Data Intensive Grids and Networks for High Energy and Nuclear Physics
Drivers of the Formation of an Information Society

Harvey B Newman
California Institute of Technology
Pasadena, CA 91125, USA
newman@hep.caltech.edu

November 2002
World Summit on the Information Society, Pan-European Ministerial Meeting, Bucharest

---

Introduction: Scientific Exploration at the High Energy Frontier

The major high energy and nuclear physics (HENP) experiments of the next twenty years will break new ground in our understanding of the fundamental interactions, structures and symmetries that govern the nature of matter and spacetime in our universe. Among the principal goals at the high energy frontier are to find the mechanism responsible for mass in the universe, and the "Higgs" particles associated with mass generation, as well as the fundamental mechanism that led to the predominance of matter over antimatter in the observable cosmos.

The largest collaborations today, such as CMS (http://cmsdoc.cern.ch) and ATLAS (http://atlasexperiment.org) who are building experiments for CERN's (http://www.cern.ch) Large Hadron Collider (LHC; http://lhc.web.cern.ch/lhc) program, each encompass 2000 physicists from 150 institutions in more than 30 countries. Each of these collaborations include 300-400 physicists in the US, from more than 30 universities as well as the major US HEP laboratories. The current generation of experiments now in operation and taking data at SLAC (BaBar; http://www-public.slac.stanford.edu/babar) and Fermilab (D0 and CDF; http://www-d0.fnal.gov and http://www-cdf.fnal.gov) face similar challenges. BaBar has already accumulated nearly a Petabyte (1 PB = 10^{15} Bytes) of stored data.

Collaborations on this global scale would not have been attempted if the physicists could not plan on excellent networks[1]: to interconnect the physics groups seamlessly, enabling them to collaborate throughout the lifecycle of the experiment, and to make possible the construction of Data Grids capable of handling massive datasets, rising from the Petabyte to the Exabyte scale within the next decade.

---

[1] As well as state of the art tools for remote collaboration, such as Caltech's VRVS system (see http://www.vrvs.org).

HEP Challenges: at the Frontiers of Information Technology

Realizing the scientific wealth of these experiments presents new problems in data access, processing and distribution, and collaboration across national and international networks, on a scale unprecedented in the history of science. The information technology challenges include:

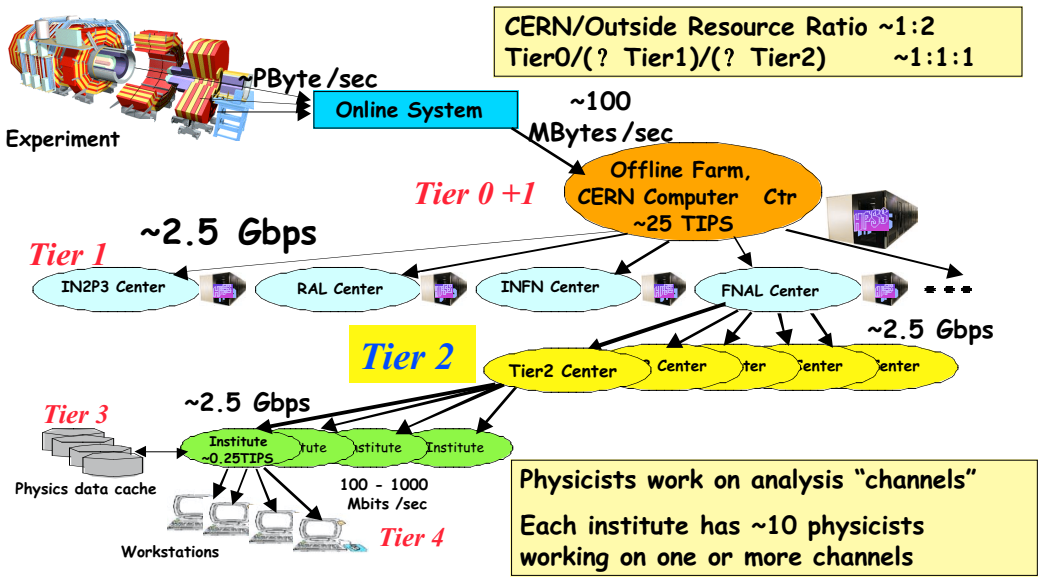
- Providing rapid access to data subsets drawn from massive data stores, rising from Petabytes in 2002 to ~100 Petabytes by 2007, and Exabytes (10^{18} bytes) by approximately 2012 to 2015.
- Providing secure, efficient and transparent managed access to heterogeneous worldwide-distributed computing and data handling resources, across an ensemble of networks of varying capability and reliability
- Matching resource usage to policies set by the management of the experimental Collaborations over the long term; ensuring that the application of the decisions made to support resource usage among multiple Collaborations sharing common (network and other) resources are internally consistent.
- Providing the collaborative infrastructure that will make it possible for physicists in all world regions to contribute effectively to the analysis and the physics results, including from their home institutions.
- Integrating all of the above infrastructures to produce the first managed distributed systems serving "virtual organizations" on a global scale

Meeting the HEP Challenges: Data Grids as Managed Global Systems

In order to meet these challenges, the LHC experiments have adopted the "Data Grid Hierarchy" concept (developed by the MONARC project (http://www.cern.ch/.MONARC) shown schematically in the figure below. This model shows data at the experiment is stored at the rate of 100 – 1500 Mbytes/sec throughout the year, resulting in many Petabytes per year of stored and processed binary data that are accessed and processed repeatedly by the worldwide collaborations searching for new physics processes. Following initial processing and storage at the "Tier0" facility at the CERN laboratory site, the data is distributed over high speed networks to ~10 national "Tier1" centers in the US and the leading European and other countries. The data is further processed and analyzed and stored at approximately 50 "Tier2" regional centers, each serving a small to medium-sized country, or one region of a larger country (as in the US, UK and Italy). Data subsets are accessed from and further analyzed by physics groups using one of hundreds of "Tier3" workgroup servers and/or thousands of "Tier4" desktops.

The successful use of this global ensemble of systems to meet the experiments' scientific goals depends on the development of Data Grids capable of managing and marshalling the "Tier-N" resources, and supporting collaborative software development by groups of varying sizes spread across the globe. Many Grid projects involving high energy physicists, including GriPhyN (http://www.griphyn.org) , PPDG (http://ww.ppdg.net) , iVDGL (www.ivdgl.org) , EU Datagrid (http://www.eu-datagrid.org), DataTAG (http://datatag.web.cern.ch/datatag) the LHC

Computing Grid project (http://lcg.web.cern.ch/LCG), and national Grid projects in Europe and Asia are working together, in multi-year R&D programs, to develop the necessary Grid systems. The DataTAG project is also working to address some of the network R&D issues and to establish a transatlantic testbed to help ensure that the US and European Grid systems interoperate smoothly.



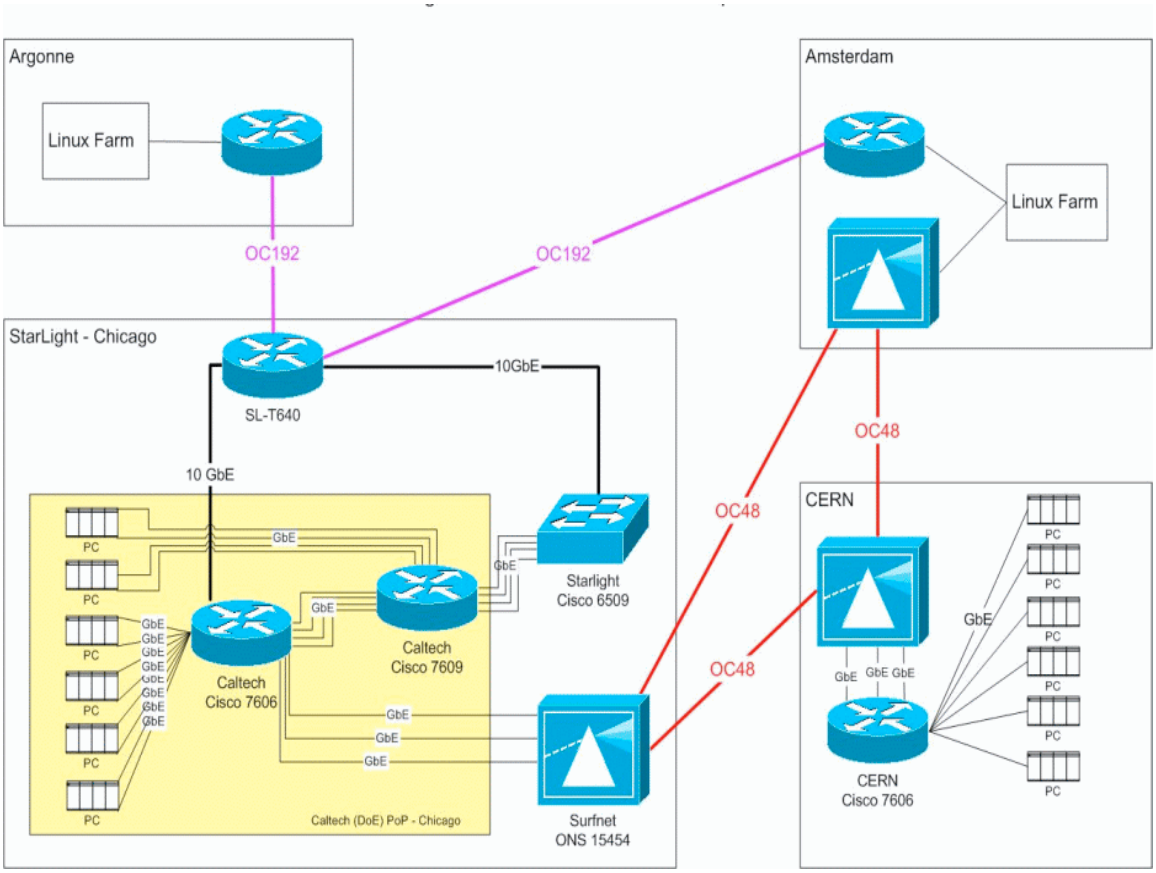
The data rates and network bandwidths shown in the figure above are per major experiment, and correspond to a conservative "baseline" formulated using an evolutionary view of network technologies. More recent estimates of the network needs indicate that the needs on the transatlantic and other network needs will reach 10 Gigabits/sec (Gbps) within the next 2 to 3 years, followed by a need for scheduled and dynamic use of 10 Gbps wavelengths by the time the LHC begins operation at CERN in 2007. In order to build a "survivable", flexible distributed system, much larger bandwidths are required, so that the typical data transactions, drawing 1 to 10 Terabyte and eventually 100 Terabyte subsamples from the multi-Petabyte data stores, can be completed in the span of a few minutes.

Completing these transactions in minutes rather than hours is necessary to avoid the bottlenecks that would result if hundreds to thousands of requests were left pending for long periods, and to avoid the bottleneck that would result from tens and then hundreds of such "data-intensive" requests per day (each still representing a very small fraction of the stored data). It is important to note that transactions on this scale correspond to data throughputs across networks of 10 Gbps to 1 Terabit/sec (Tbps) for 10 minute transactions, and up to 10 Tbps (comparable to the current capacity of a fully instrumented fiber today) for 1 minute transactions.

In order to fully understand the potential of these applications to overwhelm future planned networks, we note that the binary (compacted) data stored is pre-filtered by a factor of 10^6 to 10^7 by the "Online System" (a large cluster of hundreds to thousands of CPUs that filter the data in real time). This realtime filtering, though traditional, runs a certain risk of throwing away data from subtly new interactions that do not fit into pre-conceived existing or hypothesized theories. The basic problem is to find new interactions from the particle collisions, down to the level of a few interactions per year out of 10^{16} produced. A direct attack on this problem, analyzing every event in some depth without pre-filtering, is beyond the current and foreseen states of network and computing technologies.

US universities and laboratories engaged in high energy physics have had a leading role in these developments. The BaBar experiment at SLAC is among the largest users of national and international networks. The US contingent of the CMS experiment, including Caltech, Florida and Fermilab in particular, has led the development of the LHC distributed computing model and has had a leading role in the development, operation and planning for HENP's international networks over the last 20 years, in collaboration with LBNL (http://www.lbl.gov), SLAC (http://www.slac.stanford.edu) and FNAL (www.fnal.gov) , and more recently CERN and StarLight (www.startap.net/starlight). The physicists in the ATLAS project also have contributed to these efforts, led by the University of Michigan, Indiana and the Argonne, Berkeley and Brookhaven national labs. Caltech and UCSD recently deployed the first prototype Tier2 center in CMS, split between Caltech/CACR and SDSC, and are working closely with UC Davis, Riverside and UCLA in California, as well as the University of Florida on a diverse set of physics studies to develop optimal search strategies for the Higgs particles. Caltech and UCSD intend to use the TeraGrid, and prototype Tier1 centers at FNAL and CERN, to meet the demanding needs for simulated particle interactions in these studies.

Plans are underway to put "last mile fiber" in place between the Caltech Center for Advanced Computing Research (CACR) and the CENIC and Abilene points of presence in downtown Los Angeles, and to use OC192 (10 Gigabit/sec) wavelengths for HEP applications starting in the Spring of 2003. Similar initiatives are underway to link Fermilab to StarLight, the international peering point in Chicago, and to link the IN2P3 Computing Centre in France to CERN, using dark fibers. Caltech and CERN have recently installed servers, routers and switches at the StarLight and in Geneva to drive the development of the transatlantic networks needed by HENP, as shown in the Figure below, and to fully utilize the OC192 (10 Gbps) link donated by Level3 for iGrid2002 (September 2002) By mid-2003 the physicists at Caltech, UCSD and other universities in the CMS experiment plan are hoping to begin use of the new US optical fiber infrastructure National Light Rail (or Pacific Light Rail and a wavelength across Canada linking Seattle and StarLight), to build an intercontinental network for science using a 10 Gbps wavelength, stretching from California to Geneva and Amsterdam.



Relevance of Meeting These Challenges for Future Networks and Society

Successful construction of network and Grid systems able to serve the global HENP and other scientific communities with data-intensive needs could have wide-ranging effects on research, industrial and commercial operations. Resilient self-aware systems developed by the HENP community, able to support a large volume of robust Terabyte and larger transactions, and to adapt to a changing workload, could provide a strong foundation for the distributed data-intensive business processes of multinational corporations of the future.

Development of the new generation of systems of this kind could also lead to new modes of interaction between people and "persistent information" in their daily lives. Learning to provide, efficiently manage and absorb this information and in a persistent, collaborative environment would have a profound transformational effect on our society

Closing the Digital Divide

The world community will only reap the benefits of global collaborations in research and education, and of the development of advanced network and Grid systems, if we work to close the Digital Divide that separates the economically and technologically most-favored from the less-favored regions of the world. The ICFA Standing Committee on Inter-Regional Connectivity (ICFA-SCIC; http://cern.ch/icfa-scic) and the Internet Equal

Educational Access Foundation; http://www.ieeaf.org), among other, are working fervently towards this goal on behalf of the high energy physics and the broader research and education communities. Approaches to help close the Divide include:

- Sharing and systematization of information on the Digital Divide. ICFA SCIC for example is gathering information on these problems and developing a Web Site on the Digital Divide problems of research groups, universities and laboratories throughout its worldwide community. This will promote understanding the nature of the problems, from lack of backbone bandwidth to last mile connectivity problems to policy and pricing issues. Sharing information on examples of how the Divide has been bridged in a city, country or region; identifying common themes in the nature of the problem and the corresponding solutions methods; making technical and comparative pricing information generally available; all with help develop a general approach to solving the problem globally.
- Use (lightweight; non-disruptive) network monitoring to track the problem, and keep the research community (and the world community) informed on the evolving state of the Digital Divide. One leading example in the HEP community is the Internet End-to-end Performance Monitoring (IEPM) initiative (http://www-iepm.slac.stanford.edu) at SLAC.
- Identify and work on specific problems, country by country and region by region, to enable groups in all regions to be full partners in the process of search and discovery in science.
- Create and encourage inter-regional programs to solve specific regional problems. Leading examples include the Virtual Silk Highway project (http://www.nato.int/science/e/silk.htm) led by the DESY HEP laboratory in Germany the support for links in Southeast Asia by the KEK high energy physics laboratory in Japan (http://www.kek.jp), and the support of network connections for research and education in South America by the AMPATH "Pathway to the Americas" (http://www.ampath.fiu.edu) at Florida International University.
- Make direct contacts, and help educate government officials on the needs and benefits to society of the development and deployment of advanced infrastructure and applications: for research, education, industry, commerce, and society as a whole.
- Help start and support workshops on networks, Grids, and the associated advanced applications.
- Help form regional support and training groups for network and Grid system development, operations, monitoring and troubleshooting.