A Vision for DOE Scientific Networking Driven by High Impact Science

W. E. Johnston, W. T. C. Kramer, J. F. Leighton, LBNL
Charlie Catlett, ANL
March 15, 2002

Contributors
Michael Banda (LBNL), Julian Borrill (LBNL/NERSC), Al Kellie (NCAR), Stewart Loken (LBNL), Paul Messina (CalTech), Harvey Newman (Caltech)

Contents

1 Vision and Approach

The vision for DOE scientific networking is that major DOE applications and facility-based experiments will interconnect with widely distributed terascale supercomputing, petascale storage, high performance visualization, and remote collaborators, to dynamically create virtual laboratories. This will provide unprecedented presence and interaction for all of the participating scientists, and enable the interplay of theory, simulation, and experiment which in turn should lead to new ways of approaching science problems and new levels of scientific productivity.

In order to achieve this vision, we propose a multi-pronged approach that needs to be developed and implemented. This approach describes the framework in which DOE provides the most effective and the highest possible performance networking where needed by its high impact science. This framework will continue to provide required high quality network capabilities to the DOE and its community. At the same time, it will integrate the DOE's networking programs, facilities, and expertise into a more effective and efficient enterprise. This approach has six major components:

- DOE science programs will identify the high impact projects that set the scientific networking priorities.

  This approach is intended to tie the operation and evolution of DOE's scientific networking to DOE's high impact science projects, facilities, and teams. This entails:
  - providing network expertise to support the development and tuning of high performance, distributed science applications
  - direct involvement of distributed-application scientists in providing feedback to network engineers
  - an Office of Science program-based mechanism for prioritizing network resources based on their assessment of high impact science projects.

- A High Performance Production Network (HPPN): Provides a high bandwidth production network to enable integration of widely distributed DOE resources.

  The HPPN is the flagship Office for Science network that supports the scientific program with high quality networking and services to support distributed science.

- An Advanced Scientific Applications Pilot (ASAP) Network: Provides advanced pilot networks to develop new capabilities for high end science applications and deploy new network features.

  The ASAP Networks address key science application areas that have very bandwidth-intensive requirements *and* can benefit from early access to very high bandwidth and advanced services not yet deployable in full production. It will also bridge the gap between the network technology focused testbeds and the HPPN.

- An Advanced Network Technology Consortium Testbed (ANTCT): Provides advanced network technology testbeds to shape the next generation networks for science.

  In order to meet the needs of science that is experiencing exponential growth in the collection and analysis of data, vastly higher effective bandwidths than are available with today's technology are needed. The role of the ANTCT is to explore network technology that DOE science networks will need in five, seven, or more years.

?? Advanced Services: Provide a suite of shared middleware services for enabling distributed science applications.

The success of advanced distributed science applications depends critically on distributed middleware services in order to build realistic and reliable distributed applications. A suite of advanced services will be deployed in an integrated and non-redundant manner across the individual networks.

?? Provide improved ways to introduce new network technology that enhance networks for effective science and new middleware features.

The approach includes methods of information and technology flow between the three networks to enable timely introduction of new network services and features that improve the effectiveness of the science drivers using the networks.

Table 1.1
Characteristics of the Networks
This table indicates some of the basic differences of the three networks

Service Characteristic	**HPPN Network**	**ASAP Network**	**ANTCT Network**
Bandwidth relative to current ESnet	4 times i.e., 2.5 Gbps	16 times i.e., 10 Gbps	Defining characteristics will probably be different network architecture, protocols, etc.
Number of sites	30–50	4–6	Determined opportunistically
Maturity of applications	Full range of production applications	Limited set of early adapter applications	Experimental applications and application kernels
Reliability	99.9%	95-98%	50-80%
Mean time between failure	Months	Weeks	Days
Mean time to repair	2–4 hours	Next business day	Days to weeks

2 Example Science Drivers

DOE networking is driven from the science being done by DOE researchers and at DOE facilities and also enables science that is not otherwise possible. The particular science drivers will change and evolve over time. Below are some examples of the science areas that demand very high quality and high bandwidth networking now and in the future.

2.1 High Energy and Nuclear Physics[1]

History and Background

For 20 years, high energy physicists have relied on state-of-the-art computer networking to enable ever larger and more complex experiments. As computer networking became possible over wide areas, starting in the late 1970s, adoption by high energy physics came quickly. The practice, already prevalent, of conducting complex experiments through collaborations of groups from universities and laboratories that were often in different parts of the country created an immediate need for the new technology. In the 1980s, the connection to the emerging "HEPnet" became a required piece of infrastructure for any group serious about doing experiments in high energy physics. The network solved problems for the modest-sized collaborations of the day, and in turn enabled the evolution of the ever-larger and increasingly international collaborations that were needed to mount major collider experiments and other complex instruments. Collaborations on the scale of those now building LHC detectors could never have been attempted if they had not been able to expect the excellent international communications that make them possible.

Today, the network is needed for all aspects of collaborative work. Collaborators work together across the network to write proposals, produce and agree on designs of components and systems, collaborate on overall planning and integration of the detector, confer on all aspects of the device, including the final physics results, and provide information to collaborators and to the physics community and general public. The network supports all phases of the experimental activity. The data are acquired and written to disk or tape across at least the local area network of the host laboratory. Data move to the places in the collaboration where they are needed for processing or detailed data analysis. And finally the resulting physics papers are written, edited, and reviewed on and across the network.

Because of its need for large distributed collaborations, HEP has traditionally led the demand for research networks among the sciences. HEP developed its own national network in the early 1980s, until the multidisciplinary networks supported by DOE and, for a time, by NSF emerged. Since the installation of national backbones, HEP and other sciences have generally received support from the research network backbones. There have always been, however, specific network connections where HEP has found it necessary to support special capabilities that could not be supplied efficiently or capably enough through more general networks. The most important example is the link across the Atlantic to support U.S. experimenters at CERN, starting in the 1980s for L3 and other LEP experiments, and more recently for work on the LHC. These needs for links dedicated to HEP use are needed in special cases because HEP requirements can be large and can overwhelm those of researchers in other fields and, as in the

[1] This section was taken in part from the October 2001 report of the DOE/NSF Transatlantic Network Working Group (TAN WG) co-chaired by L. Price (ANL) and H. Newman (Cal Tech), along with individual contributions by H. Newman and S. Loken (LBNL).

case of the link to CERN, because regional networks do not give top priority to interregional connections.

The Future of HEP Networking

The experiments in HENP seek to answer some of the most fundamental questions about the nature of matter and energy in the universe. For example, current experiments at SLAC and Fermilab investigate basic symmetries between matter and antimatter and seek to understand the origin of particle mass. The experiments at LHC will continue this search into a new energy regime. To answer these basic questions, physicists must use the highest energy particle accelerators and must build complex detectors to study events resulting from the collisions of beams of particles. The answers do not come easily. To search for new particles or to study the properties of known particles, they must collect millions of events and search through them to find a small number of rare and unusual things. As a result, the experiments must deal with huge samples of data, soon measured in petabytes (10^{12} bytes), and must analyze those data at collaborating institutions around the world.

The network is crucial to the discovery process, and without robust, high performance networking, the experiments cannot achieve their scientific objectives. The current and future experiments place greater demands on the network than ever before. These demands result from the increasing complexity of experiments, the increasing size and scope of the international collaborations, and the deployment of powerful regional computing centers across the collaborations. These regional centers provide unprecedented computing power to the high energy physics teams, but their full potential can only be realized by creating the communication infrastructure to support data sharing throughout the collaboration. With the network in place, physicists will be able to carry out complex searches for events of any type from a local computer. The search will go to many different data repositories at centers across the world, taking advantage of many parallel data streams to explore much larger samples of data than would reside on a local machine. This distributed search and analysis capability will be crucial to the experiments if they are to be able to find rare and complicated events in the presence of a very large background of less interesting data.

The network is also critical for maintaining the scientific vitality of the collaboration through its support for collaboration tools. Groups from a single institution need to stay in contact with each other while some are working at the experimental facility and others are at home. This is especially important for maintaining communication between faculty and students. In addition, analysis teams working on a single physics topic will need to stay in close communication and share the results of their analyses. These teams must be able to function as a single unit as they develop software and use those programs to search for signals of new phenomena. This close collaboration demands not only real-time interactions such as video conferencing and shared analysis tools, but also requires distributed data management, workflow management, and distributed electronic notebooks to manage and track the analysis. Without the network, these teams cannot function.

Current Usage and Future Requirements

The U.S. HENP community relies on state-of-the-art networks for its major research programs, both for experiments based at Fermilab, SLAC, BNL and JLab, and for U.S. participation abroad in the LHC and other experiments at CERN and DESY. This trend continues, and in fact, the estimates for future U.S. HENP domestic and transatlantic network requirements have increased rapidly over the last two years, as documented in the October 2001 report of the DOE/NSF

Transatlantic Network Working Group (TAN WG)[2]. The increased requirements are driven by the planned deployment of distributed analysis applications, and especially the emergence of "Data Grids"[3], that are expected to meet the needs of the worldwide HENP collaborations. The LHC "Data Grid hierarchy" example (shown in Figure 2.1) illustrates that the requirements for each LHC experiment are expected to reach 2.5 Gbps by approximately 2005 at the national Tier1 centers at FNAL and BNL, and 0.6–2.5 Gbps at the regional Tier2 centers. Taken together with other programmatic needs for links to DESY, IN2P3, and INFN, this corresponds to an aggregate transatlantic bandwidth requirement rising from 3 Gbps in 2002 to 23 Gbps in 2006.

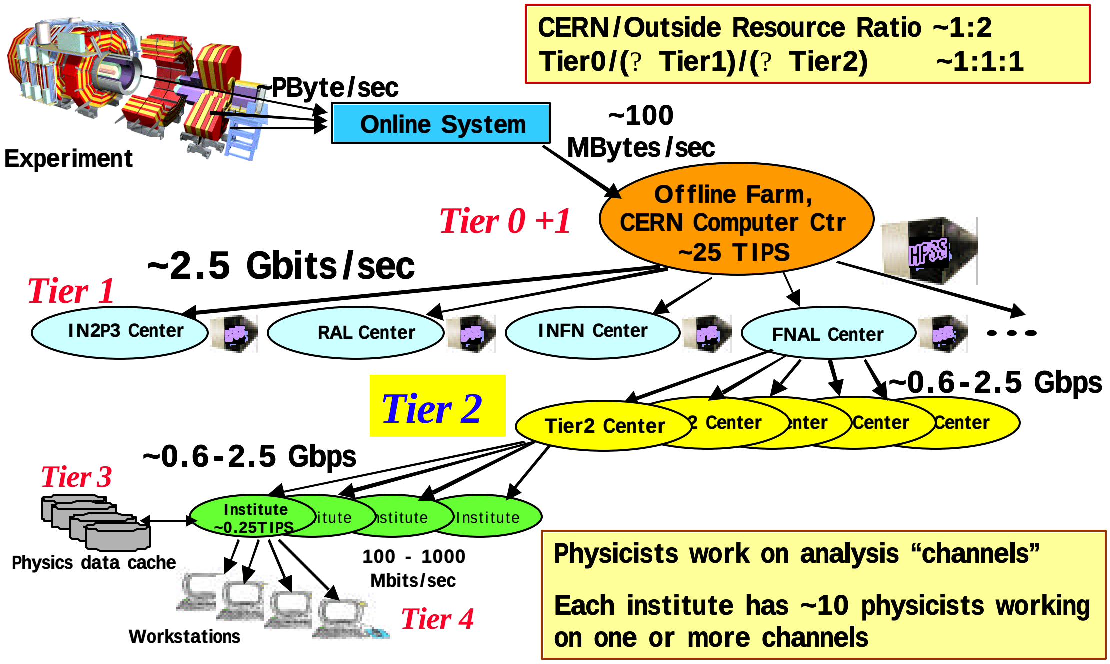


Figure 2.1.

The LHC Data Grid Hierarchy

One of the surprising results of the TAN WG report, shown in Table 2.1, is that the present-generation experiments, BaBar, D0, and CDF, have transatlantic network bandwidth needs that equal or exceed the levels presently estimated by the future LHC experiments, CMS and ATLAS. This is ascribed to the fact that the experiments now in operation are distributing (BaBar) or plan to distribute (D0, CDF in Run 2B) substantial portions of their event data to regional centers overseas, while the LHC experiments have so far foreseen only limited data distribution.

The corresponding bandwidth requirements at the U.S. HEP labs and on the principal links across the Atlantic are summarized[4] in Table 2.2.

---

[2] The report of this committee, commissioned by the U.S. DOE and NSF and co-chaired by H. Newman (Caltech) and L. Price (ANL) may be found at http://gate.hep.anl.gov/lprice/TAN. For comparison, the May 1998 ICFA Network Task Force Requirements report may be found at http://l3www.cern.ch/~newman/icfareq98.html.

[3] Data Grids for high energy and astrophysics are currently under development by the Particle Physics Data Grid (PPDG; see http://ppdg.net), Grid Physics Network (GriPhyN; see http://www.griphyn.org), and EU Data Grid (see http://www.eu-datagrid.org) projects, as well as several national Grid projects in Europe and Japan.

[4] The entries in the table refer to standard commercial bandwidth offerings. OC3 = 155 Mbps, OC12 = 622 Mbps, OC48 = 2.5 Gbps, and OC192 = 10 Gbps.

Table 2.1. Installed Transatlantic Bandwidth Requirements

	2001	2002	2003	2004	2005	2006
CMS	100	200	300	600	800	2500
ATLAS	50	100	300	600	800	2500
BaBar	300	600	1100	1600	2300	3000
CDF	100	300	400	2000	3000	6000
Dzero	400	1600	2400	3200	6400	8000
BteV	20	40	100	200	300	500
DESY	100	180	210	240	270	300
Total Bandwidth	1070	3020	4810	8440	13870	22800
U.S.-CERN BW	155–310	622	1250	2500	5000	10000

Table 2.2.
Summary of Bandwidth Requirements at HEP Labs and on Main Transoceanic Links

	2001	2002	2003	2004	2005	2006
SLAC	OC12	2 X OC12	2 X OC12	OC48	OC48	2 X OC48
BNL	OC12	2 X OC12	2 X OC12	OC48	OC48	2 X OC48
FNAL	OC12	OC48	2 X OC48	OC192	OC192	2 X OC192
U.S.-CERN	2 X OC3	OC12	2 X OC12	OC48	2 X OC48	OC192
U.S.-DESY	OC3	2 X OC3	2 X OC3	2 X OC3	2 X OC3	OC12

The levels of usage of the main links have been increasing rapidly, tracking the increased bandwidth on the main transoceanic links. Large-scale data transfers using U.S.-CERN bandwidths in the range of 20–100 Mbps were observed to be increasingly common in 2001 for BaBar, CMS, and ATLAS. Up to 1 terabyte per day was observed for BaBar, requiring the use of most of a 155 Mbps link for long periods. These high-speed transfers were made possible by the quality of the links, which in many cases are nearly free of packet loss, combined with the modification of the default parameter settings of TCP and the use of parallel data streams[5].

Network Issues

While the rapid developments shown above are encouraging, there are a number of key issues to be addressed if the networking needs of the U.S. (and worldwide) HENP community are to be met over the next five years:

?? The development of regional and local network infrastructures to allow high performance end-to-end, at speeds (throughput) from 100 Mbps now to 1 Gbps and above for individual data flows within the next few years. Campus networks, regional networks, and shared national and continental infrastructures that are not optimized for high performance use are possible areas of concern.

?? The development of tools that allow physics groups to tune the network protocols and computer system settings to achieve high network performance. The wide deployment of such tools is an achievable near-term goal of HENP in the speed range of 100 Mbps. Scaling

---

[5] See http://www-iepm.slac.stanford.edu/monitoring/bulk/ .

up to the Gbps range will require exceptionally low error rates on the links, flawless configuration and operational monitoring of routers and switches, and/or new developments of resilient network protocols[6] that are matched to these speeds.

?? The widespread deployment of high performance data flows across shared networks is a new concept, and quite different from the principles that led to the scalable Internet that has grown over the last 10–15 years. Such use by HENP and other scientific fields will entail new concepts of fairness and new modes of network operation, monitoring and management[7].

2.2 Climate[8]

Climate science will address important areas of climate system research. In particular, it is aimed at understanding and predicting the climate system. The long-term goals are simple but ambitious. They are:

?? to develop and to work continually to improve a comprehensive climate system model that is at the forefront of international efforts in modeling the climate system, including the best possible component models coupled together in a balanced, harmonious modeling framework

?? to make the model readily available to, and usable by, the climate research community at a number of centers nationally and internationally, and to actively engage the community in the ongoing process of model development

?? to use the model to address important scientific questions about the climate system, including global change (Figure 2.2) and interdecadal and interannual variability

?? to use appropriate versions of the model for calculations in support of national and international policy decisions.

We anticipate many important changes in the climate modeling enterprise over the next five years, including:

?? increasing computer power, both in the U.S. (e.g., NERSC and NCAR) and abroad, that can support more elaborate and more sophisticated models and modeling studies, using increased spatial resolution and covering longer intervals of simulated time

?? improved understanding of many of the component processes represented in the model, including cloud physics; radiative transfer; atmospheric chemistry, including aerosol chemistry, boundary-layer processes, polar processes, and biogeochemical processes; and the interactions of gravity waves with the large-scale circulation of the atmosphere

?? improved understanding of how these component processes interact

?? improved numerical methods for the simulation of geophysical fluid dynamics

?? improved observations of the atmosphere, including major advances in satellite observations.

---

[6] The standard network protocol TCP is optimized for shared network use by a large community, and not for high throughput data transfers. Individual data flows in the Gbps range over long-range networks using TCP requires packet loss rates that approach the limit of the optical data link error rates.

[7] This realization is new, and little has been written about it yet. The need for a solution is heightened by the rapid development of Grids, which implicitly assume that adequate networks of quantifiable high performance are available on demand for priority tasks.

[8] This section provided by Al Kellie (NCAR)

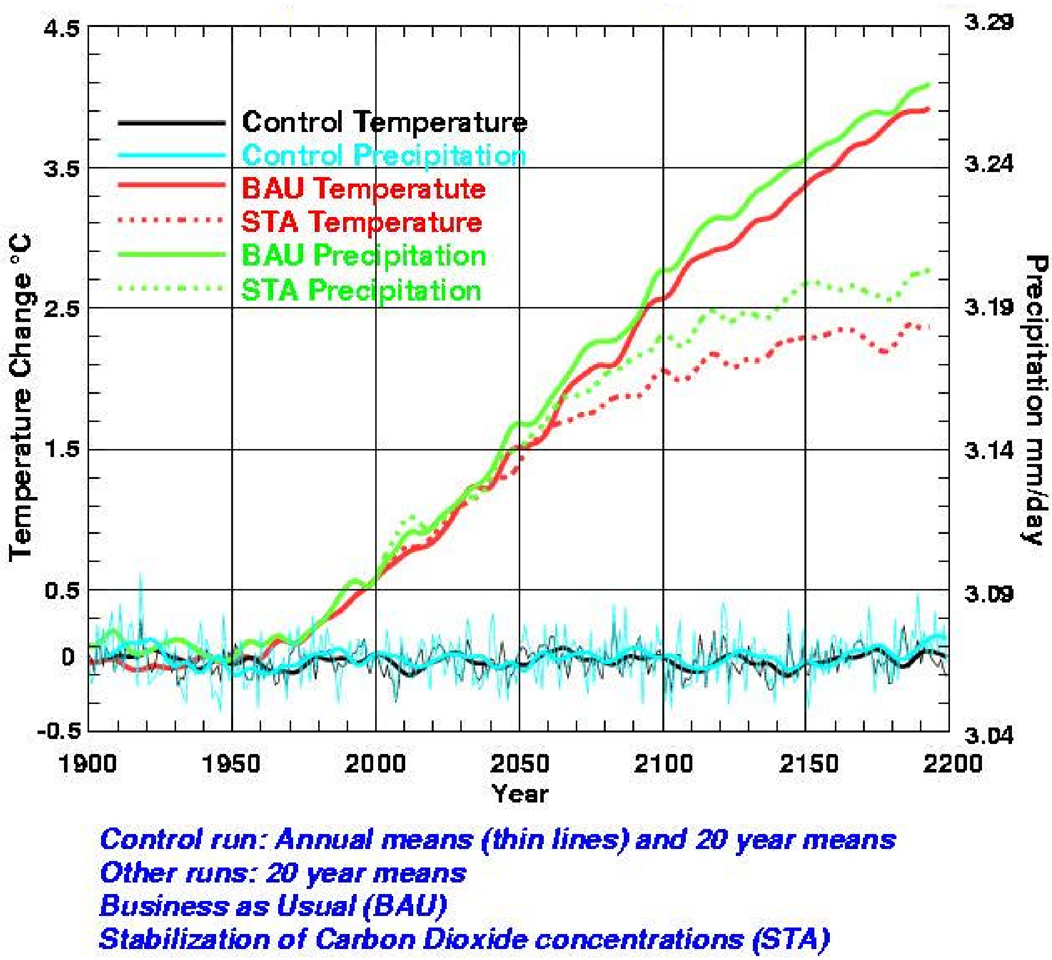


Figure 2.2. NCAR Climate Model (PCM) global temperature and precipitation changes simulated through the year 2200.

Over the next year, a new version of the Community Climate System Model, CCSM-2, will be introduced to the community of U.S. researchers. It is expected to produce improved simulations of the mean climate and climate variability and to have reduced deep-ocean drifts. Once this has been achieved, then an extended, multi-century simulation of the recent past equilibrium climate will be produced. The data will be made available to the CCSM community so that they can compare the new model simulations with previous simulations. These experiments will drive the needs for high bandwidth required to support data transfers flowing out of these major experiments. Many of the experiments will be decentralized on computers, since no single center has the capacity to provide the cycles needed to support some of these experiments. The community of researchers across the United States are committed to analyze the outputs of these scenarios, but it is essential that huge amounts of data be exchanged or gathered in order for many of the intercomparisons to be completed.

While it is important that the CCSM-2 continue to be used to study anthropogenic climate change, it is also being readied to perform a new climate of the 20th century experiment, the results of which will be compared to those produced from previous versions. The next model will run ensembles of simulations, using scenarios developed by the IPCC and others. CCSM-2 will contribute to the next National Assessment of Climate, due around 2004, and to the next IPCC report, due in 2005.

We can anticipate that the CCSM-2 will be used for some new types of experiments. The Biogeochemistry Working Group at NCAR, for example, has begun planning the Flying Leap Experiment. In it, fossil fuel carbon emissions will be specified; carbon will be actively advected through the system, dissolved in and released from the ocean, and taken up by the land surface; and atmospheric concentration of carbon will be determined as a residual of these interactive processes. How well the modeled CO_2 concentration in the atmosphere resembles observations will depend on the model components being developed. It is clear that this type of experiment will require a lot of model development and testing. It seems likely that the first experiment will require refinement and further model development and that subsequent experiments will be necessary to answer questions about the carbon budget. This work will likely continue throughout the next five years.

The next five-year period will be characterized by increased model complexity and capability, with the model being used for more experiments that have not yet been attempted. For example, these could include studies of recent climate change due to observed anthropogenic change in land surface properties, or climate change and its consequences for ecosystem succession. Exactly which experiments will be performed depends on the rate of model development and validation and the availability of computer time.

To come to grips with models simultaneously running at NERSC and NCAR and even the Earth Simulator (ES) in Japan, many tuning experiments on the order of a few days will be needed to tune the experimental design in order to have sufficient data coming out of such runs to render the analysis worthwhile. Otherwise, there is a risk that the data will overwhelm the cost of experiments to such a degree that the computations might be wasted.

The following example of climate science using of the earth simulator facility for a one-month period each year puts the data transfer issue and need for experimentation into perspective. The Earth Simulator is a 40 teraflop theoretical peak computer, built by NEC in Japan, that is available for international cooperation projects in the climate domain. It is comprised of 640 compute nodes, 8-way 8 gigaflop processors, and 65 interconnect nodes. The performance factor enhancement running on the ES is 50 times that of the NCAR system. The current production rate for data from climate science production runs is 12–20 terabytes per month. Using the ES, the potential data issue, namely staging the data back to the U.S. for analysis, is 600 terabytes for every month of experimentation. Climate science needs to experiment with high performance networking capabilities and advanced testbeds to refine the data transfer capabilities and ultimately influence the experimental design in order to maximize the impact of such climate science on policy development.

2.3 Data-Driven Astronomy and Astrophysics[9]

Motivation

Technological advances in telescope and instrument design during the last ten years, coupled with the exponential increase in computer and communications capability, have caused a dramatic and irreversible change in the character of astronomical research.

[9] This section is based on material from the Virtual Observatories of the Future conference (http://www.astro.caltech.edu/nvoconf/) and from the National Virtual Observatory white paper, also at that location. Julian Borrill (LBNL/NERSC, JDBorrill@lbl.gov) and Paul Messina, (CalTech, messina@cacr.caltech.edu) also contributed to this section.

Formerly, individual astronomers requested observing time on an instrument in order to study a few specific objects or a small region of the sky. Today, the instruments are so big and expensive that this is not practical. This has lead to a paradigm shift in how astronomy is being done, and at the same time it has vastly expanded the potential for new and discovery-based astronomy.

Large-scale surveys of the sky from space and ground are being initiated at wavelengths from radio to X-ray, thereby generating vast amounts of high-quality data. These surveys are creating catalogs of objects (stars, galaxies, quasars, etc.) numbering in billions, with up to a hundred measured parameters for each object. Yet this is just a foretaste of the much larger data sets to come.

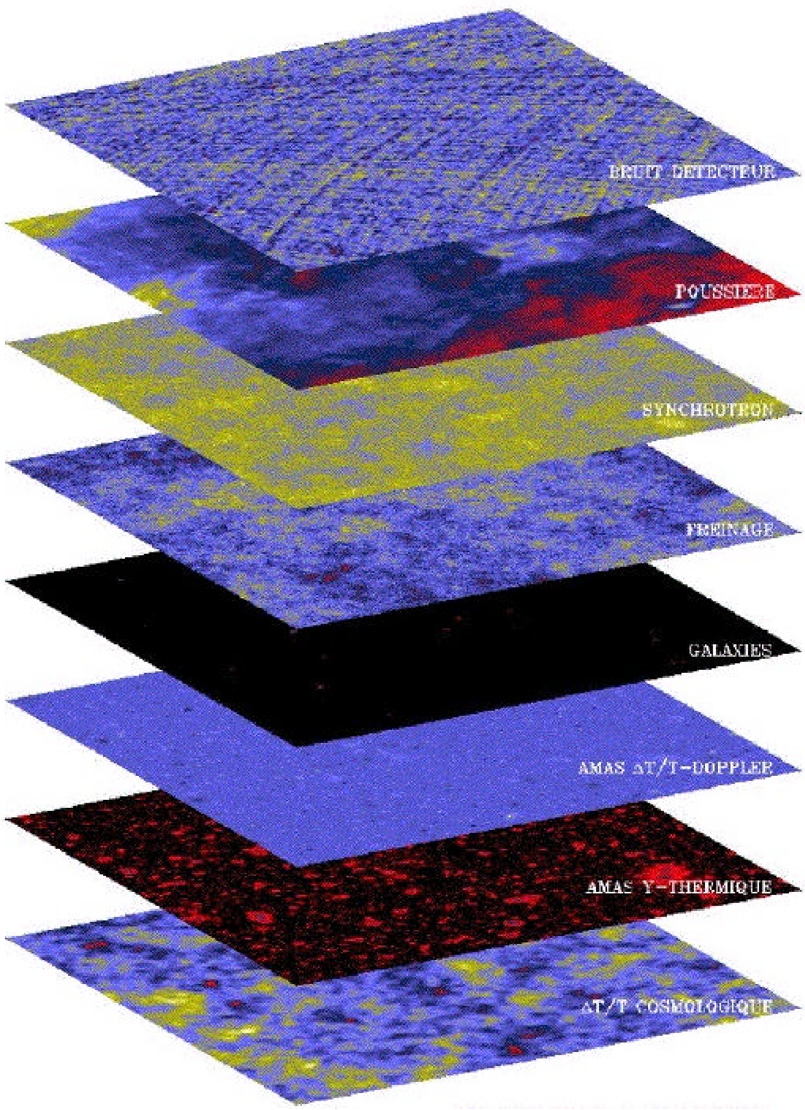
New instruments are being run all the time, taking as many observations as possible, over as much of the sky as possible. The astronomy is being done on the collected data sets rather than through direct use of the instrument. Further, this mode of operation allows for an unprecedented simultaneous analysis of high-quality observations from many instruments with different characteristics observing the same part of the sky. This has already led to some important science results that would not have been possible with single instrument observation.

This vast amount of new information about the universe, now measured in terabytes and soon in petabytes (e.g., when several proposed and recommended instruments are built), is enabling and stimulating a new way of doing astronomy. The potential for scientific discovery afforded by these new surveys is enormous. Entirely new and unexpected scientific results of major significance will emerge from the combined use of the resulting datasets, science that would not be possible from such sets used singly.

This new paradigm will enable tackling some major problems with an unprecedented accuracy. High-quality coverage over large parts of the sky in multiple wavelengths will provide data on billions of objects, and will allow discovery of new phenomena (from the analysis of statistically rich and unbiased image databases) and understanding of complex astrophysical systems (through the interplay of data and simulation). It will permit the discovery of rare objects (e.g., at the level of one source in 10 million) that may well lead to surprising new discoveries of previously unknown types of objects or new astrophysical phenomena, and it will permit the multi-wavelength identification of large statistical samples of previously rare objects (brown dwarfs, high-z quasars, ultra-luminous IR galaxies, etc.). This large coverage, periodically repeated, will allow cross-identification of "unidentified sources" (e.g., using radio, optical, and IR surveys to identify serendipitous Chandra X-ray sources), and it will allow identification of targets for specific spectrographic follow-up, as is done in supernova cosmology. The data will also provide for mapping of the large-scale structure of the universe.

Periodic re-surveys will allow for the discovery of objects and phenomena that change on observational time scales. Given that human observational time scales are minuscule on a cosmic scale, these events tend to represent something fairly dramatic. Examples include near-Earth asteroids, supernovae, gamma ray bursts, pulsars, etc.

Another class of query uniquely enabled by the multi-instrument sky surveys, and of direct relevance to DOE's mission in understanding the fundamental structure of matter, will be searches for information at all wavelengths on a particular region of the sky. As astronomers attempt to detect fainter and fainter signals, such searches will become increasingly important. For example, the spectrum of anisotropies in the polarization of the cosmic microwave background radiation (Figure 2.3) is sensitive to gravitational wave emission during the inflation of the universe, and hence probes physics at the Grand Unified Theory energy scale—energies beyond the capability of any imaginable accelerator. However, this signal is extremely faint and as yet undetected. Obtaining such a measurement will require detailed understanding of possible foreground sources. Joint searches of the surveys encompassed in projects like the National Virtual Observatory would help astronomers both to select regions of the sky with as little contamination as possible in advance of an observation, and to characterize the location and spectral dependency of whatever sources there were afterwards.



An example of the signal sources in an observation of the Cosmic Microwave Background - detector noise, dust, synchrotron, free-free, galaxies, kinetic Sunyaev-Zel'dovich, thermal Sunyaev-Zel'dovich, and the CMB itself. Understanding the impact of each of these on the total observation requires high quality data at a range of frequencies from 10GHz to 1000GHz.[10]

Figure 2.3. The cosmic microwave background power spectrum supports the model of a flat Universe.

The previous few paragraphs illustrate the types of scientific investigations that were not feasible with the more limited datasets of the past: We are at the start of a new era of information-rich astronomy. Large digital sky surveys and data archives are becoming the principal sources of data in astronomy. The very style of observational astronomy is changing: systematic sky surveys are now used both to answer some well-defined questions which require large samples of objects, and to discover and select interesting targets for follow-up studies with space-based or

[10] Image courtesy Julian Borrill, NERSC.

large ground-based telescopes. However, this vision relies completely on well-developed and highly capable software, computing, and networking infrastructure.

The Technical Challenges

This great opportunity comes with commensurate technological challenges: how to manage, search, combine, analyze, and explore these vast amounts of information, and to do it quickly and efficiently. We know how to collect many bits of information, but can we effectively refine the essence of knowledge from this mass of bits?

For example, the current data production rate of Hubble Space Telescope is about 5 gigabytes per day; but a facility recently recommended for construction by the AASC decadal survey[11]—the Large-Aperture Synoptic Survey Telescope (LSST)—could produce up to 10 terabytes per day.

The concept of a virtual observatory thus emerged. A virtual observatory would be a set of federated, geographically distributed, major digital sky archives, with the software tools and infrastructure to combine them in an efficient and user-friendly manner and to explore the resulting datasets, whose sheer size and complexity are beyond the reach of traditional approaches. It would help solve the technical problems common to most large digital sky surveys and optimize the use of our resources.

In order to realize the potential of this new paradigm, new tools and techniques are needed. For example, we can search local datasets for objects of certain characteristics, but the real gain will come when we can search over all datasets that have observations of the same part of the sky, but with different sensors, so that we can correlate many observations of the same physical object. This sort of global cross-correlation will entail a central analysis site farming out the requests for sub-regions or characteristic phenomena, and then pulling all of the results back to a large-scale facility to do the cross-correlation studies of all the available data. This requires not only catalogues that allow such searches of data locally, but analysis centers that have the network bandwidth to collect all of the data, and the computational capacity to do the analysis.

These large survey programs will produce coherent blocks of data obtained with uniform standards, and with the amount of data often measured in terabytes. This paradigm shift has been made possible not only by the increased capabilities of the new facilities that permit much faster acquisition of data, but also by the availability of computational hardware and software, and communications that make it possible to acquire, reduce, and archive this data.

Given the circumstance of archival-quality data from all branches of astronomy, physically distributed at 10–20 major archive centers and any number of ancillary datasets, together with a distributed community of thousands of scientific users, one can define what new software, computing, and communications infrastructure will be required.

The large dataset size and the geographic distribution of users and resources present major challenges in network bandwidth between the large repositories, the supercomputers used for data analysis and simulation-observation comparison studies, and the users. Hundreds of significant queries a day are expected. If such studies each extract 1% of multiple petabyte datasets (not unreasonable in several types of studies noted above, such as the LSST), then tens

[11] The National Research Council has established the Astronomy and Astrophysics Survey Committee (AASC) under the auspices of the Board on Physics and Astronomy (in cooperation with the Space Studies Board). "The committee will survey the field of space- and ground-based astronomy and astrophysics, recommending priorities for the most important new initiatives of the decade 2000–2010." http://www.nationalacademies.org/bpa/projects/astrosurvey/

to hundreds of terabytes of data a day will flow between archive, analysis, and user sites. This would require the 20 major data sites to have better than 10 Gb/s network access in order to keep up with the expected request rate.

Another scenario, proposed by astronomers who search for rare objects, suggests comparable bandwidths based on timeliness of scientific study. The anticipated study process involves pre-fetching 1 TB sized chunks of data from several (three) archives in order to do a study. If one user does this, and would like the data to arrive at the analysis site in about three hours, this requires 1 TB/hour, or 2.2 Gb/s. Assuming that only a few astronomers are doing such studies each day, one can estimate that 10 Gb/s links are needed between major archives and computing resources.[12]

Even allowing for intelligent server-side software in order to make the most efficient use of the network when interacting with end-users, it will be essential for the major virtual observatory data centers and analysis systems to be interconnected with very high speed networks.

2.4 Life Sciences

Most contemporary life sciences programs do not have the appetite for bandwidth and the requirement for reduced latency that other disciplines have acquired. Even the human genome project and its derivative functional genomic programs have comparatively modest datasets. Nevertheless, future life sciences programs, as envisioned in the Genomes to Life (G2L) initiative, will have rapidly increasing network demands. The future of biological research will increasingly focus more on *systems biology* rather than individual protein or gene research. The latter are typical of investigator-initiated programs usually found on a university campus. In contrast, the DOE Office of Science programs at the national laboratories are focused on high-throughput, multi-laboratory projects that generate large datasets that will need to be shared among several national laboratories. Those programs will be driven, in part, by the need for shared resources such as DNA sequencing at the Joint Genome Institute (JGI), mass spectroscopy at PNNL, structural analysis (crystallography) at the synchrotrons located at LBNL and ANL, and transgenic mutation analysis at ORNL. Each site may produce a dataset required by several of the other sites. Indeed, the first round of G2L awards requires multi-national-laboratory collaboration. With a principle laboratory and multiple collaborating laboratories, the G2L program mirrors the SciDAC program.

Understanding systems biology is the challenge for the next generation of biologists (*Science*, March 1, 2002). The knowledge of a single type of data will not provide the systems-level understanding required by new large-scale biology. This will be particularly important for metabolic, cellular, and tissue-level computer modeling. The different datasets required for such comprehensive models are distributed over many research sites. Real-time access to these various databases is necessary for meaningful execution of models. For example, modeling of microbial communities involved in environmental remediation might require genomic sequencing data from JGI, total proteomic data from mass spectroscopy or NMR at PNNL, and structure-function analysis from the synchrotron light sources at LBNL and ANL. While the size of each dataset might not be challenging, the number of sites that might simultaneously require those data could present networking challenges. In this context, expanded bandwidth and reduced latency will enable collaborations and transmission of datasets that would, otherwise, not be possible.

---

[12] A scenario described to Paul Messina by George Djorgovski, a Caltech astronomer who does searches for rare objects.

The G2L program will focus initially on the integration of proteomic and genomic data of microbial systems. Because bacteria are amenable to experimental manipulation, the datasets could increase in size and complexity as a function of manipulation of culture conditions, such as time and nutrients. The DOE interest in understanding the interaction of microbial communities rather than the function of a particular species in isolation further compounds the complexity of the datasets. Because approximately 40% of the earth's biomass is microbial, the curation and transfer of unique microbial datasets will be a daunting task.

Computationally modeling the complete physiology of a simple bacterium remains a challenge. Adapting such models to communities or complex mammalian systems will require sharing of additional datasets that have not yet been defined. Multiple genome comparison is an active research area. Combining these comparative data with whole proteomic and environmental data in a model will require seamless networking capabilities across several laboratories.

The foregoing discussion has been focused on the potential for biological research, particularly that supported by Office of Science, to generate and exchange large datasets. The following example describes existing large datasets and facilities that would benefit from greatly expanded network capabilities. Structural determinations from crystallographic diffraction data may only be applicable to less than half of the known proteins. Even with robotic crystal growth systems, most expressed proteins may not crystallize. This is especially true of membrane proteins and large multi-protein complexes. Resolution of these multi-protein structures is an essential requirement of systems biology, because such complexes embody much of the regulatory machinery of a cell. Recently, a joint DOE/NIH-funded cryo-electron microscopy center (R. Glaeser and E. Nogales, University of California, Berkeley) has resolved structures of large proteins that were not crystallized. In this "single particle" analysis, on the order of 10^6 to 10^7 particle images, automatically acquired, are analyzed. Each image is a randomly oriented two-dimensional rendition of the structure. Currently, the analysis of these images requires approximately one NERSC day of computation. A pre-selected dataset could be as large as one million particles, each represented as 300-by-300 pixels, 12 bits deep, for a total of something like 10^{11} bytes (0.1 terabyte). The raw dataset, i.e., digitized electron micrographs from which one then has to select the particles, might consist of 10^4 micrographs, each digitized as arrays of 6,000-by-10,000 pixels (about 0.6 terabyte). The predicted upper bound on the size of the datasets would be 1 terabyte for already selected particles, and 10 terabytes for the raw, not yet processed micrographs. Two scopes are already installed at this one center. Within the next five years, two to three scopes could be running simultaneously. Collaborations with other laboratories might involve remote instrument management that requires efficient video-rate data transfer at some stage during setup, and then constant data flow for several hours.

This same cryo-EM technique has been used for cellular tomography. While the datasets might not be as large as single particle analysis, the potential for remote instrument control is greater. Determining the precise cellular location of a molecule (recognized as a result of single particle analysis) might require repositioning the field a number of times. If done remotely, as presently planned, image sets need to be interactively shared with collaborating sites. Thus, bandwidth and latency could become limiting factors in the advancement of new field.

Quantitative and predictive biology involves complex systems analysis. The incorporation of structure-function data with kinetic data is a minimum requirement for computational modeling to evolve from a precise descriptive tool to become a tool for discovery.

3 High Performance Production Network (HPPN)

As computing and communications requirements of the DOE's programs continue to increase, improved High Performance Production Network capabilities are required to provide higher levels of network performance, to support wider network availability, and to enable use of more sophisticated applications throughout the DOE scientific facilities. Note that the High Performance Production Network is a generic term described in this section. The HPPN incorporates the majority of the current ESnet Program.

The goals of the HPPN are:

1. A high performance, science driven, production network with a targeted set of applications and facilities.
2. Increased support for end-to-end performance.
3. Directly involved science application advocates and network R&D advocates.
4. Advanced middleware services for high performance distributed applications and collaborations.
5. Broadened security focus.
6. Rich connectivity to DOE collaborators.
7. Additional focus on performance and service metrics.

These goals are discussed below, along with strategic approaches to achieving them.

Goal 1: A high performance, science driven, production network with a targeted set of applications and facilities.

The primary focus of HPPN will be high impact science. The high impact science projects and associated applications should be determined in accordance with an assessment of DOE science priorities.

Approach: The HPPN will provide focused support to the approximately 30 sites and 50–60 major projects or facilities that are strategic to DOE's success.

To facilitate high impact science, the HPPN will focus on the network capabilities and auxiliary services needed to enable the implementation and operation of high performance, high bandwidth, distributed applications, and their access to high end DOE computing, data, and instrument resources. This will include a reallocation of resources to the core science network from any commodity services that can be provided more effectively by other means.

Goal 2: Increased support for end-to-end performance.

As network-based applications grow in scope and complexity, network problems and performance issues increasingly require end-to-end (application to LAN to WAN to LAN to application) analysis to resolve.

Approach: Provide additional testing facilities and services, including measurement infrastructure, assistance in debugging and tuning applications' use of the network, and assistance in identifying and addressing local site performance issues.

In order to ensure that applications can make the best use of the network capabilities, and to continually assess the current and future needs of the high-end applications, HPPN staff will work directly with the high-end science projects, local DOE site staff, and university campus networking communities that support those projects.

Goal 3: Directly involved science application advocates and network R&D advocates.

A vital component of the success of the production networking effort has been its emphasis on network community involvement. Expanding this involvement into the research and applications communities will also help ensure success in a more effective and timely delivery of networking to science applications.

Approach: Actively solicit input from application developers and network researchers as to the directions and issues of the network.

HPPN staff will function as advocates and focal points for the high impact projects within the HPPN in order to ensure a two-way flow of information.

Goal 4: Advanced middleware services for high performance distributed applications and collaborations.

As collaboration environments and distributed science applications in DOE grow in capability and complexity, the applications have a growing need for middleware services, as well as the basic networking services, if environments such as the Grid are to become production quality and support these activities.

Approach: Work with the distributed applications and computer science community to determine what advanced service infrastructure is needed to support distributed applications.

ESnet has established a long history of trust within the DOE community, as well as a track record for providing successful services beyond networking infrastructure. Accordingly, it is well positioned to provide central or "core" enabling middleware infrastructure and services to the community as the needs evolve.

An example is that ESnet is currently participating in a project to provide both directory and PKI core services to the DOE Grid communities, e.g., DOE's HENP Grid and the DOE Science Grid. The early phase of this project will help to define the approach for this new set of core services as well as a model for related infrastructure and services.

Goal 5: Broadened security focus.

ESnet clearly delineates between security responsibility for its assets and those assets of the sites connected to it. However, in the light of the increasing number and severity of cyber-attacks on both computing and network resources and the new threat of more massive cyber-terrorist attacks, the HPPN network must provide mechanisms for assisting sites with their protection, as well as protect itself, from the broad spectrum of possible attacks.

Approach: A national emergency response capability that will form the first line of defense from a major cyber-attack.

This would anticipate a broad scale attack on the Internet in general, and HPPN sites in particular, possibly as part of a physical attack. The implementation would provide a new level of protection for the HPPN as well as for connected sites. This includes possible deployment of a wide-area intrusion detection system capability, potentially based on the LBNL's BRO, which has been successful in protecting high speed networks and helping provide forensic information to sites.

Goal 6: Rich connectivity to DOE collaborators.

Large-scale science in DOE is heavily based on collaborations, on both a national and international basis. Academic institution and global connectivity, commensurate with program requirements, will need an ongoing emphasis.

Approach: The national and international scope of DOE science requires global network connectivity.

The HPPN will continue to emphasize broad interconnectivity to other networks, academic, R&D, and commercial, on a national and international basis, driven by programmatic requirements.

Goal 7: Additional focus on performance and service metrics.

Measuring both internal performance and client service with an appropriate set of metrics will help demonstrate the success of the program as well as serve as an early warning for areas that need attention and improvement. Metrics will be coordinated with the scientific requirements.

Approach: Metrics will be established to provide measures of network bandwidth, latency, reliability, security, and utility from the user viewpoint.

An increased emphasis on measurements and metrics will enable intelligent and effective use of high-performance networking by applications, and to ensure continuous improvement of the HPPN.

4 Advanced Scientific Applications Pilot Network (ASAPN)

Pilot networks are key to the long-term success of DOE science and the vitality of the High Performance Production Network. Advanced Scientific Applications Pilot (ASAP) Networks will be application community focused. They bridge the gap between technology-focused network testbeds and key science application areas that have requirements for, *and* can benefit from, early access to very high bandwidth and advanced services not yet deployable in full production. The ASAP Networks each concentrate on a small number of key, knowledgeable users and projects and enable successes not possible otherwise.

As noted in Section 2, DOE has immediate science-driven needs that require the deployment of networks with advanced services and point-to-point bandwidth expected in the next generation and beyond of the HPPN (which today means ASAPN bandwidths of 2.5 to 10 Gb/s). However, not all HPPN clients need such services or bandwidth. In order to enable the high-end applications, but not over-engineer the entire network, DOE will deploy and coordinate ASAP Networks to a limited number of sites to support some key science themes. This provides a cost-effective solution for the breadth of DOE science.

There may be different ASAP Networks, either in parallel or for different time periods. Each would focus on a different scientifically driven area or "theme." Different ASAP Networks may exist at a given time, with DOE choosing to focus on different themes at different times (as with the NERSC "Big Splash" projects). But if resources allow, there could be more than one ASAP Network at a time, each with a different theme.

The goals of ASAP Networks are:

1. Science driven: The ASAPN drivers are very high bandwidth, distributed science applications.
2. Integrate the development of very high performance distributed science applications with their use of the very high bandwidth networks and advanced services.
3. Decrease the time it takes to make the successes of the network-oriented R&D available to bandwidth-intensive science projects.
4. Provide sufficient stability and services to attract key scientific applications and provide science teams with reasons to use the ASAPN.
5. Establish a thriving environment for network and application innovation.
6. Driven by network R&D.
7. Driven by network security R&D with the goal of ensuring that high speed networks of the future do not become channels to attack the DOE sites.
8. Driven by distributed systems middleware R&D.
9. Enhance ties to the academic and commercial communities.

These goals are discussed below, along with strategic approaches to achieving them.

Goal 1: Science driven: The ASAPN drivers are very high bandwidth, distributed science applications.

ASAP Networks first and foremost enable a coordinated set of very bandwidth intensive applications that cannot be accomplished without higher bandwidth and more advanced functions than HPPN provides at any point in time. These application themes may be just coming on-line and not yet fully functional, but require extraordinary capability as they develop their facilities. They may be precursors of permanent requirements, or they may be a limited project that will exist for a time and then end. Candidates probably include HENP, climate, and other applications.

Approach: Provide a high bandwidth IP network with new and experimental services for 4 to 6 key sites or facilities with bandwidth that is 4 to 16 times the bandwidth of the High Performance Production Network.

ASAP Networks will essentially be very high bandwidth, IP based networks, rather than evaluations of underlying technology that advanced technology testbeds provide. This frees the ASAPN to focus on enhancing scientific productivity for key areas. The goals above will be supported by a variety of network solutions that are deployable at the time the network is implemented.

ASAP Networks will support 4 to 6 facilities or sites. For the sake of discussion, two scenarios are used for the remainder of the paper; one from HENP and the other from climate. In reality, the 4 to 6 sites will be determined from DOE science projects that have the most compelling requirements and are ready to make the best use of the high bandwidth and advanced services of an ASAPN. Figure 5.1 shows the two example scenarios that ASAP Networks might support.

The first scenario involves DOE sites supporting major HENP accelerator-based experiments either with devices or by being major data repositories, simulation or analysis centers. In this example, four plausible sites could be BNL, SLAC, FNL, and LBNL. For climate research, a plausible set of sites that have significant DOE responsibilities for climate research could be ORNL, LANL, NCAR, and NERSC. Other science-driven scenarios are possible for astrophysics (with the National Virtual Observatory), fusion, materials, etc. ., potentially involving experimental computing resources or application projects at other sites such as ANL. Each scenario could have more than one useful combination of sites or facilities — particularly when geographic location is taken into account in design and cost tradeoffs. The point is that each ASAPN supports a single science theme at a given time and limit the number of sites to the most strategic or key sites of the theme. This allows engineering and operating the ASAPN to maximize the impact on the science themes in a cost-effective manner rather than having to optimize for multiple, competing requirements.

To succeed, ASAP Networks need the local sites to participate to provide connectivity to the actual science driver facilities and projects at a rate commensurate with the ASAPN bandwidth. ASAPN staff, in turn, will be committed to working with local site staff — both infrastructure and scientific — to help interface to the ASAPN and assure effective end-to-end results.

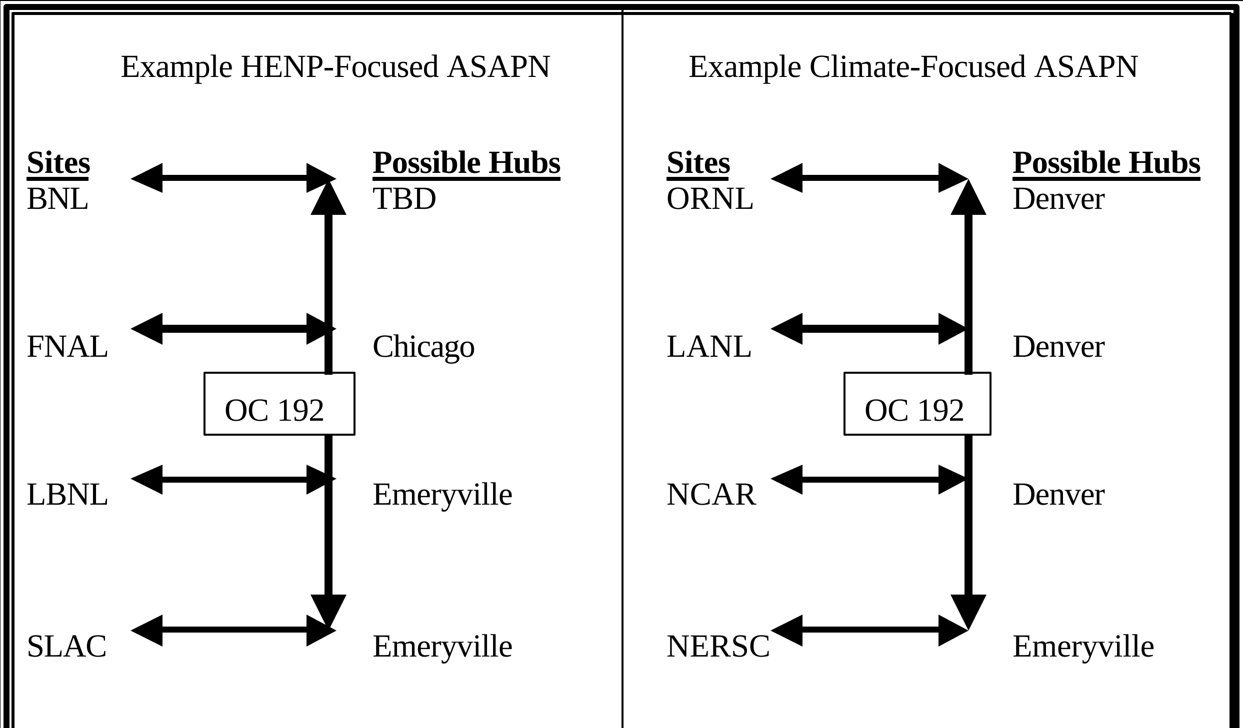


Figure 4.1. Two examples of ASAP Network application areas with possible sets of sites. On the left is a HENP-focused ASAPN, and on the right is a climate-focused ASAPN. Note that the climate example includes a non-DOE site that has significant DOE-funded efforts.

Goal 2: Integrate the development of very high performance distributed science applications with their use of the very high bandwidth networks and advanced services.

Promote the development of integrated high performance distributed science "Grand Challenge" teams that can provide a new vision for how to do science with very high speed networks and advanced distributed services, and then ensure that the vision is realized.

Approach: Build integrated science / computer science / network engineering teams to design, prototype, test, and tune very high performance distributed applications.

High performance distributed science "Grand Challenge" teams will be assembled from the application sciences, computer scientists who understand high performance middleware, and networking engineers who understand very high performance networks. These teams will take the application vision, generate requirements, identify or design supporting middleware, and optimize for the networks. The process will be iterative, with each specialty feeding information, results, and suggested modifications to all of the others.

Goal 3: Decrease the time it takes to make the successes of the network-oriented R&D available to bandwidth-intensive science projects.

A goal of ASAP Networks is to promote the enhancement and improvement of the HPPN. When technology and methods succeed in ASAP Networks, that is a primary indicator the technology will be useful and ready for implementation in the HPPN. Likewise, the expertise and experience gained in the ASAP Networks by working intensely with a subset of key DOE application areas will enhance the HPPN.

HPPN network engineers will influence the experiences and issues in the ASAPN. These issues will be areas the ASAPN focuses on, thus creating a two-way relationship that help both efforts thrive.

The ASAP Network will decrease the time it takes to make the successes of network R&D efforts available to the bandwidth-intensive science projects. ASAP Networks will be the most effective way of doing scaling and performance testing for new network technology once it is developed and proven in advanced network testbeds. Furthermore, the testing and evaluation can be done in collaboration with and benefit to real science applications in the pilot networks.

Approach: Serve as an environment to develop and test technology for the next generation High Performance Production Network.

ASAP Networks will provide technology development, expertise, and testing for the next-generation HPPN. While it is not valid to assume that whatever the last ASAP Network was becomes the next HPPN network (for example, a solution for four sites with a particular scientific application may not scale to the entire DOE portfolio), one can envision the HPPN leveraging both the ASAP Network infrastructure and possibly the ASAP vendor contracts to develop the next full HPPN.

On the other hand, the HPPN will have limitations and issues that have to be explored and dealt with in the ASAPN environment in order to make the HPPN successful. A close two-way relationship between the HPPN and ASAPN is key to the health of both efforts. ASAP Networks will enhance the flow of ideas for science, networking, and infrastructure through competition and sharing of ideas and experiences.

This will follow from the formulation of the ASPNs. The science teams have to propose what they will do with the capabilities of an ASAP Network, leading to creativity and efficiency of the science applications. Vendors will also have to present their ideas when proposing services for an ASAP Network.

Goal 4: Provide sufficient stability and services to attract key scientific applications and provide science teams with reasons to use the ASAPN.

In order to attract science users and enable them to achieve their expanded goals, ASAP Networks must have some level of advanced services. The ASAPN will also have some level of stability and persistence that is sufficient to enable productive science for early adapters.

Approach: Advanced services, stability, and persistence that is essential to attract science users will be engineered into the network

ASAP Networks will focus on certain types of science-related applications such as:

?? ultra-high data rate applications (e.g., HENP)

?? joint testbed facilities such as the Probe high speed archival storage testbeds at LBNL and ORNL

?? immersive collaboration facilities such as Access Grid / Cave and related advanced testbeds at ANL

?? high performance, distributed visualization

?? distributed computational algorithm R&D.

To support these, the ASAP Networks have to provide network and middleware services that are almost always available and stable. ASAP networks will push the limits of what can be achieved, and also evolve more rapidly than standard production networks. This requires selected sites to be more tolerant of change and work with the ASAP Network staff to implement new features in a timely manner. The key is that ASAP Network staff have direct and close relationships with the actual scientists so they can assist where necessary, but also so they have a timely feedback loop to balance the network issues with the scientific plan.

Goal 5: Establish a thriving environment for network and application innovation.

The ASAPN provides an environment where applications, middleware, and network R&D come together to solve specific, high performance science problems, and to enable innovation and new approaches in both areas.

Approach: Establish and improve infrastructure via competition of ideas.

It is possible, and maybe even likely, that the providers of the basic network fabric for ASAP Networks will differ from the one providing the HPPN. The presence of multiple vendors within the DOE network enterprise could provide diversity and competition in providing network level capacity. This could benefit not just the ASAPN but also the HPPN and other parts of the DOE network portfolio. The enhanced competition is not just for the fabric, but also for other parts of the network, such as routing equipment. On the other hand, it is possible that there would be benefits to having the vendors for the HPPN network supply the ASAPN. But even then, the shorter 3- to 4-year ASAPN timeframe before taking the next step will encourage a more robust and interesting set of possibilities for the HPPN.

Goal 6: Driven by Network R&D.

ASAP Networks enable network R&D by being more aggressive with the introduction of new technologies and functions. ASAPNs can do this in part because the coordination and integration effort for new technology is simplified when only a handful of sites are involved, not the entire DOE science community. Equally important, the clients of ASAP Networks will be applications and projects who benefit directly from the new technology, since they need the new technology to accomplish their mission.

Approach: ASAPN is a place R&D products can be evaluated with real applications and at realistic scales.

As the flow of successful experience and technology emerges from network technology testbeds (and other sources), the ASAPN will be a likely place to implement the technology. ASAP Networks provide a focused and dedicated environment to try out new technology with reasonable risk (the worst thing that can happen is that sites use the HPPN until the new technology is either corrected or backed out of the ASAPN). The ASAPNs provide two key attributes that will attract network researchers and developers: the new technology (1) will be demonstrated with real scientific applications and (2) will be used in a scale that is not achievable in a testbed.

An important component of making this and other goals achievable is that the ASAPNs be designed to assure that monitoring and internal information is provided to authorized participants. In order to make large scale applications operate properly and effectively on advanced networks, the scientific team members will have to have easily available and timely performance and monitoring information for their application data flows and the network in general. This information will also be key to network R&D participants in understanding how to improve the tools and services on the ASAP Network.

Goal 7: Driven by network security R&D with the goal of ensuring that high-speed networks of the future do not become channels to attack the DOE sites.

ASAP Networks will also be testbeds for network security R&D. The traffic on the ASAP Network will be less diverse and better characterized than the general network, giving an advantage in being able to understand abnormal traffic flows. Using the ASAP Network to ensure that high-speed networks, as they become more pervasive, do not become high-speed channels to attack the DOE sites is both achievable and necessary.

Approach: The ASAPN will serve as the testbed for the next-generation cyber-security infrastructure.

The increased bandwidth and new services used in the ASAPN will require innovations in cyber security. For example, intrusion detection systems and/or firewalls will have to perform at rates substantially above the HPPN, and one or more ASAP networks may be used to test such capabilities.. They will also have to be flexible enough to adapt to new service models, new middleware, and new application traffic patterns. These are all things that the HPPN will need eventually. Ongoing cyber security development is important to allowing the natural evolution of the HPPN and DOE science.

Goal 8: Driven by distributed systems middleware R&D.

For the most part, network technology testbeds do not provide a stable environment to develop, test, and deploy distributed systems middleware software. HPPN networks have too many service commitments to easily allow deployment and testing of newly developed middleware. The ASAP Network, with its select clients benefiting from the early use of the new technology, will provide the perfect environment to accelerate the development and use of such software to demonstrate the software's value to distributed DOE science.

By working with a limited number of high value science projects, ASAP Networks enable the easy introduction of advanced and/or experimental middleware services for high performance distributed applications and collaborations. The software could include collaboration-specific services or general network services. Once successfully demonstrated and polished for the applications in the ASAPN, the software should move into more general use in the HPPN.

Approach: Enable network and middleware R&D.

ASAP Networks will deploy network and middleware R&D products that are relied upon by the science collaborations. Many of these tools are key to enabling the high bandwidth applications. Some of these services are:

?? reconfigurable software and connections

?? facilities to gather data from inside the network

?? facilities to set up experimental situations (e.g., artificial loads, delays, etc.).

?? resource discovery services for applications

?? directory services supporting long-term homes and rooted namespaces for collaborations and virtual organizations

Experimental middleware services that are not yet ready for HPPN deployment (e.g., they may impact reliability, may not yet scale to full size, may be difficult to operate, etc.) will be available on the ASAPN to investigate distributed applications. This is particularly true if these applications have a potential for disrupting network services, thereby impacting other users of the network.

Goal 9: Enhance ties to the academic and commercial communities.

DOE science teams are composed of researchers at DOE facilities, universities, and other institutions. Likewise, ASAPN technology will come from university research communities as well as commercial organizations. The ASAPNs will serve as a beacon to attract collaborations and enhance them, and will become a key way in which academic communities participate with DOE in advanced network research and usage. The ASAPN also provides a training ground for some of the next-generation network engineers and application developers.

Approach: Peering and cooperation.

It is likely other national networks will coordinate and cooperate with an ASAP Network because their users will need access to some of the sites and services an ASAP Network provides for its science themes. ASAP Networks will peer with a limited but important set of national and international networks, particularly those who serve sites that are part of the science teams selected to use the ASAPN. It is possible that testbeds in non-networking areas, such as Probe, would also want to coordinate with an ASAPN. ASAPN could provide the opportunity to use heterogeneous equipment and to evaluate the equipment, not just in it own right, but how it compares and interacts. This could be an attraction for network R&D efforts from the commercial sector that will help contribute to the ASAPNs. Finally, ASAPNs would probably peer with the HPPN to provide access and backup.

5 Advanced Network Technology Consortium Testbed (ANTCT)

As demonstrated in the Science Drivers section, there are many areas of science where the data volumes and access requirements are increasing at a rate that will outstrip not only current networks, but current network technology as well. This is true not only for bulk data transfer, but also for collaborative uses of the remote instruments, interacting supercomputer simulations, etc.

The network requirements implied by this sort of data growth — most of which is analyzed at multiple institutions remotely from where it is generated, by an instrument or simulation — means that DOE Scientific Networking must not only address today's needs (HPPN), and the needs several years out (ASAPN), but must also devote effort to exploring the network technology that DOE science networks will need five, seven, or more years out.

This is the purpose of network technology testbeds, and DOE scientific networking staff must be involved in such testbeds in order to ensure that they are positioned to meet DOE's science networking needs in the future.

The goals of the ANTCT are:

1. Explore and influence future network technology.
2. Determine how to deliver dramatic increases in current capacity and capability to next-generation science applications and facilities.
3. Work closely with the end sites in order to achieve end-to-end high bandwidth.

These goals are discussed below, along with strategic approaches to achieving them.

Goal 1: Explore and influence future network technology.

As end-to-end (application to application) network data rates start to push towards multiple gigabits per second, it is becoming evident that both the transport protocols and the network architectures will likely have to change. Future networks will involve photonic switching of many optical channels, data units are likely to be switched instead of routed, new transport protocols will be used, etc. This will happen in order to be able to deliver effective application bandwidth and in order to manage the network itself. Potential technologies need to be explored while there is still time to influence them in directions needed to support future DOE science, and in order to gauge their impact on applications.

Approach: DOE participates as part of a national high performance network testbed consortium.

Advanced network testbeds are difficult and costly to build and maintain. The trend over the past decade is toward the cooperative development of these testbeds by consortia of government labs and universities. This will also be the approach for ANTCT.

?? Photonic switching of DWDM

One of the likely network architecture changes over the next decade will be a shift from multiplexed streams on single, very high bandwidth channels, to switched streams on large collections of lower bandwidth channels. Given the high cost of ultra-fast electronic/electro-optical switches and routers, the emerging technology that is likely to address this is photonic switching of the many optical channels created by dense wave division multiplexing (DWDM). The characteristics of this new architecture are likely to be quite different than those of current networks, and will have impact both on applications and on network management, with new and experimental switch control software, etc. Some part of the ANTCT should be configured as a photonically switched DWDM network so that this new architecture can be explored.

?? Many digital channels, fully switched and/or dynamically reconfigured

Until practical photonic switches are available, this same architecture can be realized, and therefore used for experimentation, by using digital multiplexing of single high bandwidth optical streams, and then using electronic switching of the multiple digital streams. In the first version of ANTCT, the switches are likely to be gigabit or 10 gigabit Ethernet switches operating on multiple, gigabit Ethernet channels. This should provide many of the same architectural characteristics as the photonically switched networks noted above and allow early exploration of the key architectural issues.

?? Opportunistic connection to DOE sites

Advanced network testbeds are almost always built with the cooperation and assistance of a small number of telecommunication companies. Such companies always have limited geographic coverage where they can provide access to the testbed links. The sites that have the opportunity to participate are usually those that happen to be close enough to the long-haul telecom facilities to make connection practical. Hence, participation in testbeds like ANTCT is frequently opportunistic rather than thematic, and governed by the chance coincidence of site proximity to telecom facilities that provide the testbed infrastructure.

Goal 2: Determine how to deliver dramatic increases in current capacity and capability to next-generation science applications and facilities.

With every major advance in network technology, applications have had to adapt their strategy for achieving high end-to-end throughput. The ANTCT testbed will provide an environment for a few applications to experiment with the next-generation protocols and network architecture. Realistic networks — i.e., complex testbeds — have repeatedly been shown to be critical to identifying the issues that applications will face in network environments that are different and/or much faster that the previous generation.

Another thing that typically happens with major shifts in network performance and architecture is that the bottlenecks identified and addressed in the past will be replaced by new bottlenecks in new places. The ANTCT network should provide a platform to start identifying and addressing the new bottlenecks. Realistic network infrastructure has also been shown to be critical to identifying these issues.

Approach: Explore technologies for delivering future higher bandwidths to applications and science facilities.

Network-based applications almost never behave the same when going through major changes in network bandwidth or architecture. Further, it is almost impossible to use the older generation of network to predict and plan for the application changes that will be needed for the new environment. We saw this when moving applications from the LAN environment to WAN environment, from low-speed WANs to high-speed WANs, and from low-delay, high-bandwidth terrestrial links to high-delay, high-bandwidth satellite links.

Successful network testbeds almost always have worked very hard to include a few prototype applications as part of the testbed in order to gauge and address the technology changes on applications. These applications rarely do "real" work due to the technology and reliability issues associated with testbeds, but are attracted to the testbeds by special funding for that purpose, and the lure of finding out what might be done in the next-generation networks.

Another characteristic of advanced network testbeds is that the attached end systems — the application platforms — must themselves frequently be modified in order to take advantage of the new network potential. New approaches to protocols, new operating system characteristics, new network interface hardware, etc., are typical.

Goal 3: Work closely with the end sites in order to achieve end-to-end high bandwidth.

Close cooperation of the end sites is essential in order to get the application kernels and prototypes connected to the testbed at very high bandwidth.

Approach: Include the site networking staff directly in the network testbed.

Experience has shown that one of the most effective ways to gain cooperation of the end sites is to include some of the site network staff in the testbed R&D team.

6 Advanced Services

The success of advanced distributed science applications depends critically on distributed middleware services in order to build realistic and reliable distributed applications. The emerging success of computational and data Grids is due to the fact that they supply middleware services for application security, resource discovery, advanced collaboration support, etc., that are essential for the practical construction, and reliable and scalable operations of distributed applications.

These services typically take the form of a small number of server systems and/or databases for which the science networking staff provides the persistence and management of the service infrastructure (e.g., long-lived servers). These high-value services greatly lower the barriers to building distributed application systems.

- **The science network staff would probably not get directly involved in providing the service due to the specialization and cost involved, but rather would run the enabling infrastructure.**

  The current ESnet effort in setting up and running a PKI Certification Authority server and Grid Information Service directory servers for all of DOE are examples. CA registration services that the users apply to are provided by the sites, and the Grid directory service content is determined and managed by the virtual organizations. ESNet manages the root servers.

- **ESnet, for example, is uniquely qualified to operate the infrastructure aspects of these critical services.**
  - ESnet can supply high-quality, persistent support over a long period of time.
  - Such services are applicable to large fractions of the DOE science community, and ESnet already has a customer base that includes most of DOE.
  - ESnet has the reputation of a neutral third party in these software-based services.
- **Advanced services can also be new capabilities in the network, e.g., bandwidth reservation, routing-based bulk data transport.**

7 Cost Estimates

Cost estimates at this point are extremely preliminary. The costs of the HPPN are derived from costs associated with ESnet. The costs for the ASAPN are crude estimates for the two different scenarios.

7.1 HPPN

Yearly Recurring Costs

BASELINE OPERATION

1) General Operations		$ 6.50M
2) International Communications		$ 1.20M
3) Video Conferencing		$ 0.35M
4) Domestic Communications		$ 7.00M
5) DOE DS3		$ 0.10M
6) PKI/DIR Grid Services		$ 0.25M
7) Equipment		$ 1.00M
TOTAL	$16.40M	

OC48 BACKBONE UPGRADE		$ 1.60M
	TOTAL	$ 1.60M

NEW SERVICES

1) 1 Project Leader		$ 0.30M
2) 2 Applications Specialists		$ 0.50M
3) 1 Network Eng. (measurements)		$ 0.25M
	TOTAL	$ 1.05M

HPPN TOTAL RECURRING $19.05M

7.2 ASAPN

The two network scenarios were priced using bandwidth estimates from Qwest and other costs are estimated. The bandwidth costs were very close, so the estimates below apply to either the HENP or the climate ASAPN, and are likely to be within reasonable margins with similar sets of sites. The estimates are sensitive to number and location of the sites selected for the ASAPN. Note that the staffing estimates do not necessarily assume that the ASAP networks will be operated by a single organization, and in fact a reasonable approach would be for a distributed team.

One-Time

6 - OC192 ROUTERS AND INTERFACES		$ 6.0M
TOTAL		$ 6.0M

Yearly Recurring Costs

OC192 BANDWIDTH	$ 3.0M	
TOTAL	$ 3.0M	
STAFFING		
1) 1 Project Leader	$ 0.30M	
2) 3 Network Engineers	$ 0.75M	
3) 2 Applications Specialists	$ 0.50M	
4) 1/2 Network staff per site	$ 0.50M	
TOTAL	$ 2.05M	
EQUIPMENT REFRESH	$ 2.0M	
TOTAL	$ 2.0M	
ASAPN TOTAL RECURRING		$ 7.05M

7.3 ANTCT

We understand DOE is expecting an independent proposal in this area that can serve as example costs for ANTCT type testbeds. LBNL is a participant in that proposal.