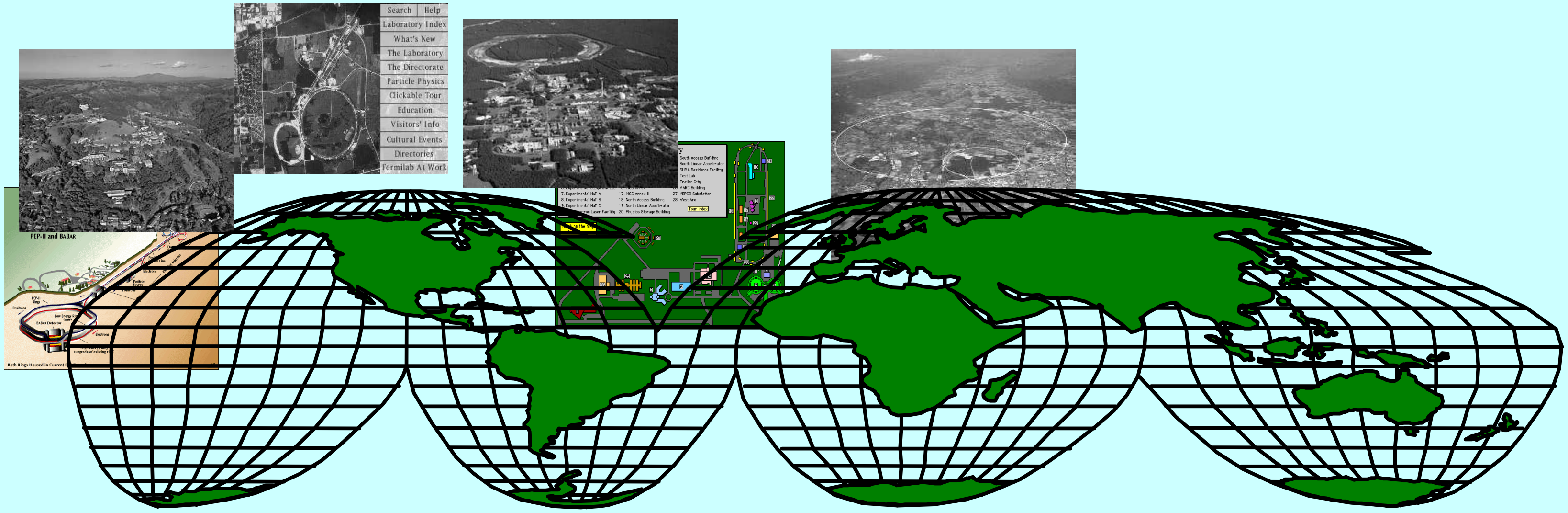
A Physics Optimized Grid Environment for Experiments (APOGEE)

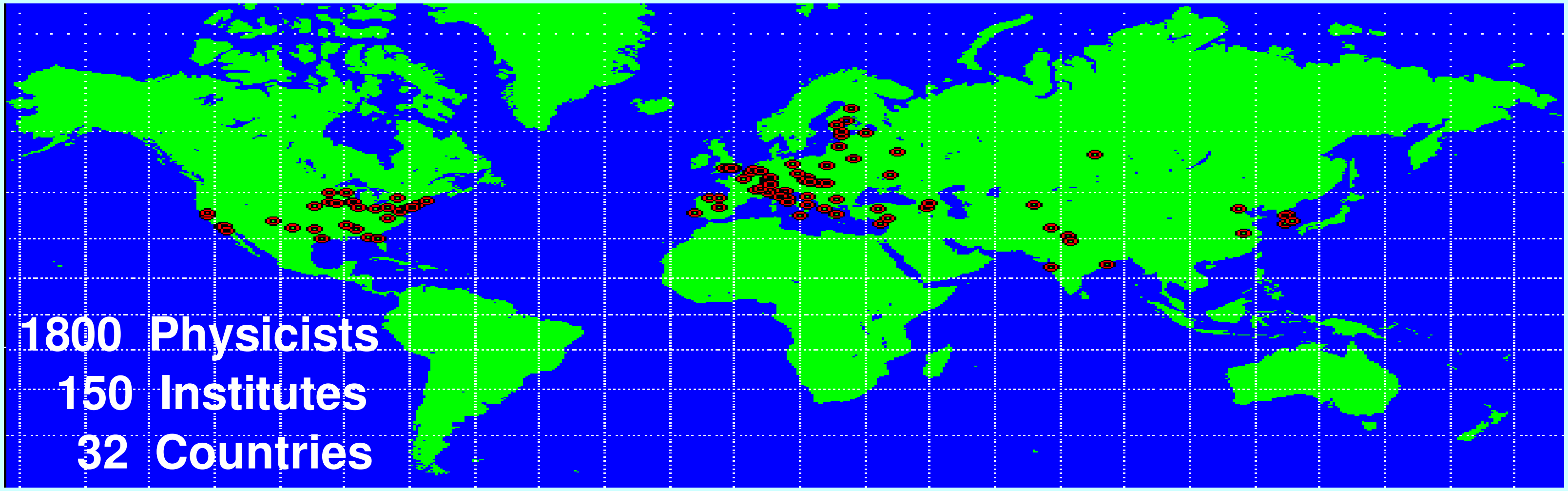


HENP Briefing

Harvey B. Newman, Caltech
DoE/Germantown,
August 18, 1999

LHC Computing Challenges

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


1800 Physicists
150 Institutes
32 Countries

Major challenges associated with:

Communication and collaboration at a distance

Distributed computing resources

Remote software development and physics analysis

R&D: New Forms of Distributed Systems

HENP (and LHC) Computing and Software Challenges

Software: Modern Languages, Methods and Tools
The Key to Manage Complexity

- ~~FORTRAN~~ The End of an Era;
  OBJECTS A Coming of Age
- *TRANSPARENT* Access To Data:
  - Location and Storage Medium Independence

Data Grids: A New Generation of Data-Intensive Network-Distributed Systems for Analysis

- A Deep *Heterogeneous* Client/Server Hierarchy, of Up to 5 Levels
- An Ensemble of Tape and Disk Mass Stores
- LHC: Object Database Federations

The Emergence of New Classes of Operating Systems

MONARC

Models Of Networked Analysis
At Regional Centers

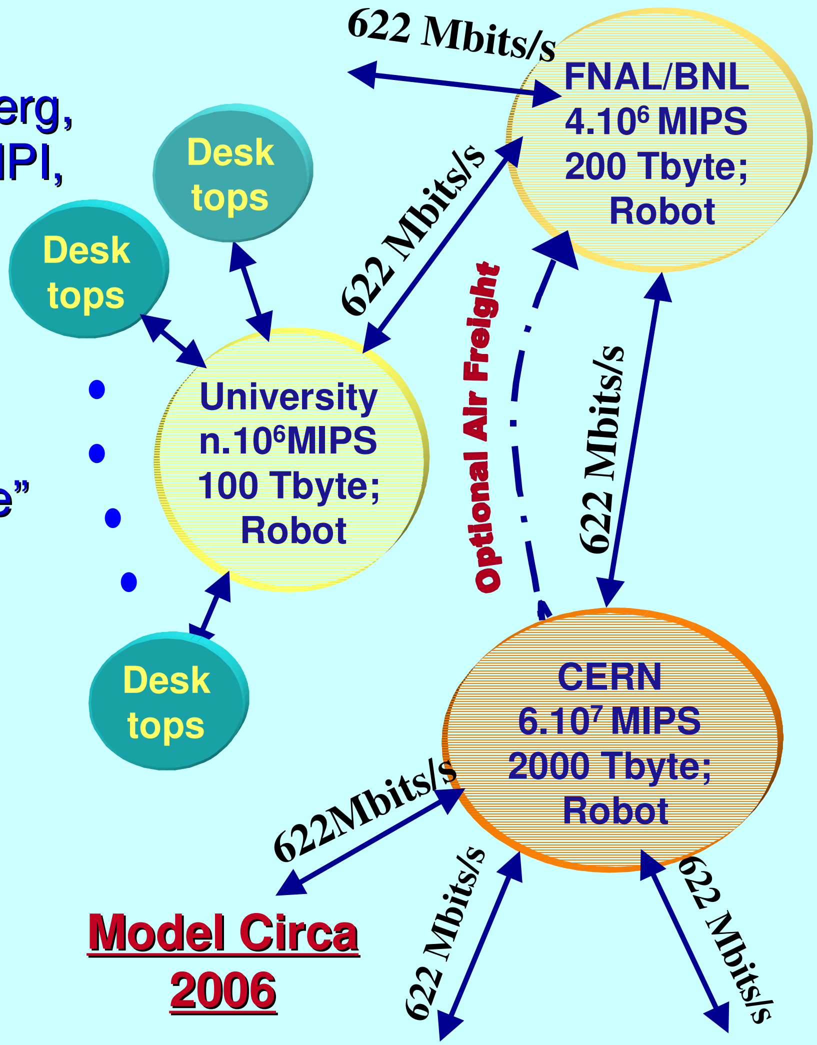
Caltech, CERN, Columbia, FNAL, Heidelberg, Helsinki, INFN, IN2P3, KEK, Marseilles, MPI, Munich, Orsay, Oxford, Tufts

GOALS

- Specify the main parameters characterizing the Model's performance: throughputs, latencies
- Develop "Baseline Models" in the "feasible" category
- Verify resource requirement baselines: (computing, data handling, networks)

COROLLARIES:

- Define the Analysis Process
- Define RC Architectures and Services
- Provide Guidelines for the final Models
- Provide a Simulation Toolset for Further Model studies



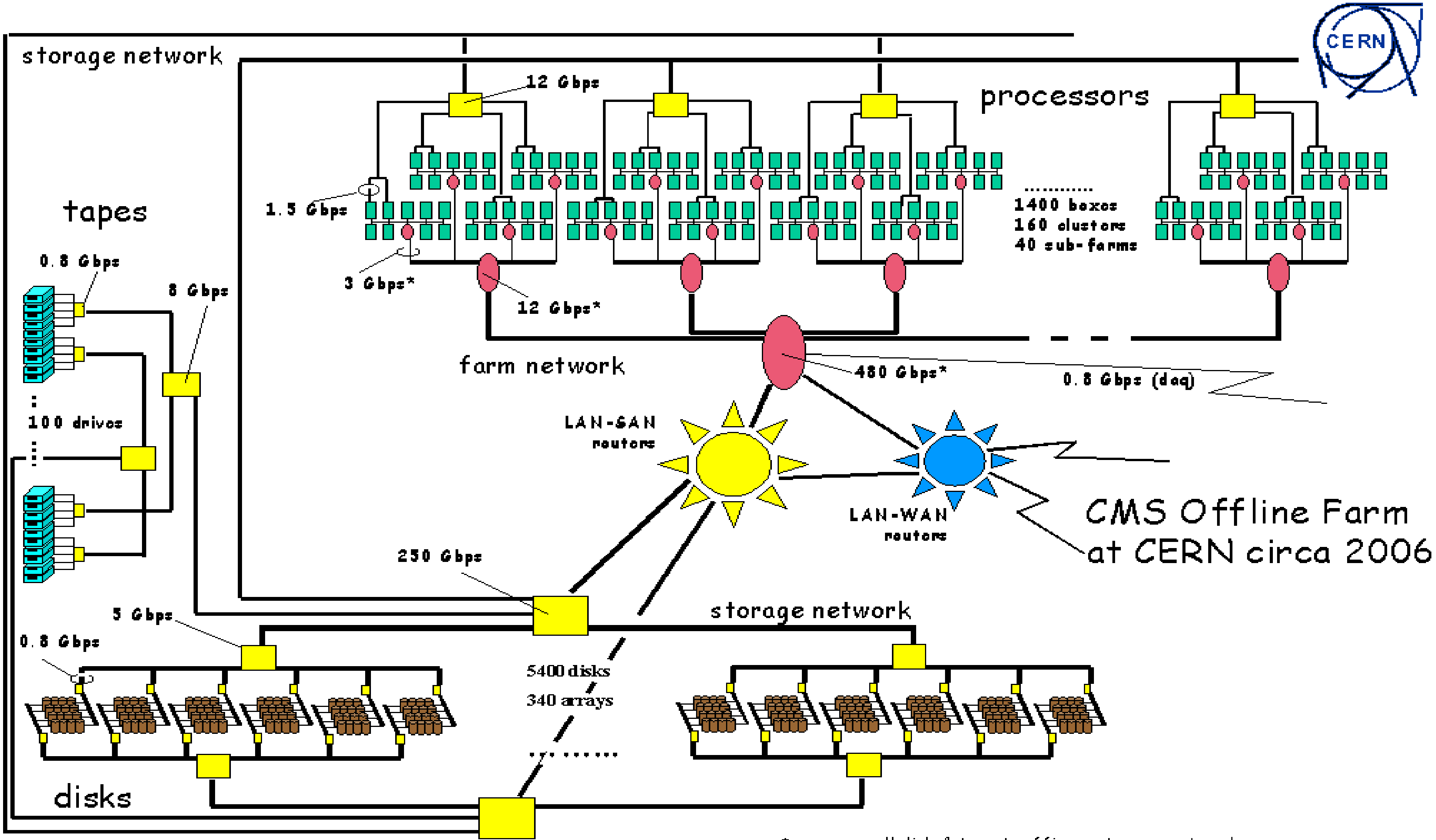
Model Circa 2006

MONARC General Conclusions on LHC Computing

Following discussions of computing and network requirements, technology evolution and projected costs, support requirements etc.

- The scale of LHC "Computing" is such that it requires a worldwide effort to accumulate the necessary technical and financial resources
- The uncertainty in the affordable network BW implies that several scenarios of computing resource-distribution must be developed
- A distributed hierarchy of computing centres will lead to better use of the financial and manpower resources of CERN, the Collaborations, and the nations involved, than a highly centralised model focused at CERN
  - Hence: The distributed model also provides better use of physics opportunities at the LHC by physicists and students
- At the top of the hierarchy is the CERN Center, with the ability to perform all analysis-related functions, but not the ability to do them completely
- At the next step in the hierarchy is a collection of large, multi-service "Tier1 Regional Centres", each with
  - 10-20% of the CERN capacity devoted to one experiment
- There will be Tier2 or smaller special purpose centers in many regions

Architectural Sketch: One Major LHC Experiment, At CERN



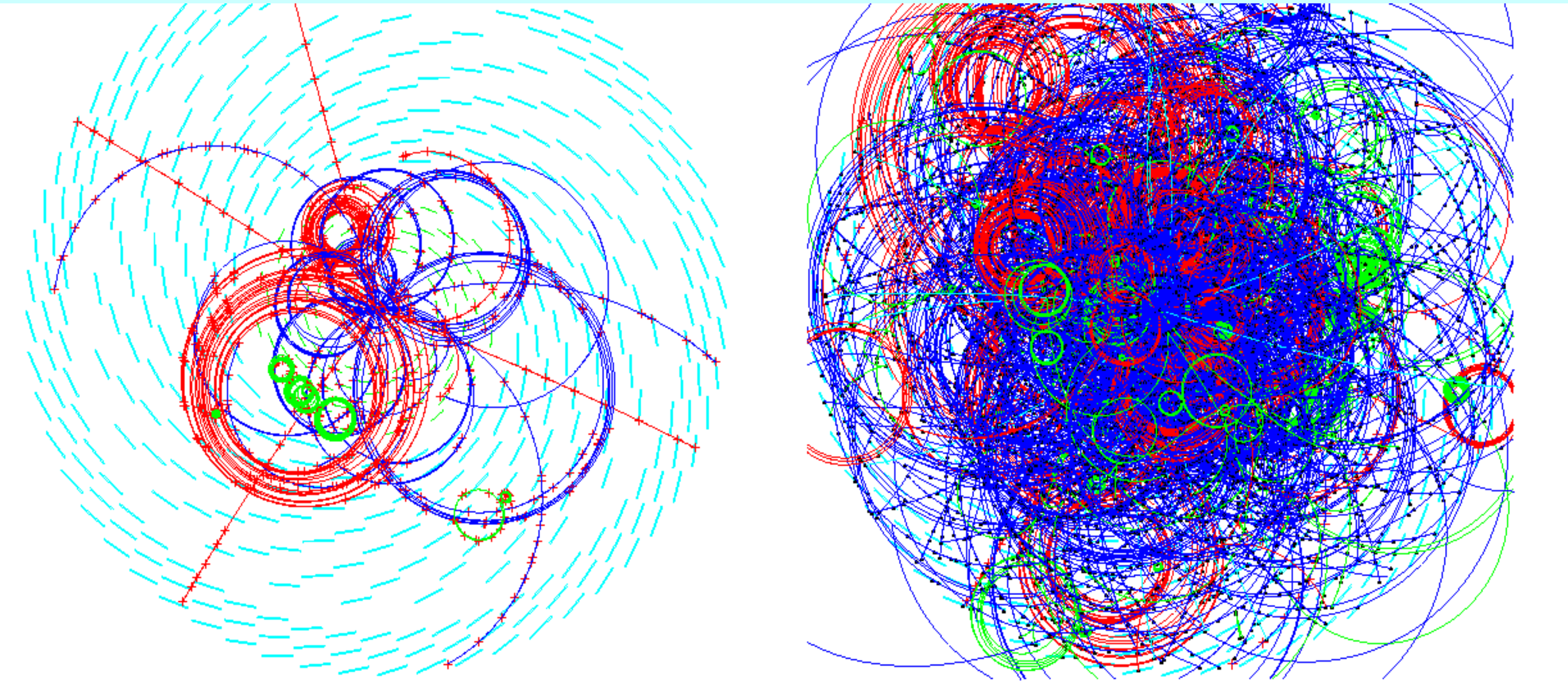
* assumes all disk & tape traffic on storage network
double these numbers if all disk & tape traffic through LAN-SAN router

lmr for Monarc study- april 1999

Challenges: Complexity

Events:

- Bunch crossing time of 25 ns is so short that (parts of) events from different crossings overlap
- Signal event is obscured by 20 overlapping uninteresting collisions in same crossing



Four Experiments:
The LHC Data Problem

ATLAS, CMS, ALICE, LHCB

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