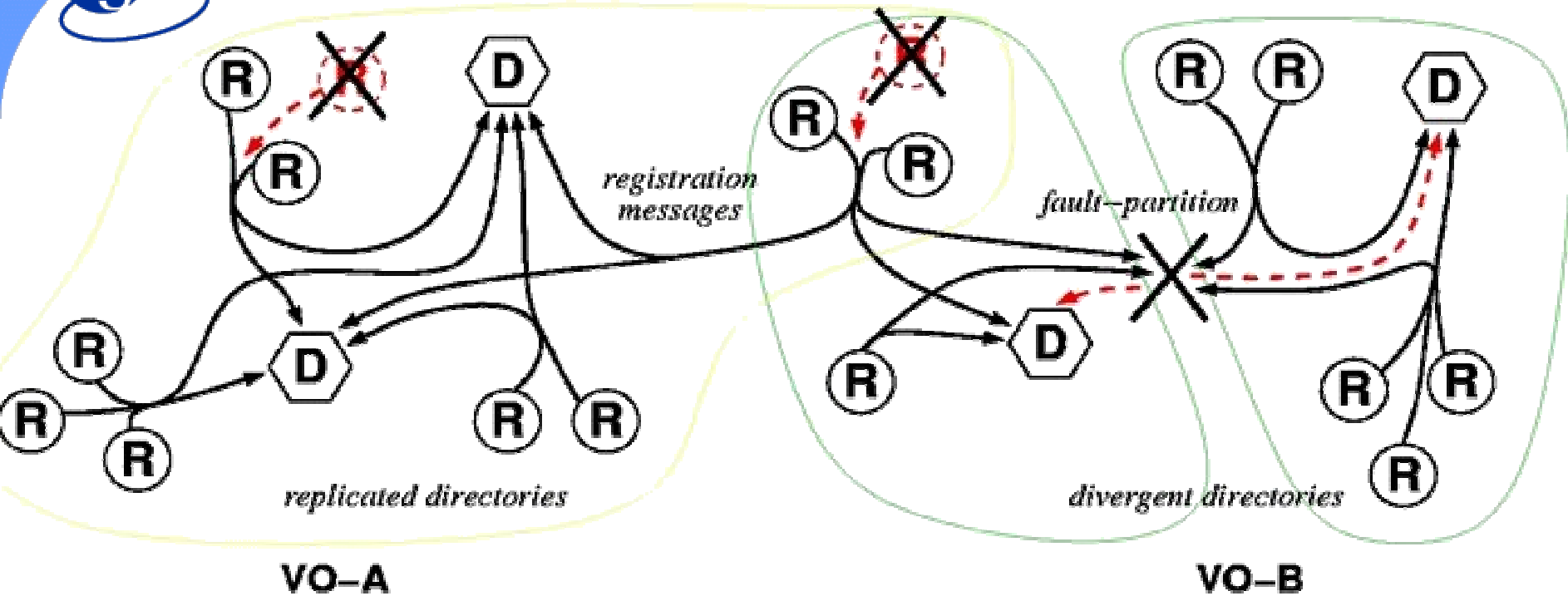
Registration & enquiry protocols, information models, query languages

– Provides standard interfaces to sensors

– Supports different "directory" structures supporting various discovery/access strategies

Karl Czajkowski, Steve Fitzgerald, others

the globus project
www.globus.org

ASCI DISCOM
Condor-G
Nimrod-G
Poznan
U. Lecce

Broker

RSL
specialization

RSL

Application

Queries
& Info

Information
Service

Ground RSL

Co-allocator

DUROC
MPICH-G2

Simple ground RSL

Local
resource
managers

GRAM

GRAM

GRAM

LSF

Condor

NQE
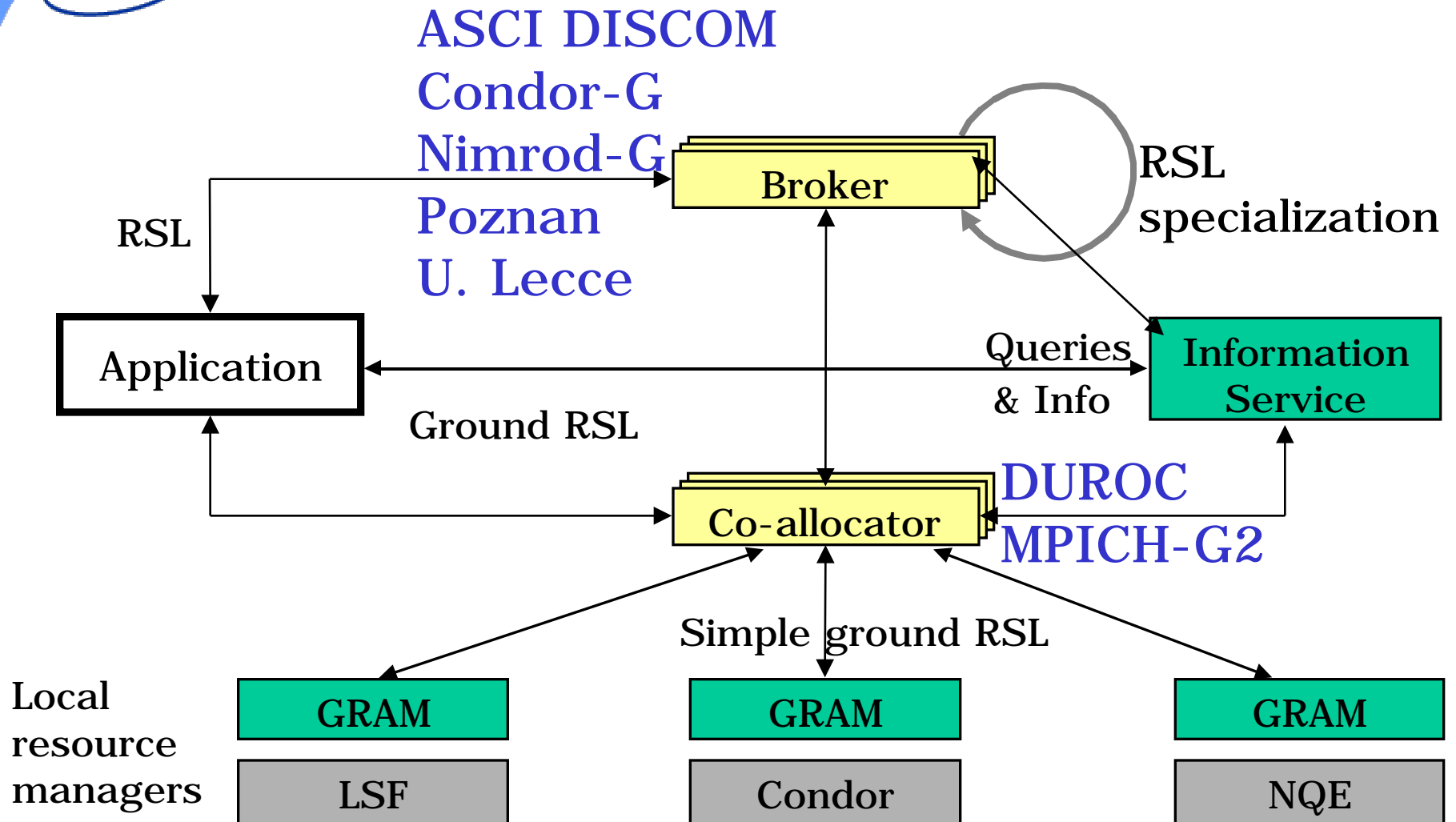
GriPhyN Project

The Virtual Data Concept

"[a virtual data grid enables] the definition and delivery of a potentially unlimited virtual space of data products derived from other data. In this virtual space, requests can be satisfied via direct retrieval of materialized products and/or computation, with local and global resource management, policy, and security constraints determining the strategy used."

Virtual Data in Action

- Data request may
  - Access local data
  - Compute locally
  - Compute remotely
  - Access remote data
- Scheduling & execution subject to local & global policies

Major Archive Facilities

Network caches & regional centers

Local sites

?
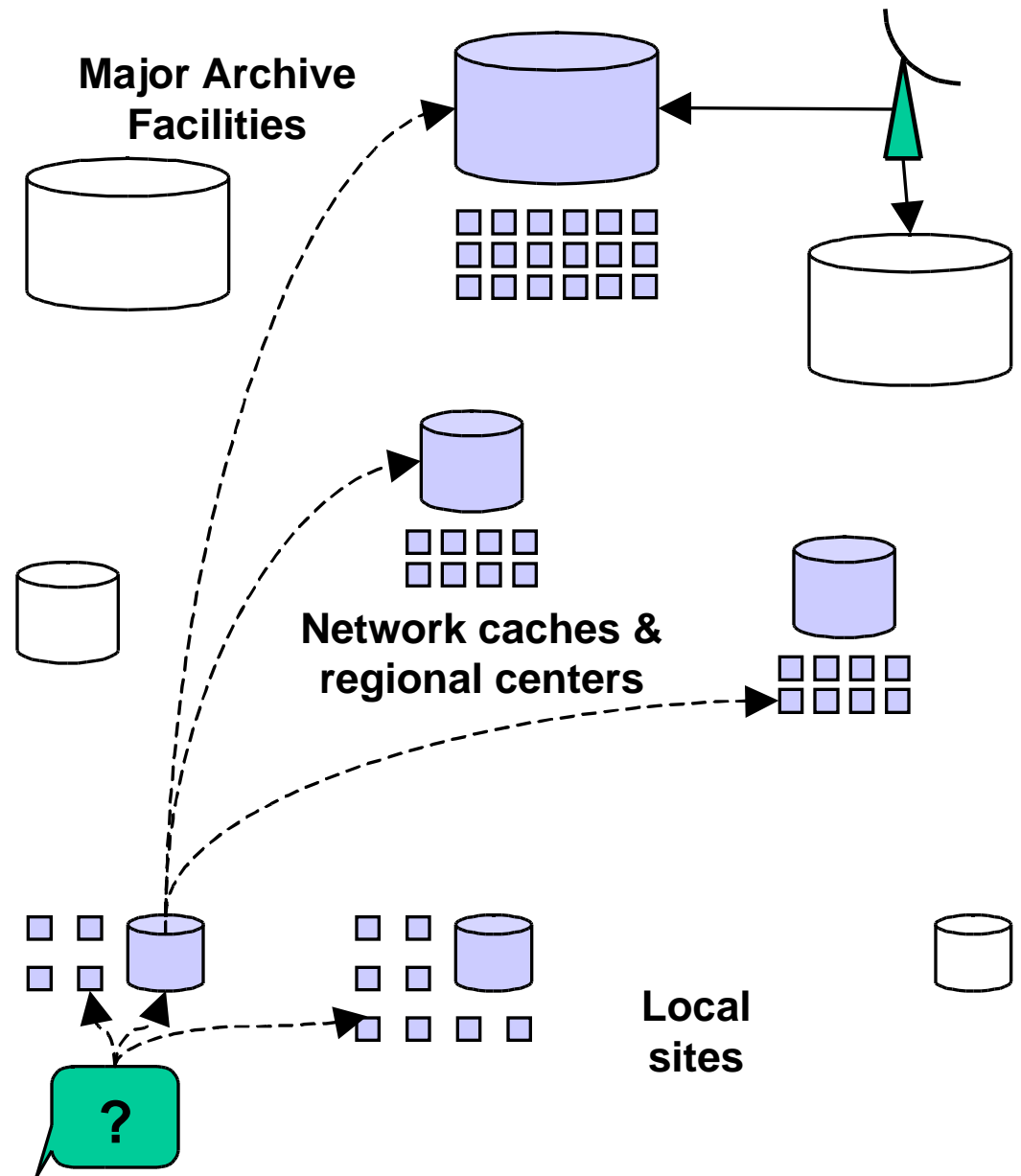
Grid Physics Network (GriPhyN)

Enabling R&D for advanced data grid systems, focusing in particular on Virtual Data concept
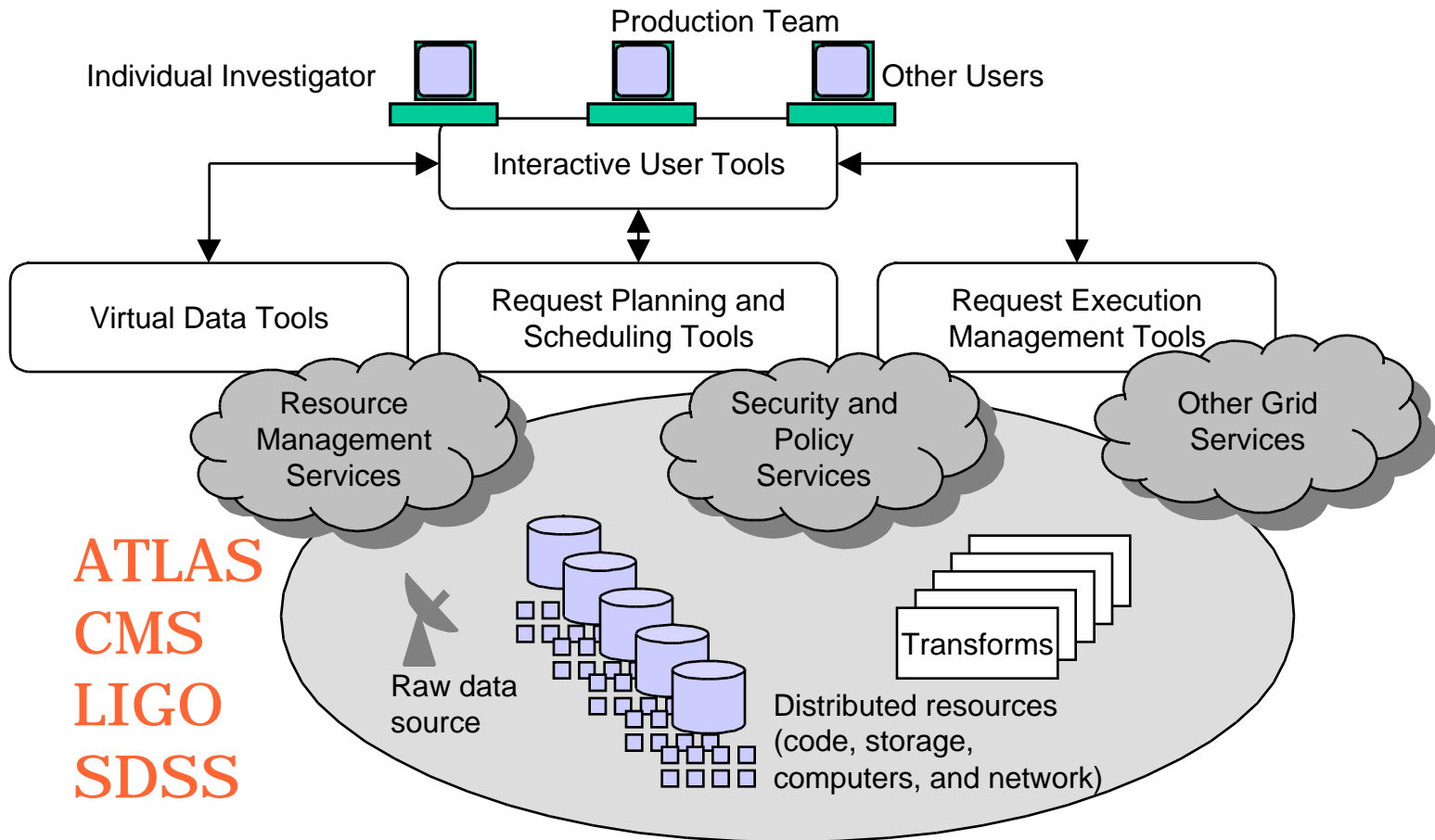
GriPhyN Status, August 2001

- Data Grid Reference Architecture defined
  - v1: core services (Feb 2001)
  - v2: request planning/mgmt, catalogs (RSN)
- Progress on ATLAS, CMS, LIGO
  - Requirements statements developed
  - Testbeds and experiments proceeding
- Progress on technology
  - DAGMAN request management
  - Catalogs, security, policy
  - Virtual Data Toolkit v1.0 out soon

GriPhyN/PPDG Data Grid Architecture

Ewa Deelman, Mike Wilde
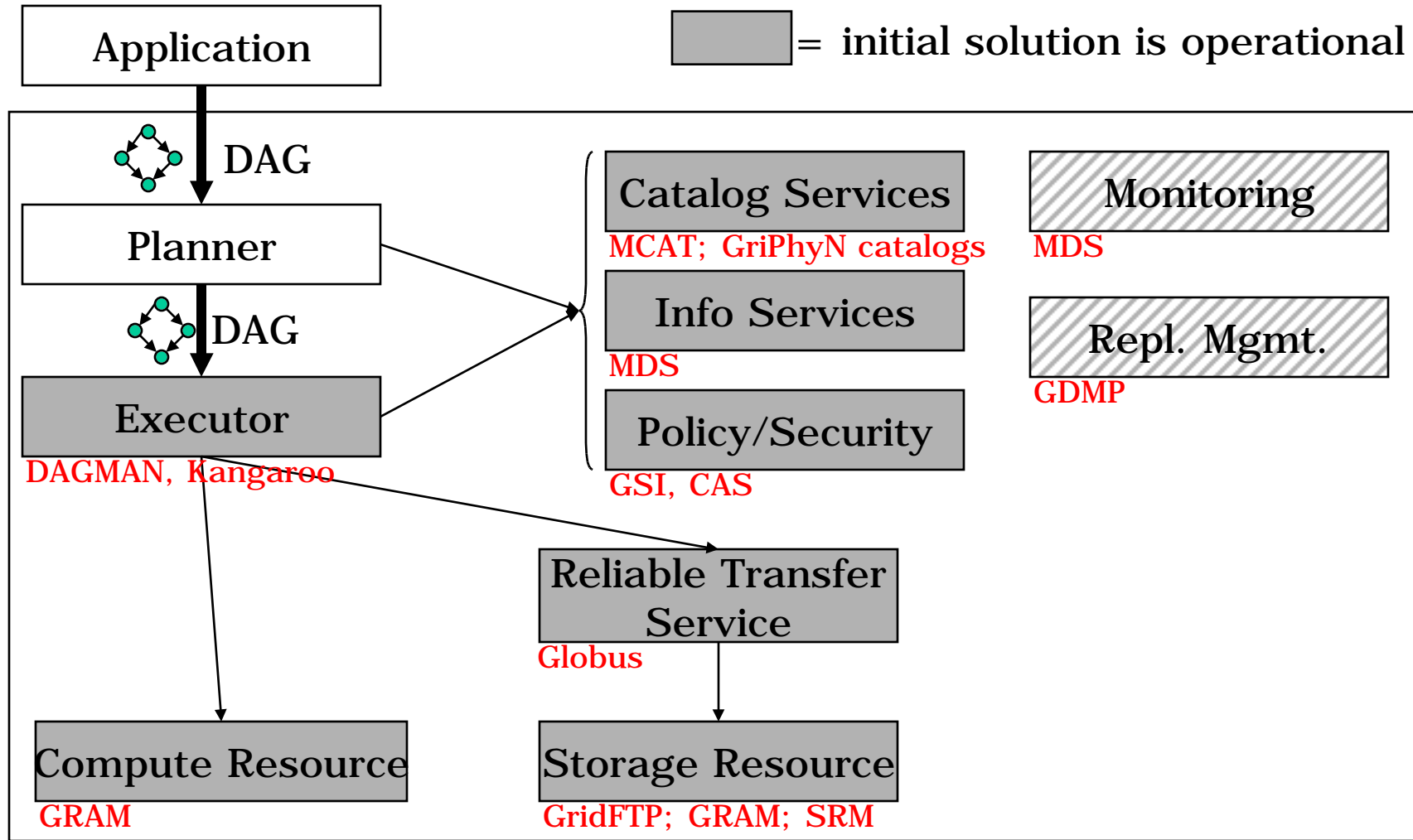
Example Data Grid Technologies

- Replica catalog
  - Centralized in place, decentralized in design
- Transformation, virtual data catalogs
  - Various prototypes
- DAGMAN (M. Livny, U.Wisconsin)
  - Reliably execute a directed acyclic graph of compute and data movement tasks on Grid

Ann Chervenak, Ewa Deelman, Miron Livny, Mike Wilde, others

Catalog Architecture

Transparency wrt materialization

Derived Metadata Catalog

App-specific-attr	id ...
...	i2,i10
...	
...	

id

Derived Data Catalog

Update upon materialization

Id	Trans	Param	Name	...
i1	F		X	F.X ...
i2	F		Y	F.Y ...
i10	G	Y	P	G(P).Y ...

Trans. name

Transformation Catalog

Trans	Prog	Cost	...
F	URL:f	10	...
G	URL:g	20	...

Transparency wrt location

Name	LObjN	...	...
X	logO1	...	...
Y	logO2	...	...
F.X	logO3	...	...
G(1).Y	logO4	...	...

Object Name

GCMS

Replica Catalog

LCN	PFNs	...
logC1	URL1	
logC2	URL2 URL3	
logC3	URL4	
logC4	URL5 URL6	

URLs for program location

Program storage

URLs for physical file location

Physical file storage

the globus project
www.globus.org

Master Condor job running at Caltech

Caltech workstation

2) Launch secondary job on WI pool; input files via Globus GASS

Secondary Condor job on WI pool

5) Secondary reports complete to master

6) Master starts reconstruction jobs via Globus jobmanager on cluster

3) 100 Monte Carlo jobs on Wisconsin Condor pool

9) Reconstruction job reports complete to master

4) 100 data files transferred via GridFTP, ~ 1 GB each

7) GridFTP fetches data from UniTree

NCSA Linux cluster

8) Processed objectivity database stored to UniTree

NCSA UniTree - GridFTP-enabled FTP server

Scott Koranda, Miron Livny, others

Trace of a Condor-G Physics Run

Pre / Simulation Jobs /
Post (UW Condor)

ooDigis at NCSA

ooHits at NCSA

Delay due to
script error
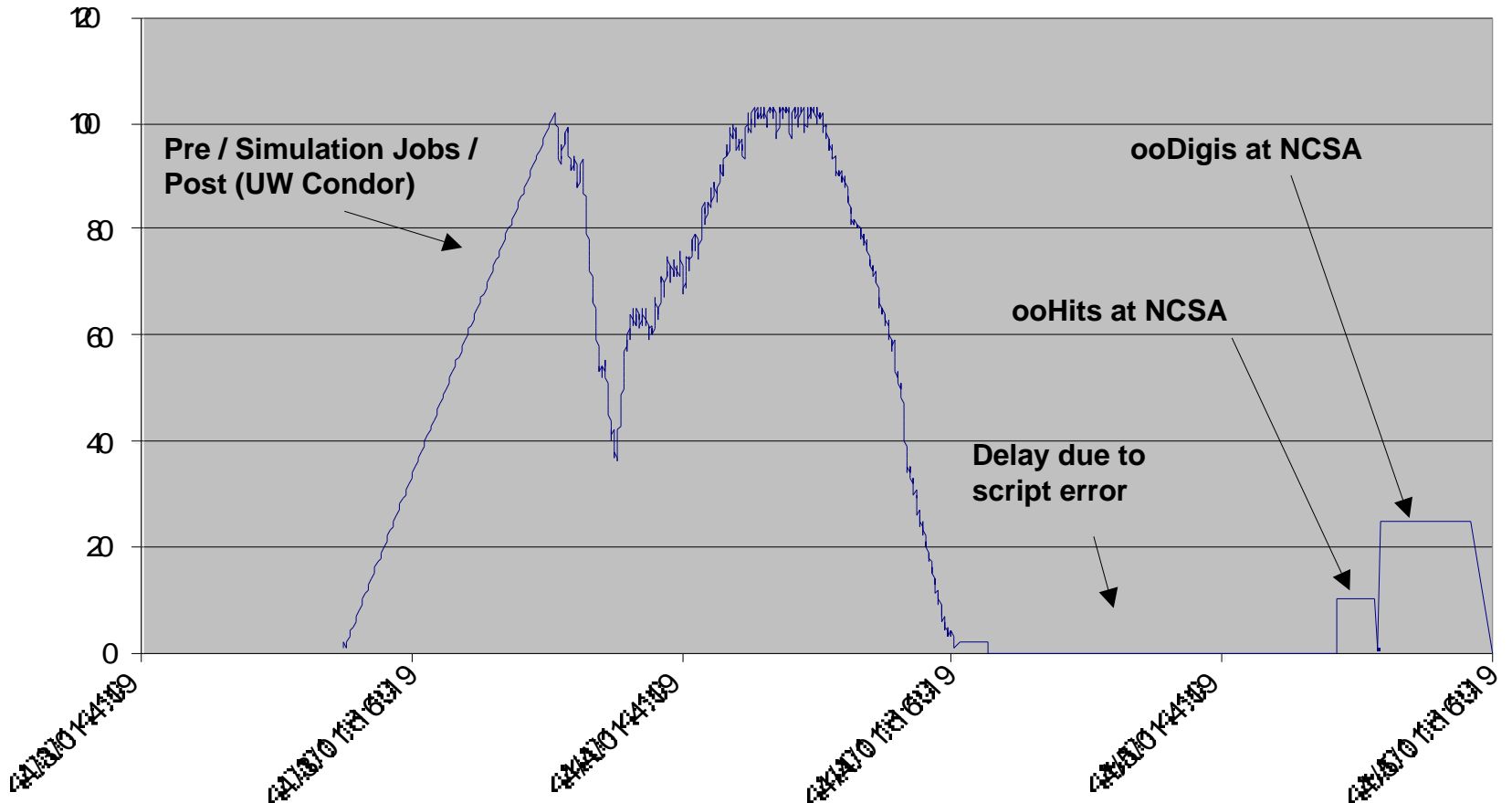
International Cooperation

International Cooperation

- Critical for long-term success
  - We hang together or hang separately
  - Scope of effort demands it
  - International science demands it
- Too early to tell if we will succeed
  - Some positives: e.g., use of Globus, GDMP
  - Still too few mutual interdependencies
- International Data Grid Coordination Cmte
- International "laboratories" seem key

iVDGL

- **International Virtual-Data Grid Laboratory**
  - A place to conduct Data Grid tests at scale
  - Concrete manifestation of world-wide grid activity
  - Continuing activity that will drive Grid awareness
- **Scale of effort**
  - For national, intl scale Data Grid tests, operations
  - Computation & data intensive computing
- **Who**
  - Initially US-UK-Italy-EU; Japan, Australia
  - & Russia, China, Pakistan, India, South America?
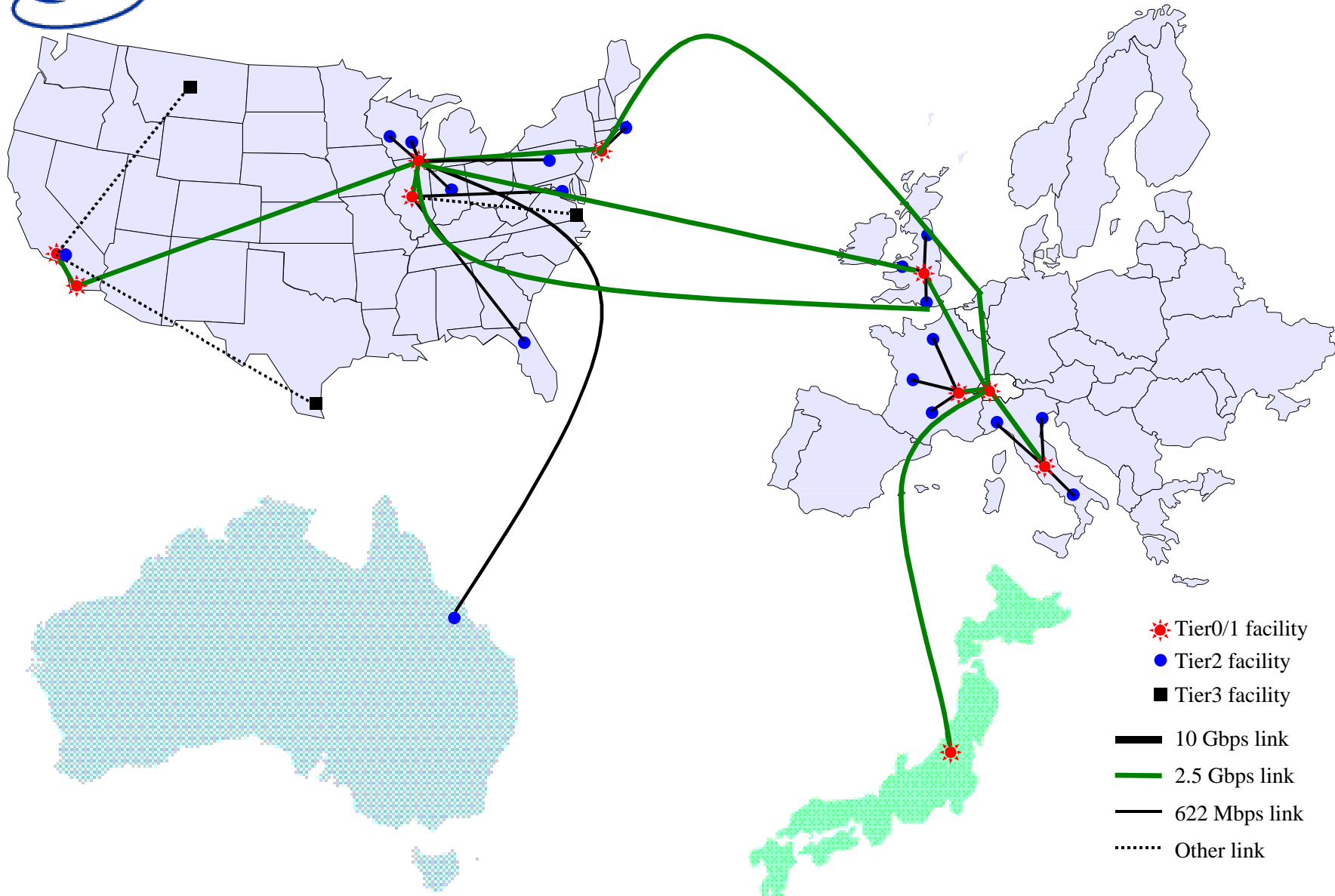 - StarLight and other international networks vital

U.S. Co-PIs: Avery, Foster, Gardner, Newman, Szalay

iVDGL Map Circa 2003-2004

Tier0/1 facility

Tier2 facility

Tier3 facility

10 Gbps link

2.5 Gbps link

622 Mbps link

Other link

Commodity Technologies

And What's This Got To Do With ...

- CORBA?
  - Grid-enabled CORBA underway
- Java, Jini, Jxta?
  - Java CoG Kit. Jini, Jxta: future uncertain
- Web Services, .NET, J2EE?
  - A significant focus within Globus project
  - GRAM-2 will use SOAP and WSDL
  - Q: What can Grid offer to Web services?
- Next revolutionary technology of the month?
  - They'll need Grid technologies too

Summary

- "Grids": Resource sharing & problem solving in dynamic virtual organizations
  - Many projects now working to develop, deploy, apply relevant technologies
- Common protocols and services are critical
  - Globus Toolkit a source of protocol and API definitions, reference implementations
- Rapid progress on definition, implementation, and application of Data Grid architecture
  - Harmonizing U.S. and E.U. efforts important