HEP Computing and GRID
(BABAR, TEVATRON and LHC)

Matthias Kasemann
FNAL and CERN

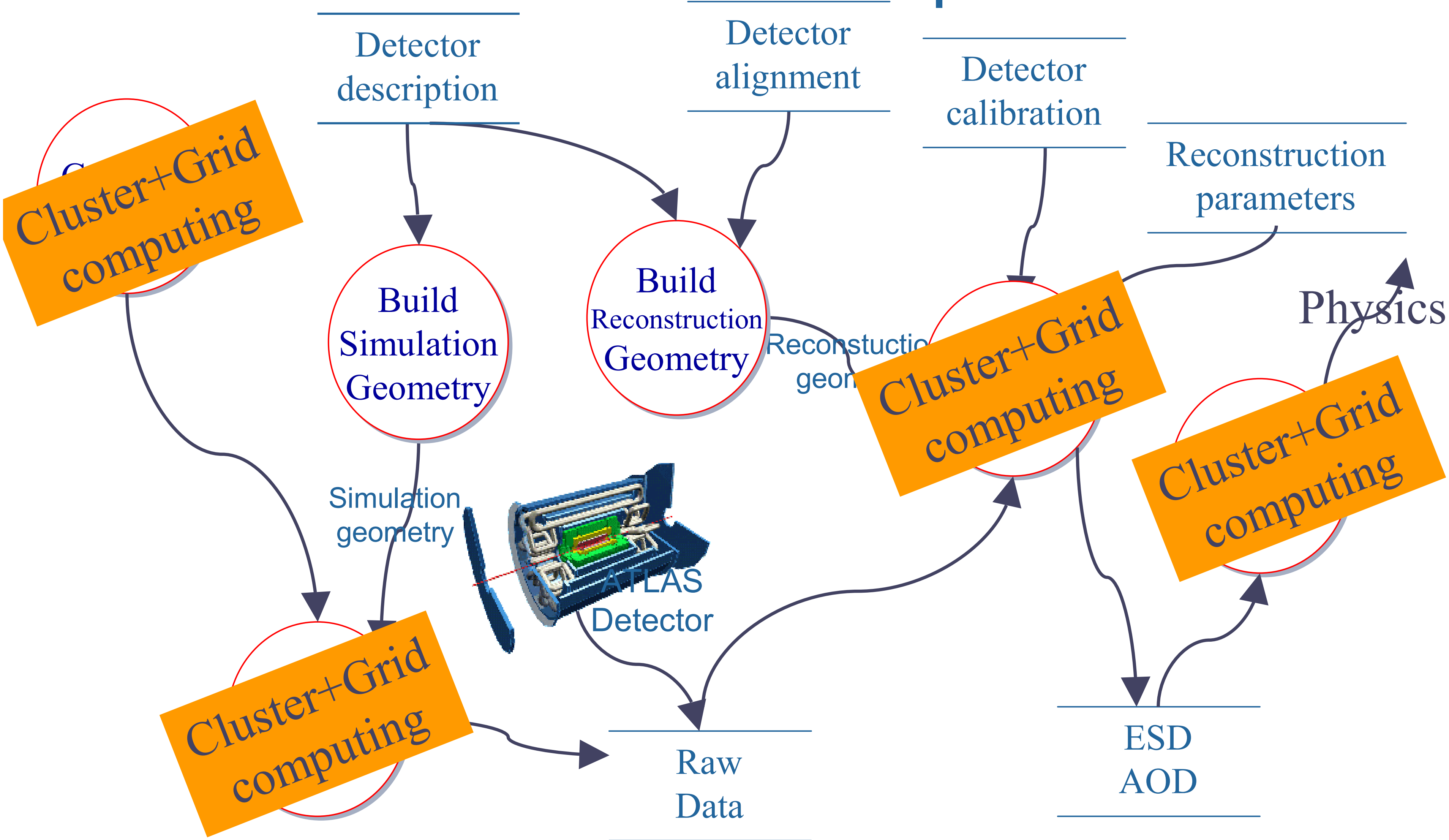
LCCWS2002, October 21, 2002

Goal for HEP analysis

- Example:
  ATLAS guiding principles
  (true for all LHC experiments):

  - Every physicist in ATLAS must have the best possible **access to the data** necessary for the analysis, **irrespective of his/her location**.
  - The access to the data should be **transparent and efficient**.
  - We should **profit from resources** (money, manpower and hardware) **available in the different countries**.
  - We should **benefit from the outcome of the Grid projects**.

HEP analysis chain:
common to most experiments



BaBar computing: truly distributed

- Distributed Computing and Analysis:
  - TierA sites in SLAC and LYON,
    - both have full set of analysis data in objectivity
    - SLAC: site for first reconstruction
  - TierA site at RAL for ROOT based analysis data distribution
  - TierA site in Padova ready for data reprocessing (initially)
  - MC production distributed over 15 sides (incl. Lyon), stable
  - people have free and transparent access to analyze in Lyon and SLAC
- Data copies at TierA sites <u>improve access performance</u>
- Questions to assess:
  - manpower is a serious issue to solve and maintain problems and maintain two analysis branches!
  - Re-evaluate (streamline?) the data formats used for analysis.
  - how many copies of the data do they need on disk for performance?
  - how to best use the 4 TierA sites?

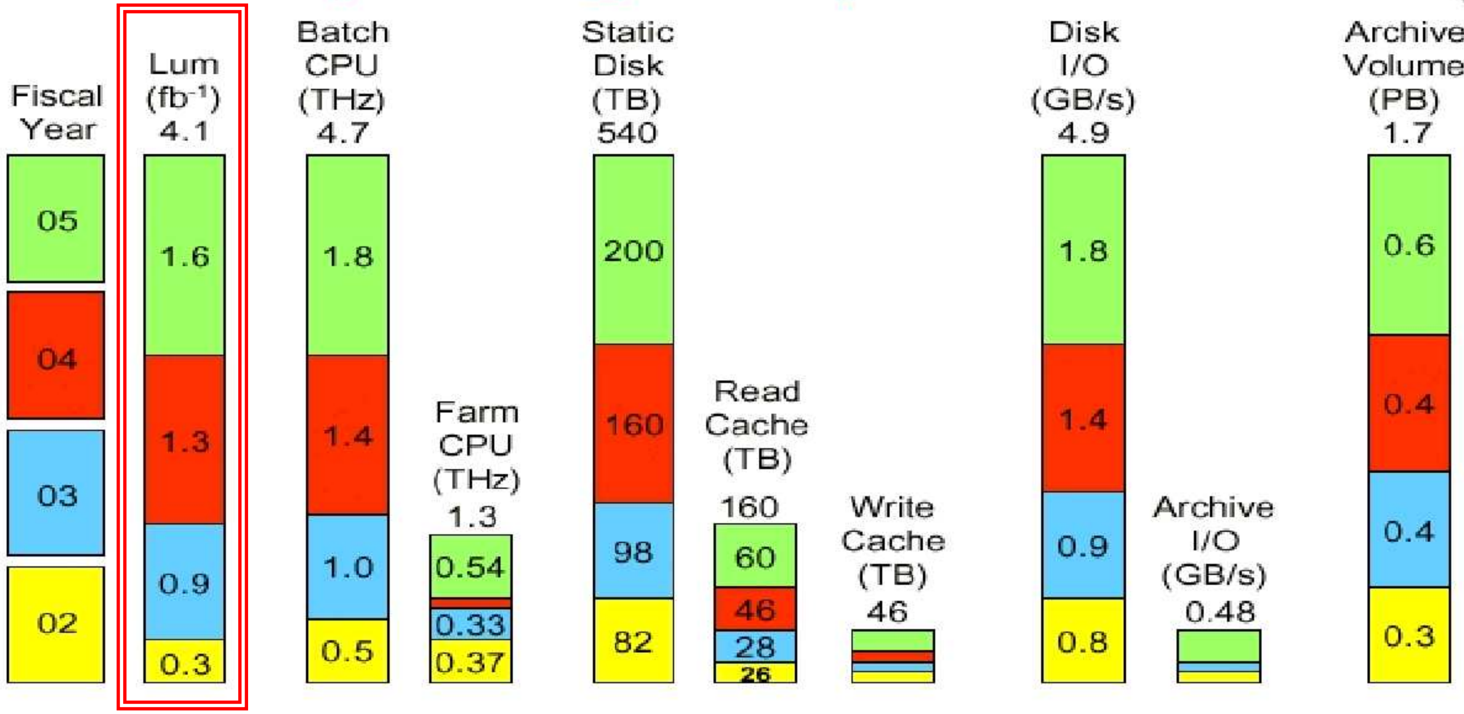
Scale of CDF & D0 Computing

- Scope of Run2a computing: **per experiment:**
  - ~ 12-15 MB/sec raw data rate
  - ~ 12 MB/sec into reconstruction farms
  - ~ 4 - 16 MB/sec out of reconstruction farms
  - ~ 150 MB/sec – total offline capacity for data movement

  - Raw data ~ 150 TB /yr / experiment
  - Total datasets up to 500 TB /yr /experiment
  - Central disk storage now ~ > 150 TB (growing!)

- Computing hardware and infrastructure **cost**:
  - Initial investment (1997-2002) ~ $15M / experiment
  - Operating and upgrades: ~$3M / yr / experiment

CDF
Computing Requirements



Requirements set by goal:

200 simultaneous users to analyze secondary data set (10^7 evts) in a day

Need ~700 TB of disk and ~5 THz of CPU by end of FY'05:

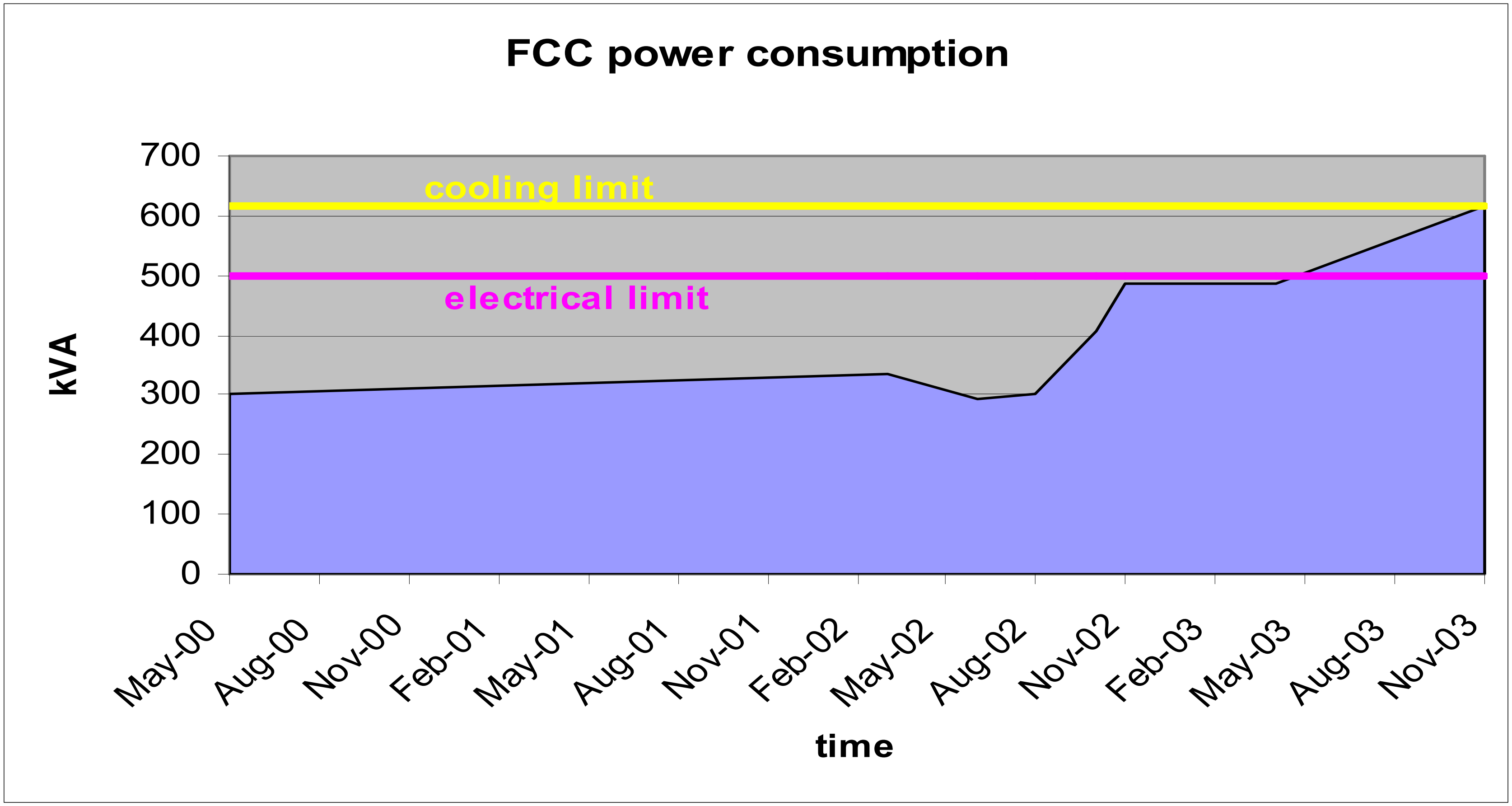
→ need lots of disk→ need cheap disk → IDE Raid

→ need lots of CPU→ commodity CPU → dual Intel/AMD

Mark Neubauer/MIT *ACAT'02*


Fermilab
Michigan St.
Michigan
Indiana
Kansas
Oklahoma
Arizona
Texas
Boston
Lancaster
Imperial
NIKHEF
Prague
Lyon
Wuppertal
Munich
SAM data transfer
Monte Carlo files
Data Grid status in March 2002
5 remote Monte Carlo generation sites + more coming
15 SAM stations for remote analysis + more coming

Setting up fabrics: FCC power and cooling



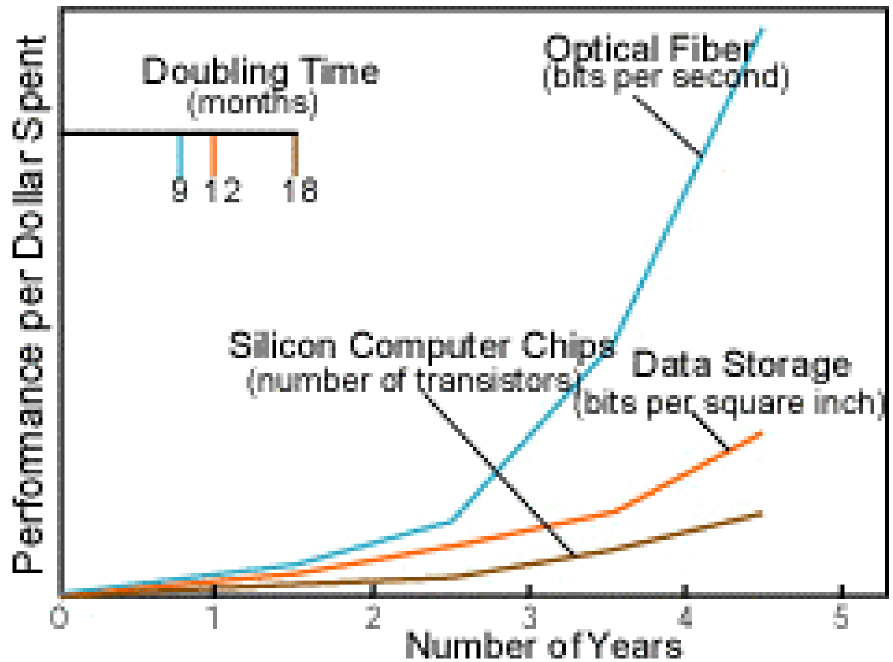
- **We have some work in front of us!!**

Why compute on a Grid?

- Network vs. computer performance
  (observed + forecasted)
  - Computer speed doubles every 18 months
  - Network speed doubles every 9 months
  - Difference = order of magnitude per 5 years

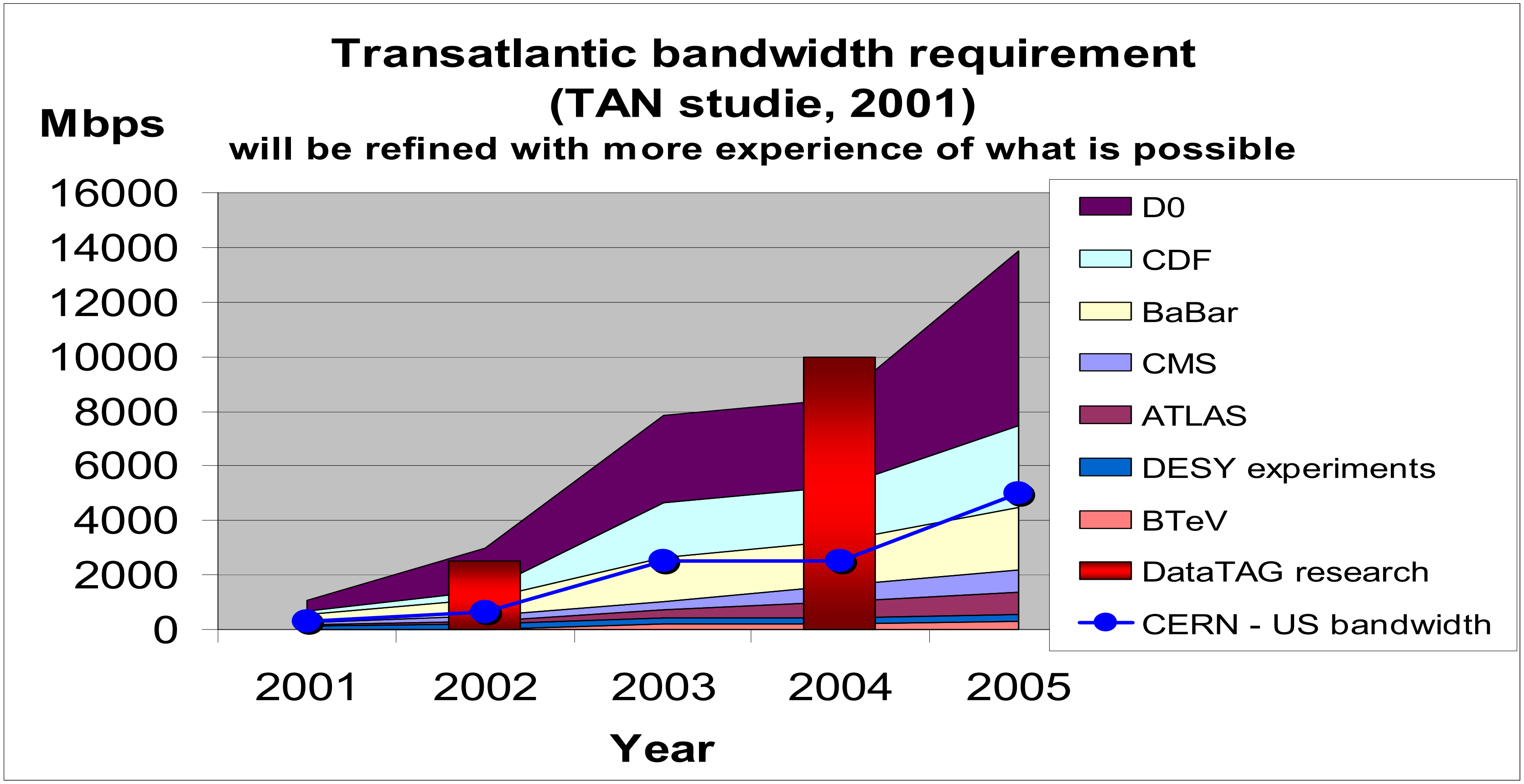
- 1986 to 2000
  - Computers: x 500
  - Networks: x 340,000

- 2001 to 2010
  - Computers: x 60
  - Networks: x 4000



Moore's Law vs. storage improvements vs. optical improvements. Graph from **Scientific American** (Jan-2001) by Cleo Vilett, source Vined Khoslan, Kleiner, Caufield and Perkins.

HEP Networking needs: e.g. EU-US



Installed bandwidth, Maximum Link occupancy of 50% assumed
See: ***http:/gate.hep.anl.gov/lprice/TAN***

Project: DataTAG 2.5 Gbps Research Link in Summer 2002;
10 Gbps Research Link in ~2003 or Early 2004

HENP Major Links: Bandwidth Roadmap (Scenario) in Gbps

Year	Production	Experimental	Remarks
2001	0.155	0.622-2.5	SONET/SDH
2002	0.622	2.5	SONET/SDH DWDM; GigE Integ.
2003	2.5	10	DWDM; 1 + 10 GigE Integration
2005	10	2-4 X 10	λ Switch; λ Provisioning
2007	2-4 X 10	~10 X 10; 40 Gbps	1st Gen. λ Grids
2009	~10 X 10 or 1-2 X 40	~5 X 40 or ~20-50 X 10	40 Gbps λ Switching
2011	~5 X 40 or ~20 X 10	~25 X 40 or ~100 X 10	2nd Gen λ Grids Terabit Networks
2013	~Terabit	~MultiTerabit	~Fill One Fiber

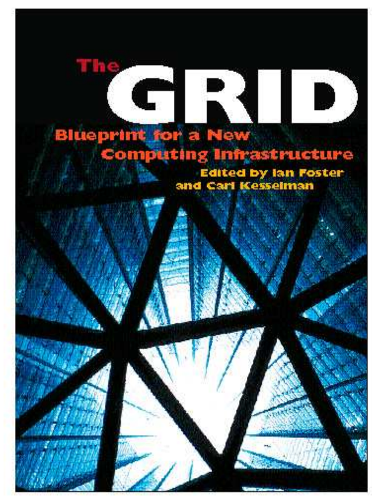
From: ICFA SCIC, H. Newman, Feb, 2002

The Grid vision of computing

- Flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resource
  - From “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”
- Enable communities (“virtual organizations”) to share geographically distributed resources as they pursue common goals -- *assuming the absence* of…
  - central location,
  - central control,
  - omniscience,
  - existing trust relationships.

Globus Grid Computing—the Next Internet
by John Roy/Steve Milunovich

The Internet was first a network and is now a communications platform. The next evolutionary step could be to a platform for distributed computing. This ability to manage applications and share data over the network is called “grid computing.”



1999

Grid Projects: Current Status

- **Several major Grid projects** in scientific & technical computing/research & education set up.
- **Considerable consensus** on key concepts and technologies
  - Open source Globus Toolkit™ is a de facto standard for major protocols & services
  - Far from complete or perfect, but out there, evolving rapidly, and large tool/user base
- Industrial interest emerging rapidly
- Concepts map very well with HEP style of collaborating
  - Globally spread participation in experiments
  - Many funding sources
  - Widely spread expertise
  - 'transparent' access to data

Improve scientific result by

- **easy data access**
- **broad participation**

Grid Technology Area
Leveraging Grid R&D Projects

Many national, regional Grid projects -- GridPP(UK), INFN-grid(I), NorduGrid, Dutch Grid, …

US projects

European projects

Grid Technology Area
Leveraging Grid R&D Projects

- significant R&D funding for Grid middleware
- risk of divergence
  → requires substantial coordination effort
  and interfacing work to HEP effort
- global grids need standards
- useful grids need stability
- hard to do this in the current state of maturity
  - Extensive testing and
    prototyping program required

We (HEP) feel we have no choice than to participate!!

How do we solve problems?
Q: is this valid for HEP?

- Communities committed to common goals
  - Virtual organizations map well to HEP collaborations
  - Teams with heterogeneous members & capabilities
- Distributed geographically and politically
  - No location/organization possesses all required skills and resources
- Adapt as a function of the situation
  - Adjust membership, reallocate responsibilities, renegotiate resources
- Online negotiation of access to services (dynamically):
  - who, what, why, when, how
- Establishment of applications and systems able to deliver multiple qualities of service
- Autonomic management of infrastructure elements
- Open, extensible, evolvable infrastructure

What is a "virtual" dataset?

- Tracking the derivation of experiment data with high fidelity
- Transparency with respect to location and materialization
  - Track all data assets
  - Accurately record how they were derived
  - Encapsulate the transformations that produce new data objects

- Resulting data access possibilities are:
  1. **Access data at storage site**
  2. **Copy dataset to requesting site**
  3. **Recreate dataset at requesting site**

Computing for the LHC experiments

A new Project has been setup at CERN:
the **LHC Grid Computing Project (LCG)**

The first phase of the project: 2002-2005

- preparing the prototype computing environment, including
  - support for applications – libraries, tools, frameworks, common developments, …..
  - global grid computing service
- Shared funding by Regional Centers, CERN, Contributions
- Grid software developments by national and regional Grid projects

Phase 2: 2005-2007
construction and operation of the initial LHC Computing Service

LCG - Fundamental goal:

The experiments have to get the best, most reliable and accurate physics results from the data provided by their detectors

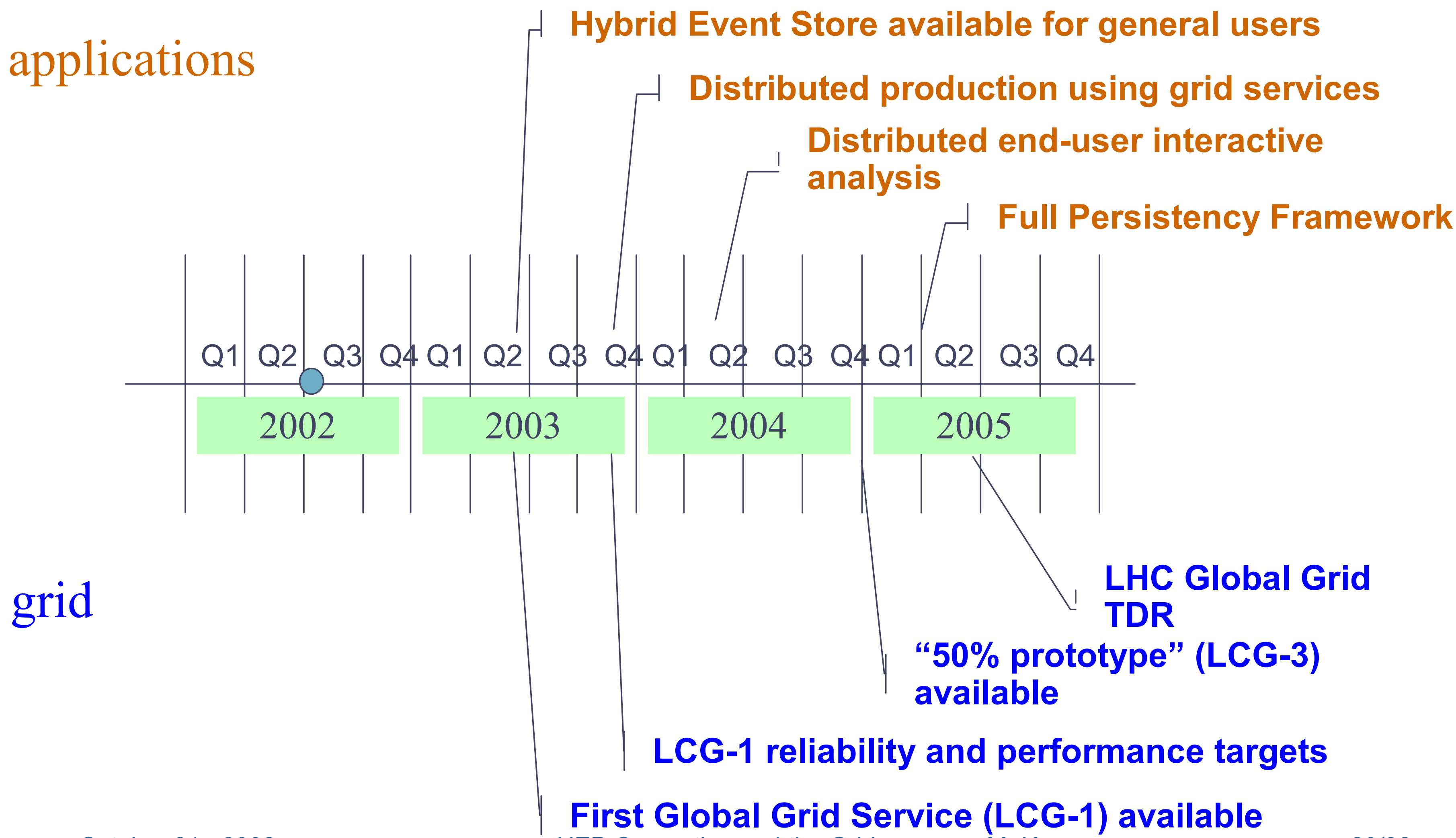
Their computing projects are fundamental to the achievement of this goal

The LCG project at CERN was set up to help them all in this task

Corollary

Success of LCG is fundamental to success of LHC Computing

Proposed LCG Phase 1 L1 Milestones (more as project evolves)



LCG:
Steps towards LHC computing

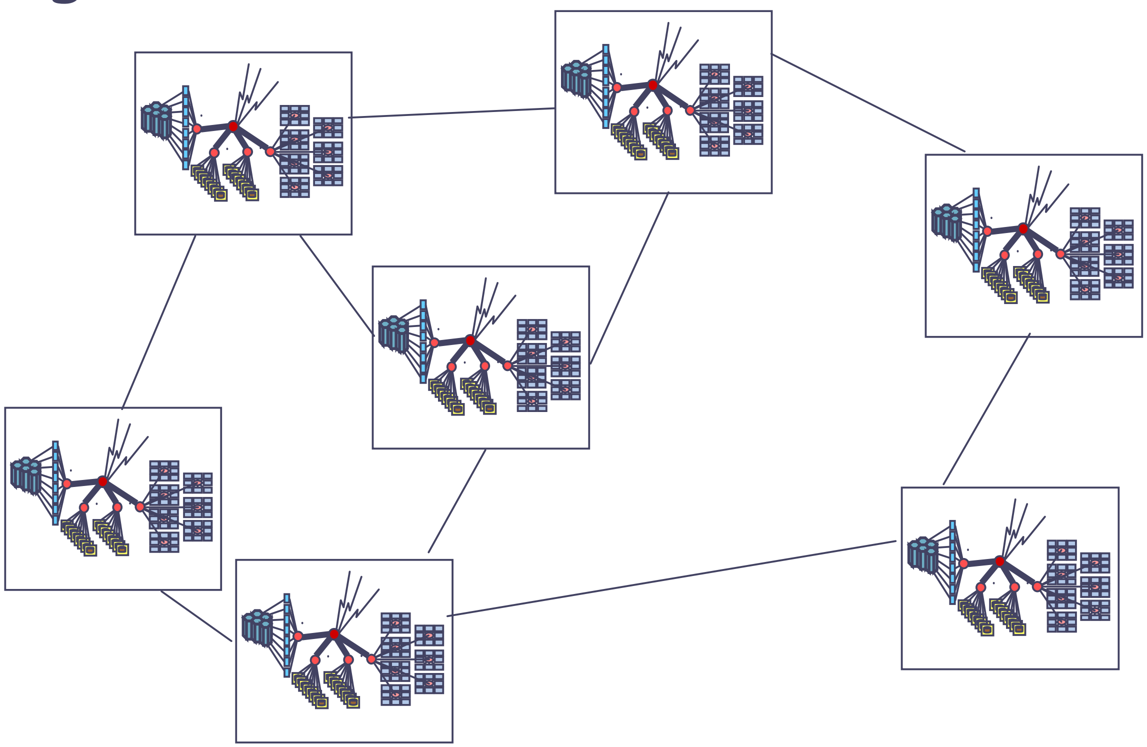
- Prepare and deploy the LHC Computing Environment
  - Not just another GRID technology project
  - Applications - provide the common components, tools and infrastructure for the physics application software
  - Computing system – fabric, grid, global analysis system
  - Deployment – foster collaboration and coherence
- Validate the software by participating in Data Challenges using the progressively more complex Grid Prototype
  - Phase 1 - 50% model production grid in 2004 (2005)
- Produce a TDR for full system to be built in Phase 2
  - Software performance impacts on size and cost of production facility
  - Analysis models impact on exploitation of production grid
- Maintain opportunities for reuse of deliverables outside the LHC experimental programme

The "virtual" LHC Computing Center

- The aim is to build
  - a general computing service -
  - for a very large user population -
  - of independently-minded scientists -
  - using a large number of independently managed sites!
- This is NOT a collection of sites providing pre-defined services
  - it is the user's job that defines the service
  - it is current research interests that define the workload
  - it is the workload that defines the data distribution
- DEMAND - Unpredictable & Chaotic
- But the SERVICE had better be Available & Reliable

Building a Grid for LHC

Collaborating Computer Centers



Building a Grid for LHC

→The "virtual" LHC Computing Center

Collaborating Computer Centers

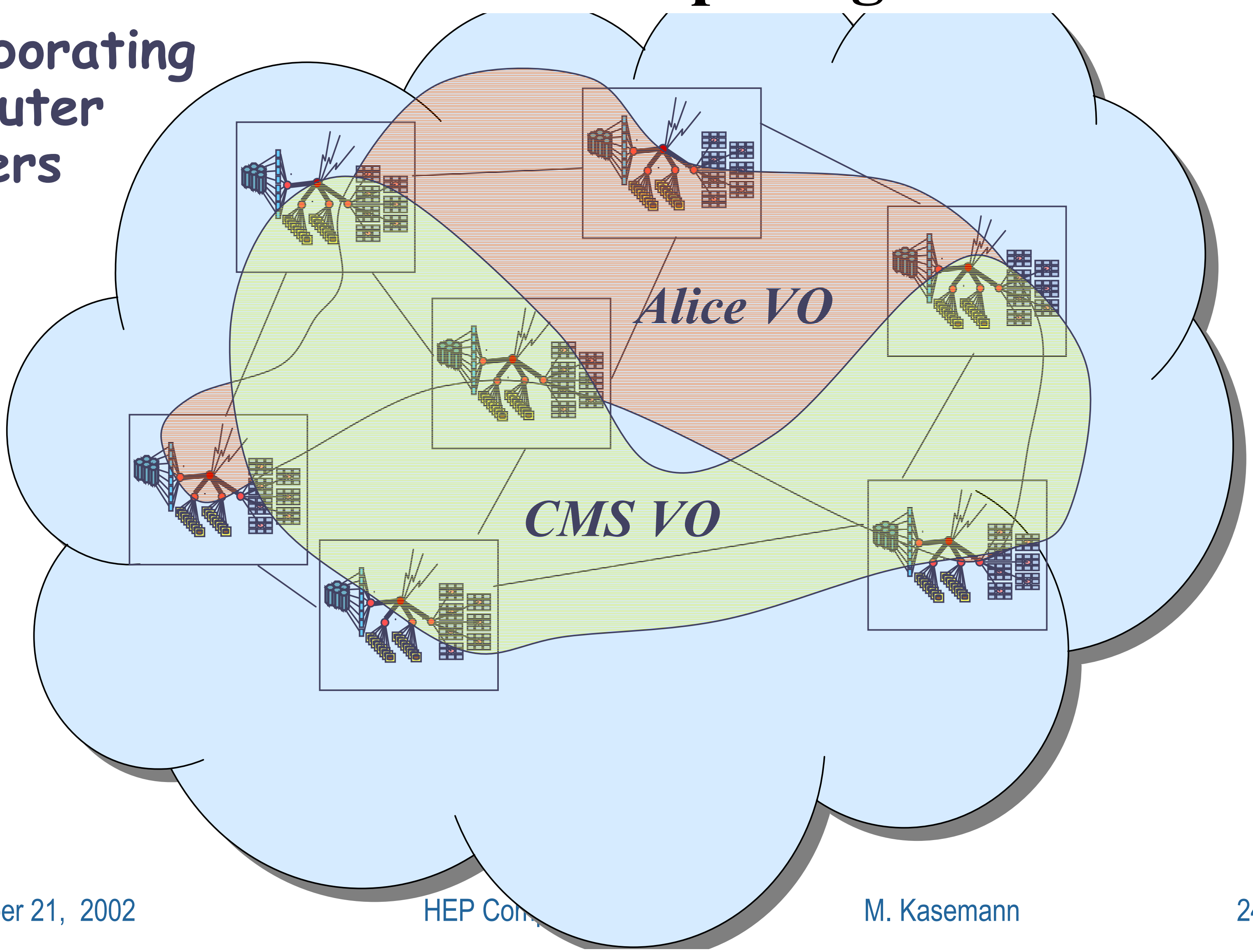


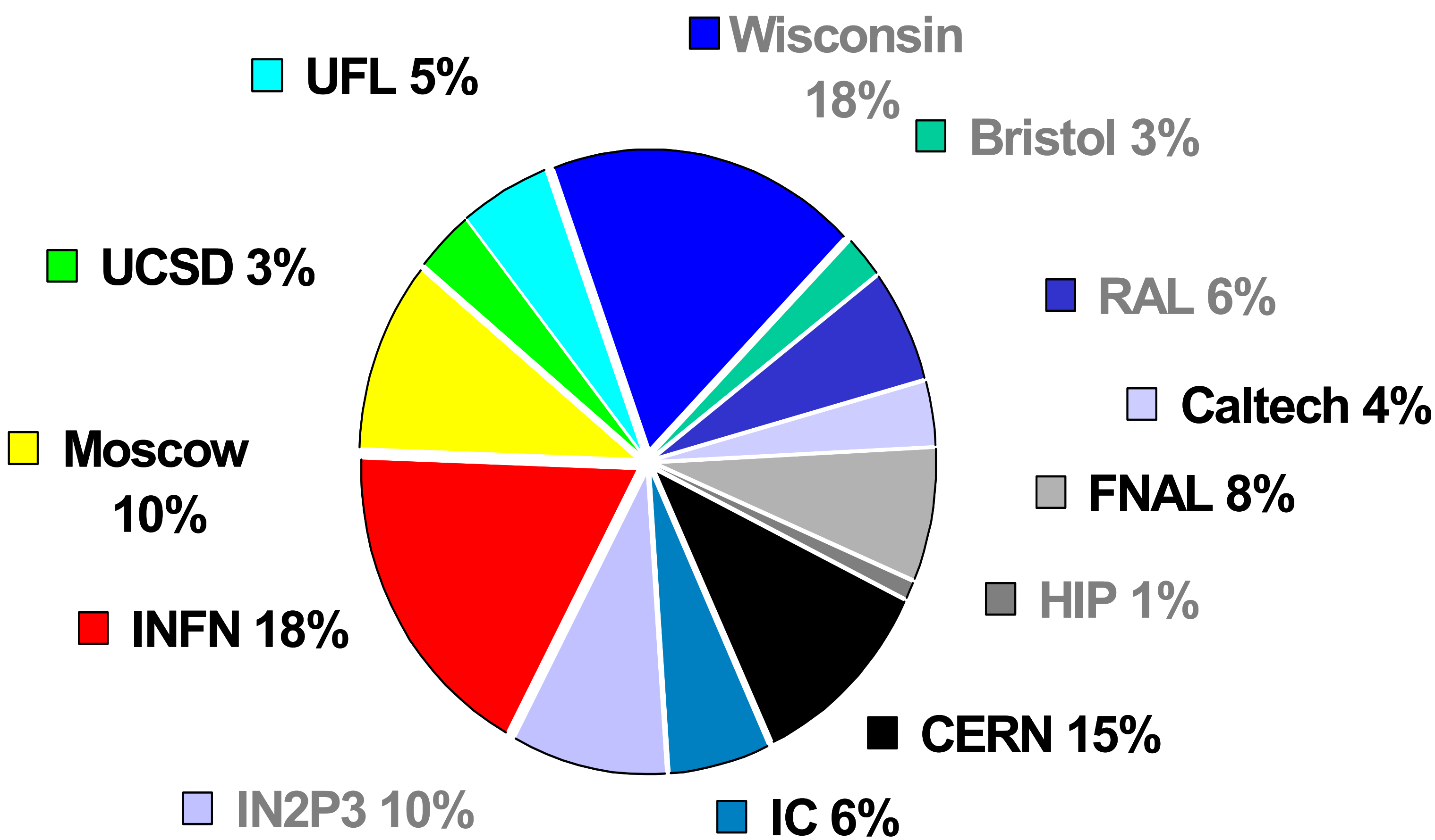
Table 1 - Estimated Capacity in Regional Centres taking part in the LCG Phase 1 Global Grid

year	*2002*	*2003*	*2004*	*2005*
Processing capacity (K-SI2000)				
CERN	200	400	700	1,400
Other Tier 1 Centres	450	1,300	2,700	5,200
Other Regional Centres	610	1,600	2,200	2,300
Total	**1,260**	**3,300**	**5,600**	**8,900**
CERN as % of total	***16%***	***12%***	***13%***	***16%***

- Data comes from centers reporting resources to the Grid Deployment Board
- This is a subset of the centers providing capacity for physics data challenges
- Some of the capacity will be used for non-LHC experiments

Spring02: CPU Resources

Most Resources not at CERN
(CERN not even biggest Single Resource)



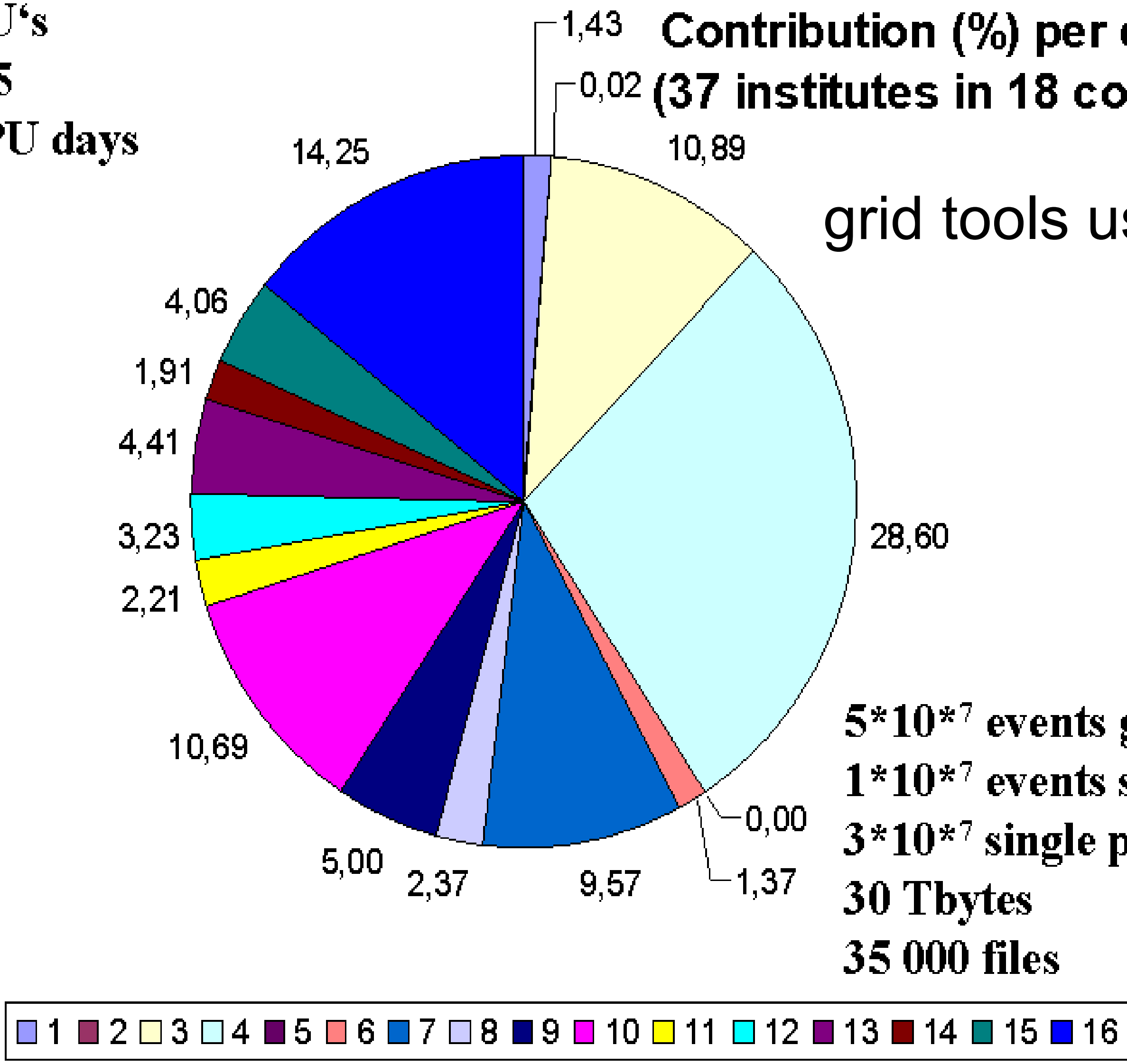
6 million events
~20 sites

ATLAS DC1 Phase 1 : July-August 2002

3200 CPU's
110 kSI95
71000 CPU days

Contribution (%) per country
(37 institutes in 18 countries)

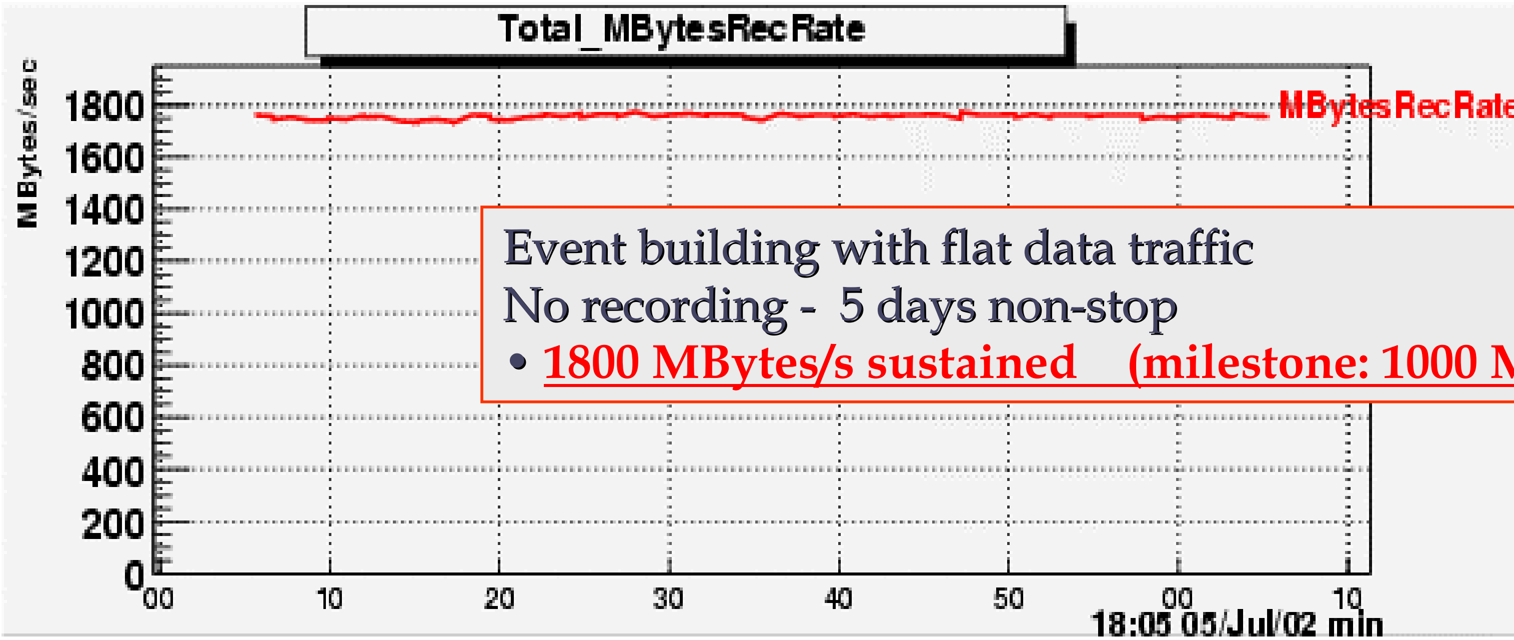
grid tools used at 11 sites



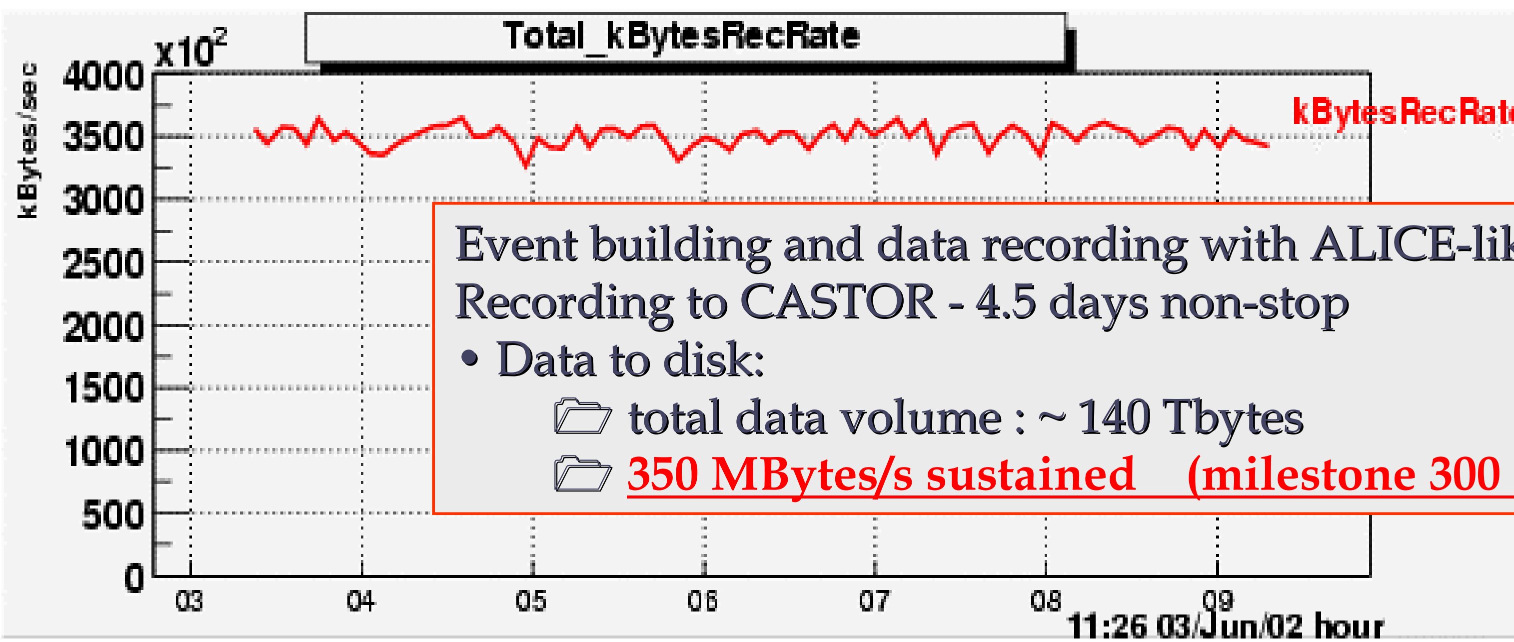
5*10*7 events generated
1*10*7 events simulated
3*10*7 single particles
30 Tbytes
35 000 files

Alois.Putzer@cern.ch

ADC IV performances – Period 1



Event building with flat data traffic
No recording - 5 days non-stop
- 1800 MBytes/s sustained (milestone: 1000 Mbytes/s)

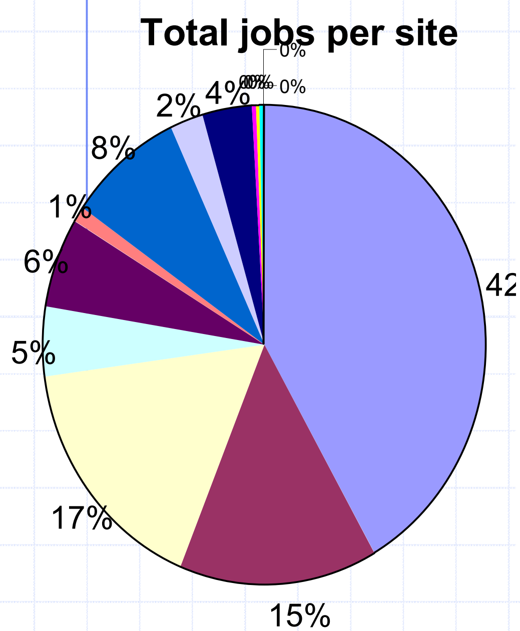


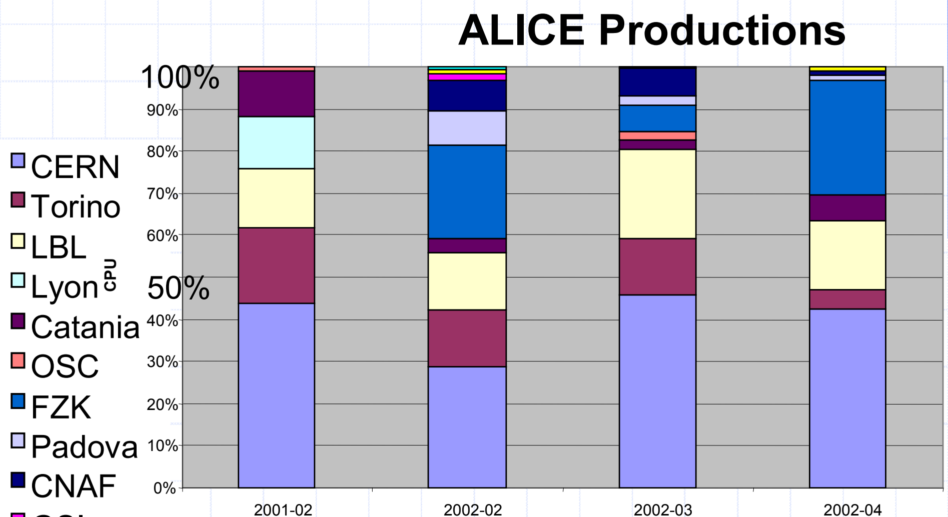
Event building and data recording with ALICE-like data traffic
Recording to CASTOR - 4.5 days non-stop
- Data to disk:
  - total data volume : ~ 140 Tbytes
  - 350 MBytes/s sustained (milestone 300 MBytes/s)

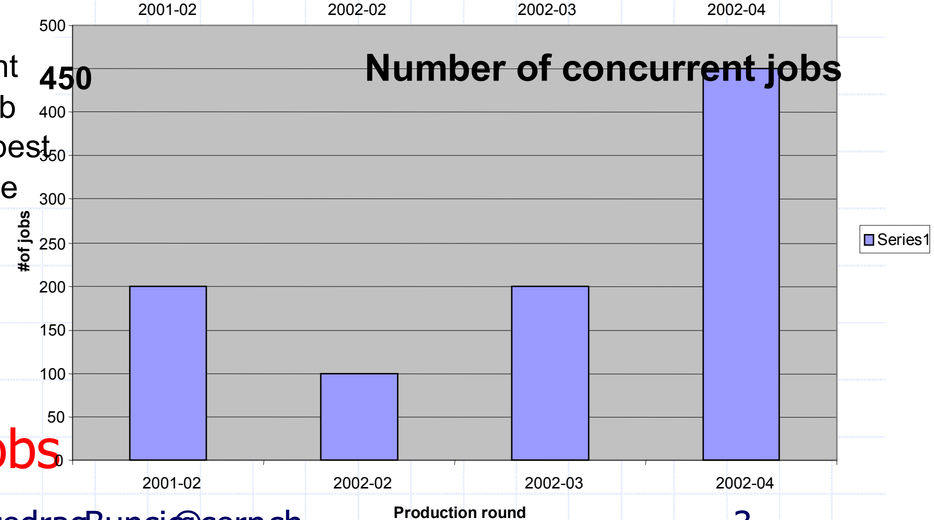
RUNNING PHYSICS DATA CHALLENGES WITH ALIEN

AliEn Production Status

@GRID







15100 jobs, ~12CPUh/job,

~1GB output/job

up to 450 concurrently running jobs

Today: ~900 Si2000/box
In 2007: 7k Si2000/box

Table 3: Summary of Computing Capacity Required for all LHC Experiments in 2008

	CERN Tier 0	CERN Tier 1	CERN Total	Other Tier 1	Total Tier 1	CERN as % of Tier 1	Total Tier 0 + 1	CERN as % of total Tier 0 + 1
Processing (K SI2000)	12,000	8,000	20,000	49,000	**57,000**	14%	**69,000**	29%
Disk (PetaBytes)	1.1	1.0	2.1	8.7	**9.7**	10%	**10.8**	20%
Magnetic tape (PetaBytes)	12.3	1.2	13.5	20.3	**21.6**	6%	**33.9**	40%

LHC experiments foresee *(…but funding dictates)* –

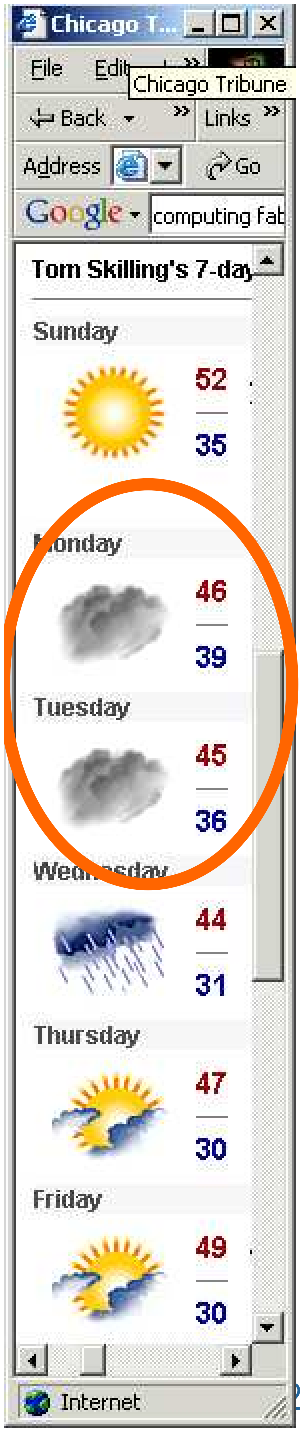
- Worldwide distributed computing system
- Small fraction of the analysis at CERN
- Batch analysis – using 12-20 large regional centers
  - have to use the resources efficiently
  - establishing and maintaining a uniform physics environment
- Data exchange and interactive analysis involving tens of smaller regional centers, universities, labs

Expected cost at CERN: 80M CHF *(60M in the budget for now)*

HEP Computing & Grid: Conclusions

- Experiments need substantial resources to perform computing and analysis (10 - 20% of detector costs)
- Central laboratories not able to provide all required resources
  - It is essential to fund and operate this in a distributed way (using Grid ideas and technology)

- LHC computing based on models emerging and experiences gained in current HEP experiments
- Multi-laboratory and –funding agencies computing projects needs formal project structure:
  - Project- and Project-Management Plan
  - well defined scope with identified resource estimate, schedule, deliverables, milestones and risk assessment
  - Project Organization
  - Project Oversight



LCCWS: a (warm) welcome!

- We have learned to setup, master and use cluster of commodity computers for HEP analysis in the different labs and Universities

- *"The goals of this workshop are to disseminate knowledge of design, construction and operational techniques and foster closer interactions between computer scientists and physicists.*
  - *technology review since the last meeting;*
  - *what are the issues about running a fabric within a grid"*

- We will have good work[...] conditions

HEPIX weather will be worse….

cold and chilly;-(