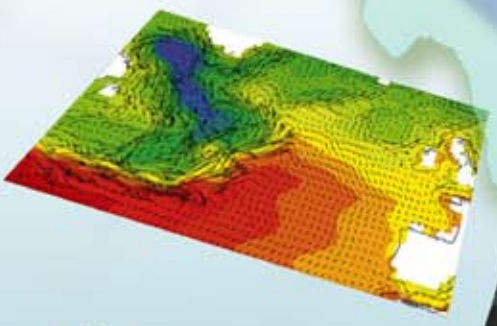
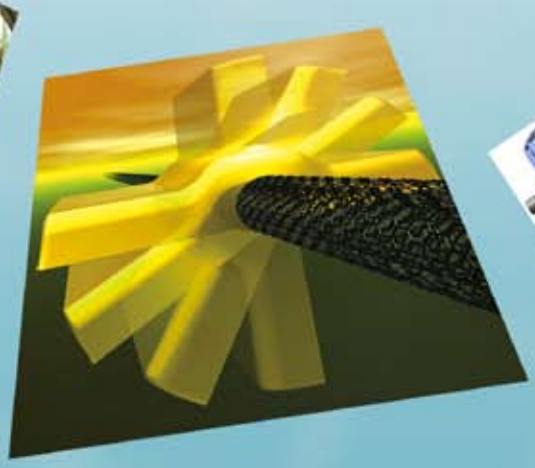
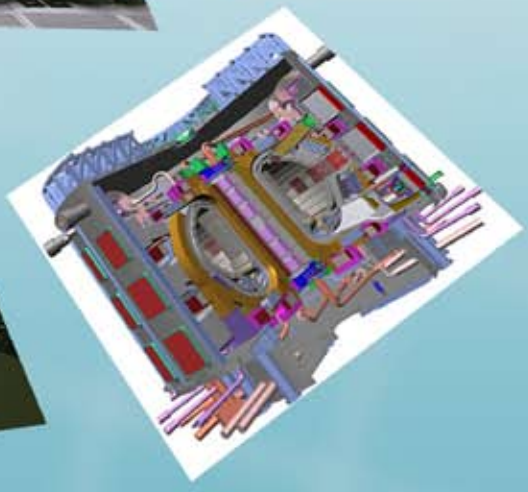
FES Science Network Requirements

ESnet
Energy Sciences Network

Report of the Fusion Energy Sciences Network Requirements Workshop Conducted March 13 and 14, 2008





BERKELEY LAB

Office of Science
U.S. DEPARTMENT OF ENERGY

FES Science Network Requirements Workshop

Fusion Energy Sciences Program Office, DOE Office of Science
Energy Sciences Network
Gaithersburg, MD – March 13 and 14, 2008

ESnet is funded by the U.S. Dept. of Energy, Office of Science, Advanced Scientific Computing Research (ASCR) program. Vince Dattoria is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This work was supported by the Directors of the Office of Science, Office of Advanced Scientific Computing Research, Facilities Division, and the Office of Fusion Energy Sciences.

Participants and Contributors

Rich Carlson, Internet2 (Networking)
Tom Casper, LLNL (Fusion – LLNL)
Dan Ciarlette, ORNL (ITER)
Eli Dart, ESnet (Networking)
Vince Dattoria, DOE/SC/ASCR (ASCR Program Office)
Bill Dorland, University of Maryland (Fusion – Computation)
Martin Greenwald, MIT (Fusion – Alcator C-Mod)
Paul Henderson, PPPL (Fusion – PPPL Networking, PPPL)
Daniel Hitchcock, DOE/SC/ASCR (ASCR Program Office)
Ihor Holod, UC Irvine (Fusion – Computation, SciDAC)
William Johnston, ESnet (Networking)
Scott Klasky, ORNL (Fusion – Computation, SciDAC)
John Mandrekas, DOE/SC (FES Program Office)
Doug McCune, PPPL (Fusion – TRANSP user community, PPPL)
Thomas NDousse, DOE/SC/ASCR (ASCR Program Office)
Ravi Samtaney, PPPL (Fusion – Computation, SciDAC)
David Schissel, General Atomics (Fusion – DIII-D, Collaboratories)
Yukiko Sekine, DOE/SC/ASCR (ASCR Program Office)
Sveta Shasharina, Tech-X Corporation (Fusion – Computation)
Brian Tierney, LBNL (Networking)

Editors

Eli Dart, ESnet Engineering Group – dart@es.net
Brian Tierney, LBNL – bltierney@lbl.gov

Contents

1 Executive Summary

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy Office of Science, the single largest supporter of basic research in the physical sciences in the United States of America. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 20 years.

In March 2008, ESnet and the Fusion Energy Sciences (FES) Program Office of the DOE Office of Science organized a workshop to characterize the networking requirements of the science programs funded by the FES Program Office.

Most sites that conduct data-intensive activities (the Tokamaks at GA and MIT, the supercomputer centers at NERSC and ORNL) show a need for on the order of 10 Gbps of network bandwidth for FES-related work within 5 years. PPPL reported a need for 8 times that (80 Gbps) in that time frame. Estimates for the 5-10 year time period are up to 160 Mbps for large simulations. Bandwidth requirements for ITER range from 10 to 80 Gbps.

In terms of science process and collaboration structure, it is clear that the proposed Fusion Simulation Project (FSP) has the potential to significantly impact the data movement patterns and therefore the network requirements for U.S. fusion science. As the FSP is defined over the next two years, these changes will become clearer. Also, there is a clear and present unmet need for better network connectivity between U.S. FES sites and two Asian fusion experiments – the EAST Tokamak in China and the KSTAR Tokamak in South Korea.

In addition to achieving its goal of collecting and characterizing the network requirements of the science endeavors funded by the FES Program Office, the workshop emphasized that there is a need for research into better ways of conducting remote collaboration with the control room of a Tokamak running an experiment. This is especially important since the current plans for ITER assume that this problem will be solved. .

2 Workshop Background and Structure

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy Office of Science, the single largest supporter of basic research in the physical sciences in the United States of America. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 20 years.

In March 2008, ESnet and the Fusion Energy Sciences (FES) Program Office of the DOE Office of Science organized a workshop to characterize the networking requirements of the science programs funded by the FES Program Office. These fusion programs included the DIII-D National Fusion Facility at General Atomics, the Magnetic Fusion Energy program at Lawrence Livermore National Laboratory, the MIT Plasma Science & Fusion Center and C-Mod Tokamak, Fusion Simulation Computing at Oak Ridge National Laboratory (ORNL), the Princeton Plasma Physics Laboratory (PPPL), and U.S. ITER Construction efforts at ORNL.

Workshop participants were asked to codify their requirements in a "case study" format. A case study includes a network-centric narrative describing the science being done. The narrative describes the instruments and facilities necessary for the science and the process by which the science is done, with emphasis on the network services needed and the way in which the network is used. Participants were asked to consider three time scales in their case studies – the near term (immediately and up to 12 months in the future), the medium term (3-5 years in the future), and the long term (greater than 5 years in the future). The information in the narrative is also distilled into a summary table, with rows for each time scale and columns for network bandwidth and services requirements.

3 DOE Fusion Energy Sciences Programs

Introduction

The mission of the Fusion Energy Sciences (FES) program is to support fundamental research for developing the knowledge base for a new and attractive form of energy based on the nuclear fusion process, the same process that gives rise to the energy of the sun and stars. In addition, FES supports research focusing on the underlying sciences of plasma physics, the study of the fourth state of matter which is the central component of magnetically confined fusion systems, as well as the emerging field of high energy density physics (HEDP)—the state of matter encountered in inertially confined fusion energy systems. Related work contributes to understanding astrophysics, geosciences, industrial low temperature plasma processing, turbulence, and complex self-organizing systems.

To carry out its mission, FES supports research activities involving over 1,100 researchers and students at approximately 67 universities, 10 industrial firms, 11 national laboratories, and 2 Federal laboratories, distributed over 31 states. These activities include efforts in experiment, theory, and advanced computation, ranging from single-investigator research programs to large-scale national and international collaborative efforts.

Major Facilities

At the largest scale, the FES program supports world-class magnetic confinement facilities that are shared by national teams of researchers to advance fusion energy sciences at the frontiers of near-energy producing plasma conditions. Each of the major facilities offers world-leading capabilities for the study of fusion-grade plasmas and their interactions with the surrounding systems.

The Department's three major fusion physics facilities are: the DIII-D Tokamak at General Atomics in San Diego, California; the Alcator C-Mod Tokamak at the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts; and the National Spherical Torus Experiment (NSTX) at the Princeton Plasma Physics Laboratory in Princeton, New Jersey.

The three major facilities are operated by the hosting institutions but are configured with national research teams made up of local scientists and engineers, researchers from other institutions and universities, and foreign collaborators.

In addition to the FES major facilities, a range of small innovative experiments at Universities and National Laboratories are exploring the potential of alternative confinement concepts.

ITER

U.S. participation in ITER is a Presidential Initiative to build and operate the first fusion science facility capable of producing a sustained burning plasma. The mission of ITER is to demonstrate the scientific and technological feasibility of fusion energy for peaceful

purposes. ITER is designed to produce 500 MW of fusion power at a power gain Q > 10 for at least 400 seconds, and is expected to optimize physics and integrate many of key technologies needed for future fusion power plants. The seven ITER parties (China, European Union, India, Japan, Russia, South Korea, and United States) represent over half of the world's population. The European Union is hosting the site for the international ITER Project at Cadarache, France. Through ITER, the FES program is pushing the boundaries in large-scale international scientific collaboration.

International Collaborations

In addition to their work on domestic experiments, scientists from the United States participate in leading edge scientific experiments on international fusion facilities in Europe, Japan, China, South Korea, the Russian Federation, and India—the ITER members, and conduct comparative studies to enhance our understanding of the underlying physics. These facilities include the world's highest performance tokamaks (the Joint European Torus [JET] in the United Kingdom and the Japan Torus 60 Upgrade [JT-60U] in Japan), a stellarator (the Large Helical Device [LHD] in Japan), a superconducting tokamak (Tore Supra in France), and several smaller devices. In addition, the United States is collaborating with South Korea on KSTAR, the Korean Superconducting Tokamak Advanced Research Project, and with China on research using the new long-pulse, superconducting, advanced tokamak EAST. These collaborations provide a valuable link with the 80% of the fusion research that is conducted outside the United States and provide a firm foundation to support ITER activities.

Advanced Computations

High Performance Computing has played an important role in fusion research since the early days of the fusion program. The National Energy Research Scientific Computing Center (NERSC) –the flagship scientific computing facility for the Office of Science—started as the Magnetic Fusion Energy Computer Center (MFECC).

Currently, most of the FES advanced computational projects are supported under the auspices of the Office of Science (SC) Scientific Discovery through Advanced Computing (SciDAC) program. The goal of the FES SciDAC projects is to advance scientific discovery in fusion plasma physics by exploiting the emerging capabilities of terascale and petascale computing and associated progress in software and algorithm development, and to contribute to the FES long term goal of developing a predictive capability for burning plasmas.

The current FES SciDAC portfolio includes eight projects which are set up as strong collaborations among 29 institutions including national laboratories, universities, and private industry. Of these, five are focused on topical science areas while the remaining three—which are jointly funded by FES and the Office of Advanced Scientific Computing Research (ASCR) and are known as Fusion Simulation Prototype Centers or proto-FSPs—focus on code integration and computational framework development in the areas of edge plasma transport, interaction of RF waves with MHD, and the coupling of the edge and core regions of tokamak plasmas.

The success of the FES SciDAC projects combined with the emerging availability of even more powerful computers and the need to develop an integrated predictive simulation capability for the needs of ITER and burning plasmas, have led OFES to propose a new computational initiative, the *Fusion Simulation Project* (FSP). The FSP, to be initiated in 2009, will develop experimentally validated computational models capable of predicting the behavior of magnetically confined plasmas in the regimes and geometries relevant for practical fusion energy, by integrating experimental, theoretical, and computational research across the FES program and taking advantage of emerging petascale computing resources.

3.1 General Atomics' Energy Group: DIII-D National Fusion Facility and Theory and Advanced Computing

Background

The DIII-D National Fusion Facility at General Atomics' site in La Jolla, California is the largest magnetic fusion research device in the United States. The research program on DIII-D is planned and conducted by a national (and international) research team. There are more than 500 users of the DIII-D facility from 92 worldwide institutions including 41 Universities, 36 National Laboratories, and 15 commercial companies. The mission of DIII-D National Program is to establish the scientific basis for the optimization of the tokamak approach to fusion energy production. The device's ability to make varied plasma shapes and its plasma measurement system are unsurpassed in the world. It is equipped with powerful and precise plasma heating and current drive systems, particle control systems, and plasma stability control systems. Its digital plasma control system has opened a new world of precise control of plasma properties and facilitates detailed scientific investigations. Its open data system architecture has facilitated national and international participation and remote operation. A significant portion of the DIII-D program is devoted to ITER requirements including providing timely and critical information for decisions on ITER design, developing and evaluating operational scenarios for use in ITER, assessing physics issues that will impact ITER performance, and training new scientists for support of ITER experiments.

General Atomics also conducts research in theory and simulation of fusion plasmas in support of the Office of Fusion Energy Sciences overarching goals of advancing fundamental understanding of plasmas, resolving outstanding scientific issues and establishing reduced-cost paths to more attractive fusion energy systems, and advancing understanding and innovation in high-performance plasmas including burning plasmas. The theory group works in close partnership with the DIII-D experiment in identifying and addressing key physics issues. To achieve this objective, analytic theories and simulations are developed to model physical effects, implement theory-based models in numerical codes to treat realistic geometries, integrate interrelated complex phenomena, and validate theoretical models and simulations against experimental data. Theoretical work encompasses five research areas: (1) MHD and stability, (2) confinement and transport, (3) boundary physics, (4) plasma heating, non-inductive current drive, and (5) innovative/integrating concepts. Members of the theory group are also active in several SciDAC Fusion Simulation Project (FSP) prototype centers and fusion SciDAC projects.

Numerical simulations are conducted on local Linux clusters (46 and 114 nodes) as well as on computers at NERSC and NCCS.

Current Local Area Network Requirements and Science Process

The General Atomics' connection to ESnet is at 1 Gigabits per second (Gbps) with major computing and storage devices connected by a switched 1 Gbps Ethernet LAN. Network connectivity between the major computer building and the DIII-D facility is by dual Gbps circuits. The major data repositories for DIII-D comprise approximately 30 TB of online storage with metadata catalogues stored in a relational database. Network connectivity to offices and conference rooms is at 100 Mbps on a switched Ethernet LAN. There are approximately 2000 devices attached to this LAN with the majority dedicated to the DIII-D experiment.

Like most operating tokamaks, DIII-D is a pulsed device with each pulse of high temperature plasma lasting on the order of 10 seconds. There are typically 30 pulses per day and funding limits operations to approximately 15 weeks per year. For each plasma pulse, up to 10,000 separate measurements versus time are acquired and analyzed representing several GB of data. Throughout the experimental session, hardware/software plasma control adjustments are debated and discussed amongst the experimental team and made as required by the experimental science. The experimental team is typically 20–40 people with many participating from remote locations. Decisions for changes to the next plasma pulse are informed by data analysis conducted within the roughly 15 minute between-pulse interval. This mode of operation requires rapid data analysis that can be assimilated in near–real–time by a geographically dispersed research team.

The highly distributed nature of the DIII-D National Team requires the usage of substantial remote communication and collaboration technology. Five conference rooms are equipped with Polycom H.323 videoconferencing hardware. The DIII-D control room has the ability to use Access Grid, VRVS/EVO, and software-based H.323 for remote videoconferencing. Additionally, scientists utilize a variety of technology to communicate with audio/video to the desktop. The DIII-D morning operations meeting is automatically recorded and published with podcasting capability. A Jabber server is operating for Instant Messaging and is tied into the tokamak's operations for real-time status updates.

Current Wide Area Network Requirements and Science Process

The pulsed nature of the DIII-D experiment combined with its highly distributed scientific team results in WAN traffic that is cyclical in nature. Added onto this cyclical traffic is a constant demand of the collaborative services mostly associated with several different types of videoconferencing with the majority H.323 based. To assist in collaborative activities the DIII-D web site has been transitioned to a Wiki-based system that is available to the DIII-D team worldwide. As the collaborative activities associated with DIII-D continue to increase there becomes an increasing usage of collaborative visualization tools by off site researchers.

The scientific staff associated with DIII-D is very mobile in their working patterns. This mobility manifests itself by traveling to meetings and workshops, by working

actively on other fusion experiments around the world, and by working from home. For those individuals that are off site yet not at a known ESnet site the ability to efficiently transition from a commercial network to ESnet becomes very important. Therefore, ESnet peering points are becoming a critical requirements area.

Local Area Network Requirements – the next 5 years

Although the operation of DIII-D will remain similar for the next five years it is anticipated that the rate of acquiring new data will continue to increase. From 2002 to 2007 the amount of data taken per year increased by eight fold. To keep up with this demand plus the increased usage of collaborative technologies, even within the local campus, discussion have begun to increase the major LAN backbone to 10 Gbps and the connections to selected desktop increased to 1 Gbps.

Wide Area Network Requirements – the next 5 years

As stated previously, the General Atomics' connection to ESnet is at 1 Gigabits per second (Gbps). It is presently in the working plan to increase this connection to 10 Gbps.

For DIII-D, the need for real-time interactions among the experimental team and the requirement for interactive visualization and processing of very large simulation data sets will be particularly challenging. Some important components that will help to make this possible include easy to use and manage user authentication and authorization framework, global directory and naming services, distributed computing services for queuing and monitoring, and network quality of service (QoS) in order to provide guaranteed bandwidth at particular times or with particular characteristics.

Presently, the DIII-D scientific team is actively involved in the start-up phases of the EAST tokamak in China and the KSTAR tokamak in the Republic of Korea. Over the next 5 years, the operation of these tokamaks will become routine and it is anticipated that the remote participation of DIII-D scientists will increase. It is also possible that one of these tokamaks can be operating at the same time as DIII-D, putting an increased strain on the WAN. Therefore, how ESnet peers with particularly China and South Korea will become increasingly important.

Beyond 5 years – future needs and scientific direction

In the outlying years, it is anticipated that the Chinese and South Korean tokamaks will be fully operative with a rich diagnostic set and ITER, located in France, will be close to coming on line. With DIII-D operating to assist ITER it is possible to imagine the DIII-D scientific team working on numerous tokamak simultaneously placing a further strain on the WAN and creating a need for efficient peering to our Asian and European partners.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
Near-term	• DIII-D Tokamak • Collaboration on other experiments • SciDAC/FSP prototype simulation and modeling • Assistance in ITER construction	• Real time data access and analysis for experimental steering • Shared visualization capabilities • Remote collaboration technologies	• 1 Gbps backbone • 100 Mbps to desktop • Remote collaboration with some desktop usage	• 1 Gbps ESnet connection • PKI CA • Strong rapid support of collaborative activities
5 years	• DIII-D Tokamak • Collaboration on other experiments • SciDAC/FSP modeling • ITER construction support and preparation for experiments	• Real time data analysis for experimental steering combined with simulation interaction • Real time visualization interaction among collaborators across US	• 10 Gbps backbone • 1 Gbps to desktops • Significant desktop usage of remote collaboration	• 10 Gbps ESnet connection
5+ years	• DIII-D Tokamak • Collaboration on other experiments • SciDAC/FSP modeling • ITER experiments	• Real time remote operation of the experiment • Comprehensive simulations	• 10 Gbps backbone • 10 Gbps to select desktops • Collaborative visualization to desktop routine	• > 10 Gbps ESnet connection

3.2 LLNL Magnetic Fusion Energy Program

Background

LLNL has active experimental, modeling and theoretical efforts supporting fusion energy science research national and international programs.

Experimental:

> LLNL's main experimental effort is concentrated at the DIII-D National Fusion Facility with efforts in both core and edge physics. Smaller efforts contribute to both the NSTX and Alcator C-Mod research programs. A staff of on-site scientists is maintained at the DIII-D and NSTX facilities. This is augmented by remotely participating scientists at LLNL working on experiments, computations and modeling efforts. LLNL diagnostic instruments at DIII-D include a Motional Stark Effect (MSE) diagnostic for current profile measurements and a variety of visible light and infra-red cameras. We also provide data acquisition systems,

operations support and program direction in edge, advanced tokamak (AT) physics and ITER startup similarity experiments. LLNL has an active role in the ITER diagnostic development in the areas of IR cameras and the MSE system. LLNL experimentalists also participate in international experimental programs at JET in England and Asdex-U in Germany.

Modeling and Design:

LLNL has active modeling efforts in support of experiments in the U.S. and for international collaborations.

The UEDGE and BOUT codes provide modeling for the DIII-D experiments in the areas of pedestal, scrape-off-layer, and divertor physics. Expertise from the LLNL edge physics program (experiments and modeling) provide support for ITER design activities assessing the effects of edge-localized modes (ELMs), tritium retention and divertor performance.

The CORSICA code provides modeling support for the DIII-D AT program. This is currently focused on ITER startup similarity experiments, profile feedback control development and hybrid discharge analysis. CORSICA is actively used in support of the current ITER experiment design review activities where it is used to assess the adequacy of the poloidal and central solenoid field coil systems, vertical stability and shape controller performance, stability properties and viability of operating scenarios. CORSICA has been installed at ASIPP in China for development of scenarios and controllers for the EAST experiment and at Seoul National University for controller development for the KSTAR experiment in Korea.

Theory:

LLNL continues to provide support and development of the BOUT, UEDGE and CORSICA codes used for modeling experiments and ITER design activities. In addition, LLNL has been developing TEMPEST, a 5D (3 spatial and 2 velocity space coordinates) kinetic code, for exploring pedestal and edge physics. LLNL participates in the FACETS code development, the P-TRANSP project and the Edge Simulation Laboratory.

Current Local Area Network Requirements and Science Process

Most current fusion experiments are pulsed devices that generate plasmas for a few seconds to study the basic physics of confined plasmas. Typically there are a few discharges/hour during daily operations. Multiple special-purpose scientific instruments (laser scattering, microwave and RF detectors, imaging and camera systems, probes, etc.) are used to take measurements to diagnose the quality of a discharge. All of this data is digitized and stored for post-shot analysis leading to multiple GB of data per day. Fusion experiments are currently actively feedback-controlled in real time to achieve the highest performance possible. Many channels of the digitized diagnostic data are also used as part of the feedback system and must be fed to the control system in near real time. This is used to control magnets, high-power heating systems and instrumentation systems. All

these systems rely on the availability of robust and high-speed network technology in the laboratory.

Theoretical simulations and modeling activities rely on high-performance compute clusters to study the basic science. Some of these clusters are local to a given laboratory and use high-performance networking for parallelized computations. Compute clusters connected to the experimental devices also provide analysis of data between shots.

Current Wide Area Network Requirements and Science Process

All U.S. magnetic fusion experiments (DIII-D, C-Mod and NSTX are the major experiments in the US) are national collaborative efforts. Staff and data are shared among the laboratories and universities across the US. Remote access to data, diagnostic instruments and control room operations is regularly used. With research staff spread around the country, the access to network-based (e.g. H.323) services for remote participation in meetings and experimental operations is necessary and routinely used. The international flavor of fusion energy research has led to joint experimentation among the US, Japanese, and European programs (now being extended to China and Korea) where researchers participate in experiments via worldwide network connections. International databases are accessible via wide area network connections. The final design/verification effort for the ITER experiments has led to audio-video conferences at all hours of the day.

Consortia of scientists located at multiple sites across the U.S. are developing the generation theoretical of codes needed simulate the underlying physics of confined plasmas. This leads to compute clusters at multiple sites and sharing data and software development via wide-area-network connections.

Local Area Network Requirements – the next 5 years

The next generation experiments coming into operation use superconducting coils that are capable of steady state or long-pulse operation rather than the short pulses on current devices. This will dramatically increase the network bandwidth capacity needed to handle instrumentation and control systems used during operations. In addition, scientific instruments are becoming considerably more sophisticated with imaging, cameras, and three-dimensional techniques under development and coming into use. Detailed measurements will be needed for validation and comparisons with the new theoretical simulations soon to be available.

Wide Area Network Requirements – the next 5 years

The MFE program is relying more and more on the collaboration among sites, both within the U.S. and internationally. Over the next 5 years, considerable progress on the construction of ITER in France is expected and this will lead to a greater emphasis on the experiments and theory in support of its mission. With this focus towards ITER, current experiments in the U.S. and abroad will concentrate more on the requirements for operation of ITER. It is expected that sharing of existing experiments (data and operations) and audio-video meetings will increase.

Theoretical codes now under development will be operational on this time scale. This will drive the network requirements for sharing huge data sets for multi-dimensional simulations of plasma turbulence and the needs for visualization and comparison with experimental measurements.

Beyond 5 years – future needs and scientific direction

In this time frame the ITER experiment will be operational. Operation on a 24-hour basis with groups from around the world is part of its planned operation. Remote participation is integrated into the controls and data access system (CODAC) currently being specified and developed. Efficient operation will require robust and reliable communications: data access, conferencing and shared environments. A priori simulation/modeling of discharges and real-time analysis of data streaming into the control room is required and accessible from a wide area network to all participants. All of this must be done with a strong security infrastructure consistent with the needs for licensing of a nuclear facility (ITER will have a modest tritium inventory).

Concurrent with the ITER operation, participant countries will maintain some level of autonomous experiments for development of ideas and instrumentation needed by ITER. This information will feed into the international program and shared among the various groups.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
• Near-term	• Multiple experiments operating at various sites. • Shared access to data, operations and meetings • Large groups and individual researchers accessing data and information	• Real-time data analysis for controls applications • Rapid, between shot analysis for directing the operations • Shared workspace • A/V communication for meetings and operations	• Few Gbytes/shot at ~4shots/hour	• Remote access to operations and data archives at 500 Mbps • Audio/video meetings and monitoring operations
• 5 years	• Steady-state experiments coming on line in China and Korea • Sophisticated instrumentation for understanding turbulence • Integrated shared workspace	• New parallelized theory codes coming on line • Validation with new experiments and measurements	• 10's Gbytes per shot at ~4shot/ hour rate • streaming bursts of data	• Desktop integration of all service: data, audio/video, shared workspace • 1-2 Gbps bandwidth
• 5+ years	• ITER on-line operations 24 hour access	• Full device modeling • Remote operations	• Real-time streaming of data and real-	• Real-time wide-area access to international operations

	• Shared operations on ITER and U.S. experiments • Access to instruments and CODAC system	developing for ITER • Shared operations on existing devices • Monitoring operations • Off-site data analysis feeding back to operations in near real time	time analysis	• Interactive access to operations and data analysis internationally • Integrated workspace with data and audio/video • 10 Gbps bandwidth

3.3 MIT Plasma Science & Fusion Center and C-Mod Tokamak

Background

The Plasma Science & Fusion Center (PSFC) is a large interdisciplinary research center located on the MIT campus. The largest activity at the center is the Alcator C-Mod Tokamak, one of three major experimental facilities in the U.S. domestic fusion program. Research is carried out in the areas of turbulent transport, plasma-wall interactions, MHD and RF heating and current drive. A significant portion of machine time is devoted to answering questions connected to design and operation of the ITER device, now under construction in Cadarache, France. The C-Mod team is international, with collaborators at more than 35 institutions in the U.S. and abroad. C-Mod is also an important facility for graduate training with about 30 students currently carrying out thesis research at any given time. The PSFC has a number of smaller research facilities as well, including LDX, (Levitated Dipole Experiment) and VTF (Versatile Toroidal Facility), an experiment studying collisionless magnetic reconnection. Research on these devices has relevance outside the fusion program, particularly to space and astrophysical plasma physics. It is worth noting that LDX is a joint experiment between MIT and Columbia University, with each institution contributing personnel and other resources.

The Plasma Science division at the PSFC carries out a broad program of theory and computational plasma physics. The computational work emphasizes wave-plasma interactions and turbulent transport. The division makes extensive use of the DOE facilities at NERSC and NCCS while operating locally, two mid-sized computer clusters with 48 and 256 nodes respectively. PSFC researchers are actively involved in several large national SciDAC and FSP (fusion simulation project) collaborations.

Current Local Area Network Requirements and Science Process

The PSFC has ~1,500 network attached devices, more than half associated with the C-Mod team and experiment. The infrastructure is switched 1 Gbps Ethernet, with Gigabit connectivity to most desktops. The C-Mod experiment is directly supported by ~10 Linux servers and approximately 60 Linux workstations. All experimental data is maintained online, with approximately 12 TB currently archived. Higher-level data is maintained in SQL databases, which hold several million records.

The C-Mod experiment conducts 30-40 “shots” per day, each storing 2-3 Gbyte of data. The team works in a highly interactive mode, with significant data analysis and display carried out between shots. Typically 20-40 researchers are involved in experimental

operations and contribute to decision making between shots. Thus a high degree of interactivity with the data archives and among members of the research team is required.

Current Wide Area Network Requirements and Science Process

The PSFC WAN connection is through a T3 (45 Mbps) ESnet link. In cooperation and coordination with the MIT campus network, this link is currently being upgraded to 1 Gbps. The fiber infrastructure being deployed would allow this link speed to be increased further, with only moderate effort and expense. The WAN link is shared by OSC researchers at MIT, particularly the Lab for Nuclear Sciences (LNS). The older link has been heavily utilized and occasionally congested.

Multi-institutional collaborations are a critical part of the research carried out at the PSFC. In addition to remote researchers who use facilities at the MIT, scientists and students at the PSFC are actively involved in experiments at laboratories around the world. As noted above, the theory groups are involved in several nation-wide computational projects and rely on use of remote super-computers. MIT also supports the MDSplus data system, which is installed at about 30 facilities world-wide.

All groups at the PSFC make active use of collaboration technologies. Three conference rooms are set up for videoconferencing and used for all regular science and planning meetings. In addition, videoconferencing is available from the C-Mod control room and used to support remote participation. In recent years, 5-10% of runs were led by off-site session leaders. Videoconferencing software is also installed on a number of office computers.

The PSFC makes significant use of the ESnet provided collaboration services. The H.323 video conference facilities are used for both scheduled and ad-hoc collaboration. Data/screen sharing are regularly used to broadcast visuals from presentations. Over the next five to ten years we would like to see expansion of these services in both technical and support areas. Room and person based paradigms both need to be provided for, with recognition that the needs for these classes of users differ significantly.

We have begun deploying SIP based VoIP systems (hardware and software) to support the next generation of collaboration tools. Taking advantage of an MIT pilot program, we have been able to integrate these tools into the normal work flow. One aim is to improve ad hoc interpersonal communications, which, we believe, limits the effectiveness and engagement of remote participants.

Local Area Network Requirements – the next 5 years

Overall, operations on our experiments will be similar over the next 5 years. Data rates on the C-Mod experiment have increased by a factor of 10 roughly every 6 years. To provide for this ongoing expansion, we have begun planning for upgrades to the local area network. To improve performance, it will be segmented into several routing domains with a 10 Gbps backbone. We will provide 10 Gbps links to servers and switches. Workstation and desktop connectivity will remain at 1 Gbps for the near future.

Working with the MIT Information Services group, we expect to continue and expand SIP based tools to fully integrate our data and telecommunications networking. Over this period a complete migration from traditional telephony to VoIP is anticipated.

Wide Area Network Requirements – the next 5 years

Collaboration with off-site researchers will likely grow over the next 5 years. For example, PPPL will be switching their focus from Tokamak to Stellarator research, likely increasing demands on the remaining two Tokamak facilities. Planning will also begin for the next large U.S. experiment. It is expected that this would be a national facility run by a broad consortium. Activities in support of ITER construction will be centered on the U.S. ITER Program Office and will probably not drive much additional traffic to MIT.

Extension of SIP based tools to off-site collaborators will be supported by the PSFC in the short run. However, as this technology continues to expand into labs and universities, we will need to find mechanisms to support peering between sites. Eventually, this may become a commercial telecommunications function, but we expect significant gaps in coverage for the near future.

Beyond 5 years – future needs and scientific direction

As we approach ITER operations (about 10 years from now), there may be increased network traffic associated with preparation for the research program, data challenges and diagnostic development.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
• Near-term	• C-Mod Tokamak • Collaboration on other national and international facilities • Simulation and modeling	• Incoming and outgoing remote participation on experiments • Use of remote supercomputers, remote databases • Active use of collaboration technologies	• 10 Gbps backbone • 1 Gbps to each end-user	• 1 Gbps
• 5 years	• C-Mod Tokamak • Collaboration on other national and international facilities • Prep work for ITER • Simulation and modeling	• Increased use of collaboration technologies including SIP/VoIP • Involvement in development of next major U.S. expt.	• 10 Gbps backbone • 10 Gbps to important servers and clients • 1 Gbps to each end-user	• 1 Gbps
• 5+ years	• C-Mod Tokamak • Collaboration on other	• All of above plus preparation for ITER,	• 10+ Gbps backbone • 10+ Gbps to	• 10 Gbps

	national and international facilities including next major U.S. experiment • Research on ITER • Simulation and modeling	research on ITER	important servers and clients • 10 Gbps to each end-user	

3.4 Leadership Class Fusion Computing at ORNL

Background

ORNL is one of the two centers (with ANL) for leadership class computing. Fusion researchers have now been using the National Center for Computational Science (NCCS). Currently there are three funded fusion projects at ORNL (Pat Diamond "Verification and Validation of Petascale Simulation of Turbulent Transport in Fusion Plasmas", Fred Jaeger, "High Power Electromagnetic Wave Heating in the ITER Burning Plasma", and Jeff Candy "Gyrokinetic Steady State Transport Simulations". There are also three extra director's discretion projects for fusion, W. Lee "Validation Studies of Global Microturbulence Simulations vs. New Short-Wavelength Measurements on the National Spherical Torus (NSTX) Experiment", Don Batchelor "Integrated Simulation of Tokamaks", and R. Waltz "Coupled ITG/TEM-ETG."

FES at ORNL includes the following types of workflow and data generation:

- Simulation Monitoring of GTC: need to a move reduced amount of data from ORNL/NERSC to clusters near users desktop machines
- Bulk data movement workflow: sometimes need to move large amounts of data between ORNL and NERSC, and possibly other sites such as MIT or PPPL.
- Data Analysis workflows: Typically for large data, this will be done over the LAN. For the WAN, this will just transmit a subset of data, probably through movies, that will not require much network bandwidth.

The projects that generate the largest amount of data at ORNL are GTC (Gyrokinetic Toroidal Code), CPES, Chimera, and S3D. One of the things all of these codes have in common is that they ALL are switching over to the ADIOS software developed by the end-to-end group at ORNL. This is a componentization of the I/O layer and has allowed researchers to synchronously write over 25 GB/sec on the Cray XT system at ORNL; which is approximately 50% of the peak performance. The biggest problem with this software is that it is now allowing researchers to write unheard of datasets sizes per simulation. For example, GTC and GTS simulations will write out over 150TB per run (600 TB for 1 month of runs). This requires very advanced state of the art networking techniques along with workflow automation techniques, fast access to HPSS, and fast methods to move information over the WAN when required.

Current Local Area Network Requirements and Science Process

Currently ORNL has just finished upgrading to the 250 teraflop Cray XT system with disk access speeds of 40 GB/sec. Early access simulations from GTC will approach 150

TB of data per 1.5 days of simulation. This equates to 1.2 GB/sec (9.6 Gbps) LAN access. The ORNL end-to-end analysis cluster is capable of ingesting data from our Cray at 4 GB/sec, and we have measured over 2 GB/sec for real-time analysis of data. By looking at the 'dashboard' system from CPES/SDM/ORNL and the data movement that must take place for real-time analysis, we envision that we will need about 1 GB/sec for more complex operations that show up on the dashboard system.

Current Wide Area Network Requirements and Science Process

Currently, simulation data produced from large scale simulations are analyzed at scale only at the large supercomputing centers. Typically scientists look at reduced datasets when they bring the data 'back-home'. For the GTC and CPES scientist, this typically means looking at only the scalar field from their simulations, or a five dimensional mesh from the simulations. Currently, the 3D+time data generated from these codes typically approach about 1 TB/day. This leads to the requirement that the WAN needs to move 1TB/day with QoS (~300 Mbps for 8 hours). More complex operations for CPES and GPSC take place back at ORNL, where they have access to much larger systems for analysis. Using the SDM/ORNL/CPES dashboard system, flash video is produced for all variables, and we have seen that scientists want to look at around 9 variables at a single time. Each video can output data at around 20KB/sec, so 9 videos means that they need 180 KB/sec of data. Supporting a large amount (100) of simultaneous users looking at different shots equates to approximately 2 MB/sec of data (with QoS). This will support collaborative access to data.

Local Area Network Requirements – the next 5 years

Coupled simulations from just GTC and XGC could write data at around 200 GB/minute. Looking at the early science run from the GTC team, they will write 60 GB every 60 seconds for over 2K timesteps. When we couple this with XGC, we envision that this data rate will at least double. The local area network must be able to move this data over to the HPSS in a timely fashion. We also need to move the data over to analysis machines. We note that there will be a shared file system at ORNL, but most likely the analysis operations will take place on another machine. This means we need to sustain 2 GB/sec in the LAN. Experimental data can most likely be saved on large disks, along with HPSS. Given that disk storage on the petaflop machine will exceed 10PB, we envision all data on disk. Once ITER becomes a reality, we understand that one shot can encompass 4TB of data, but exascale computing will be upon us by then. Most likely ORNL will have 10 exabytes of storage on disk. Thus, storage on local disk should not be a problem for large supercomputing centers, even for ITER sized experimental data. However, we should make sure that access to permanent (HPSS or HPSS-like storage) should be fast. This means that we would want to make sure that it is easy to get one shot of data (4TB) in one hour or less, which requires 1 GB/sec bandwidth. For simulation data, we believe that this number will require us to obtain 300 TB of data in at least 12 hours (1/2 day), so that scientist can analyze their data on the same day they want to look at it (7 GB/sec).

Wide Area Network Requirements – the next 5 years

In five years ORNL will move from the 250 teraflop computer (2008) to the petaflop computer (2009) to a sustained petaflop computer. The disk rates will move from 40 GB/sec to 150 GB/sec to 1 TB/sec. The increase of data from the simulations will also increase, but in a much more linear fashion over the next 5 years. We believe this will be the case because codes will be coupled, and the amount of data that will need to be processed over the WAN will most likely only quadruple in the next five years. If we still use web 2.0 technologies in the dashboard, then the WAN requirement will not increase by much from this area. The WAN requirement can increase for the reduced datasets that users want to interact with. We envision looking at reduced sets of the data (reduced by 100x). A full size dataset in 5 years will approach 5PB, which means we will want to move 50 TB of data during the lifetime of the run (5 days); i.e. we will need 10TB/day = 120 MB/sec (960 Mbps) sustained per simulation. If multiple people are looking at data (5 people) then the requirement will be 600 MB/sec (4.8 Gbps).

Beyond 5 years – future needs and scientific direction

If trends continue and new codes get coupled with experimental data, then we will see a further explosion of data beyond five years. We believe ITER will generate 40TB of data in 1 hour, and simulations will generate 1PB of data in a day (42 TB/hour); then workflow automation becomes even more crucial for scientific understanding of the data. The human visual system will not be able to compete with these large data generation rates, so advanced data analysis routines must take place. Indexing of data must also take place during the data generation, since computers can only take advantage of the added bandwidth if the algorithms can achieve a high level of parallelism. Analysis routines will need to be parallelized, and fast access of information must be achieved for scientific insight. Estimates for WAN will most likely be a 10x increase of data over the previous numbers.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
Near-term	• ORNL petaflop computer • GTC, XGC with delta-f and full-f	• PIC simulations. End to end workflows for analysis and verification and validation.	• 16 Gbps	• 16 Mbps for dashboard environments with QOS. • Up to 800 Mbps for moving most of the data from the largest simulations to another system.
5 years	• Coupling of GTC with XGC	•	• 56 Gbps	• 8 Gbps
5+ years	• ITER	•	• 96 Gbps	• 80 Gbps

3.5 Princeton Plasma Physics Laboratory (PPPL), Princeton, NJ

Background

The mission of PPPL is to pursue a scientific program of theory and experiment to develop an understanding of magnetically confined plasmas, and create innovations to make fusion power a practical reality.

Current Local Area Network Requirements and Science Process

The PPPL network consists of a mix of 10 and 100 Mbps desktop connections, linked to a core switch providing 1 Gbps connections to servers and a set of high performance clusters. Current connectivity provides for terminal access to compute servers and midrange visualization capabilities for data. Applications use of LAN services such as NFS occasionally saturate portions of the internal network, causing intermittent performance and reliability issues.

Current Wide Area Network Requirements and Science Process

PPPL's current WAN connection to ESnet is an OC3 (155 Mbps) connection. With basic data copying techniques (scp, single-thread ftp), users achieve a data download capability of about 1 MB/sec. Using more advanced tools such as multi-stream ftp users can achieve ~10 MB/sec, with saturation of the OC3 degrading connectivity for other uses of the wide area network. Other documented performance problems with access to PPPL networks over the OC3 have also occurred. The current limited connectivity precludes extensive downloads of remote experimental data, effectively allowing VNC web streaming of limited visualization data sets and/or terminal sessions to remote sites (with very limited graphics capability) as the only means of live collaboration. Remote operations of experiments have been done within the U.S. and overseas (to JET, Europe), based on a very select subset sampling of downloaded data and graphics.

Local Area Network Requirements – the next 5 years

Under current plans, all desktops will be connected at 100Mb/s with connectivity to core internal services and switches upgraded to 2 Gbps. Topological changes to LANs will address performance bottlenecks as they are encountered. Greater use of LANs for control of PPPL experiments is contemplated; this will likely pose issues of quality of service, latency, as well as additional security concerns—available bandwidth is expected to be sufficient.

Wide Area Network Requirements – the next 5 years

In order to reach long term goals for PPPL remote operation of ITER tokamak experiments, effective data connectivity to current generation overseas experiments will need to be achieved as an intermediate step. The KSTAR tokamak (South Korea) has commenced operations and will reach its full parameters within the next 7 years. Collaboration on KSTAR represents a near term opportunity to prepare to meet the long term ITER requirements.

Data rates on KSTAR are planned to be as follows:

- Medium length pulse regime: 20s shots, 30 shots per day, 3.5 TBytes/day
- Long pulse regime: 300s shots, 10 shots per day, 17.6 TBytes/day

Current PPPL connectivity (OC3) operating at 10 MBytes/sec would require 4 days to download the complete experimental data archive for a single days' worth of medium length pulse data, and 20 days to download the archive for a single days' worth of long pulse data from KSTAR.

Data management strategies can greatly mitigate the load, if appropriate software engineering investments are made. It is not clear that the entire dataset needs to be downloaded. Even so, upgrade of the PPPL OC3 wide area network connectivity remains a clear urgent requirement.

Beyond 5 years – future needs and scientific direction

ITER data rates (in ~10 years) are projected as follows:

- Long pulse regime: 400s shots, 24 shots per day, 40 TBytes per shot.
- 960 TBytes/day – approaching 1 PByte.

Connectivity of 100 Gbytes/sec – about 5000 times the current saturated OC3 capability – would be required to stream ITER data to a PPPL ITER control room in near-real-time. Connectivity of about 11 Gbytes/sec – over 1000 times the current optimized OC3 capability for a single use – would be required to copy the entire single day ITER data archive over a 24 hour period.

However, live participation in experiments with prompt data analysis capability, including timely results of analysis feeding into the decision making process for subsequent shots, requires a near-real-time capability.

Real time transfer of the entire experimental archive to all collaborating sites is not likely to be feasible. Data management strategies can greatly mitigate this anticipated load, if appropriate software engineering investments are made (and they will have to be).

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
Near-term	• NSTX • DIII-D • C-Mod	• U.S. Fusion experiments • ~100 Gbytes/shot • ~40 shots/day	• 1 Gbps	• 1 Gbps
5 years	• NCSX (stellarator) • KSTAR (tokamak in South Korea)	• U.S. and overseas fusion experiments	• 2 Gbps	• 10 Gbps
5+ years	• ITER (Europe)	• Overseas fusion experiment	• 10 Gbps	• 100 Gbps

3.6 PPPL Computational Science Networking Requirements

Background

PPPL is involved in all of the Fusion Energy Science (FES) SciDAC projects. These include, for example, CEMM (Center for Extended MHD Modeling), GPSC (Gyrokinetic Particle Simulation Center), APDEC (Applied Partial Differential Equations Center), and CSWPI (Center for Simulation of Wave-Plasma Interactions). A common element among all these projects is the use of scientific codes to investigate physical phenomena relevant to OFES on parallel platforms ranging from Linux clusters, to capacity computing at NERSC, and in some instances on leadership class facilities at ORNL. For example, CEMM codes M3D and NIMROD investigate nonlinear MHD for macroscopic physics at device scales; the GPSC code GTC (Gyrokinetic Toroidal Code) is used to investigate the physics of plasma microturbulence in a tokamak, and the APDEC code AMRMHD is used to investigate the MHD phenomena associated with pellet refueling of a tokamak. In all instances, it is fair to state that these codes are state-of-the-art computational engines, which operate at the leading edge of supercomputing science. The purpose of this case study is to determine the networking requirements today and in the future to enable the computational scientists to perform their tasks efficiently and effectively.

Compared with the other fusion codes at PPPL, GTC and the other 5-D gyrokinetic codes generate the most massive amounts of data, and execute in a nearly-perfectly scalable fashion as demonstrated by weak scaling studies up to thousands of processors on a wide variety of hardware platforms (IBM SP, Cray XT4, Blue Gene etc.). The data throughput of GTC may be assumed as an upper bound on the data generated by fusion codes at PPPL and thereby implicitly implying that the networking requirements that satisfy the needs of GTC, will meet the needs of other fusion codes. The ensuing discussion will, therefore, focus mostly on the needs of GTC although we will point out requirements by M3D for the sake of comparison.

Current Local Area Network Requirements and Science Process

A *typical* GTC simulation today employs a billion particles, and a mesh of approximately 50 million points, and is run for 10,000 time steps. Storage requirements for each time step is: 10^9 x 8 x 12 variables (for particles) + $5x10^7$ (mesh size) x 8 x 4 (variables) = 97.6 GB (Gigabytes). If every time step is stored, we quickly approach one PB (Petabytes) of storage requirement. In practice not every time step is saved, and each GTC run results in 10TB (Terabytes) of data. There are no local area network issues, which hamper this process. There is ongoing work in collaboration with ORNL (see section 3.4) in which GTC will be switching over to the ADIOS software to enable it to synchronously output over 25 GB/sec on the Cray XT at ORNL. The ensuing data sets will exceed 150 TB/simulation. As section 3.4 points out, this requires advanced local networking (up to 10Gbps LAN access speeds) and fast access to HPSS.

Current Wide Area Network Requirements and Science Process

Given the size of the data from a typical GTC simulation, and the fact that data analysis is performed in parallel on the remote supercomputing site, there is little reason to transfer the data to local disk. In the event the data transfer is undertaken, usually a smaller subset

of data or the results of data analysis (~100 GB) are transferred using the tool *bbcp*. Transfer rates of approximately 40 MB/sec have been empirically observed, and the entire data transfer takes less than one hour. PPPL's WAN connection to ESNet is an OC3 (155 Mbps) connection. So, insofar as the present modus operandi is concerned, this poses no undue burden on the scientist. However, there is a strong thrust towards simulation monitoring to ensure valuable supercomputing resources are not wasted. Furthermore, quick analysis of the simulation data will become important in the context of widely dispersed collaborations and for the information sharing amongst codes in the proposed Fusion Simulation Project (FSP). Real-time simulation monitoring will require either remote visualization (and associated latency issues of X-windows) or quick conversion to images/flash videos and rendering these with tools such as the ORNL dashboard. It is estimated that the latter will require 2 MB/sec for QOS. Another pertinent issue, which arises in the context of the FSP, is quick transfers over the WAN from one supercomputer site to another are anticipated. A typical GTC restart checkpoint files take about 45 minutes to move from ORNL to NERSC (about 40 Mbps). Transfer rates at 1 GB/sec will reduce this time to a more acceptable 100 seconds.

As mentioned earlier, GTC poses an upper bound on networking requirement. Nonetheless, we now highlight the requirements for M3D – an MHD code, which employs a finite element discretization in each poloidal plane and finite-difference or Fourier decomposition in the toroidal direction. A *typical* production M3D run on Franklin (CRAY XT4 at NERSC) writes 373 MB per output file containing 12 time slices (redundant information such as mesh connectivity is written once so that one time slice is not 373/12 MB but rather 50 MB). The output frequency is typically every 45 minutes. Over the duration of the simulation 400 such files are written and saved to HPSS. The total simulation run time is 45K CPU hours on 72 nodes. The data is currently transferred using scp (which will be changed to bbcp) from NERSC to PPPL at a rate of 1 Mbps. Due to the relatively small amount of data; there are no significant networking bottlenecks. Again, as mentioned for GTC, improvements in networking speeds for GTC will automatically result in improvements for M3D and allow remote simulation monitoring.

Local Area Network Requirements – the next 5 years

In five years, the number of particles and mesh used by GTC will increase by at least an order of magnitude. It is anticipated that the GTC code will write out 1TB for every time slice saved. An important LAN requirement will be to move this data efficiently from scratch space to HPSS. Over the next five years M3D will double its resolution along the radial, poloidal and toroidal directions with a consequent increase in data per time step by a factor of eight. Furthermore, the number of time steps per simulation will also double. LAN requirements are expected to scale linearly with the size of the simulation.

Wide Area Network Requirements – the next 5 years

Simulation data will easily increase by at least an order of magnitude (factors of 10-16 expected) over the next five years which means that to achieve the same wall clock time in transferring the data, the network speed will have to improve by an order of magnitude over the next five years. If the FSP becomes a reality, this will imply a massive

collaborative effort between computational physicists dispersed across the United States. A key component of the FSP will be state-of-the-art research massively parallel codes, which feed into a production component (e.g. gyrokinetic codes providing first principle quantities into analysis codes such as TRANSP). Also, coupled simulations (e.g. coupled gyrokinetic – MHD simulations as undertaken by the CPES proto-FSP project) will be commonplace requiring fast transfer over LANs and WANs.

Beyond 5 years – future needs and scientific direction

While it is hard to predict reliably how the science process will evolve beyond five years, one may safely conjecture that the data explosion will continue perhaps nonlinearly and simulation codes such as GTC will routinely produce 1PB/simulation. A paradigm shift may be required in the manner in which data is scientifically queried. From a networking viewpoint, the network speeds may not have to increase at rates comparable to the growth rates of the data because intelligent reduction of data, feature extraction, compression etc. will be used to reduce the data to a state suitable for human comprehension.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
• Near-term	• GTC on NERSC and ORNL Supercomputers (Cray XT4). • M3D on NERSC (Cray XT4).	• PIC simulations. Post-processing on remote supercomputer sites. • Simulation Monitoring. • MHD code. Post processing on remote supercomputer sites.	• 10 Gbps	• 16 Mbps for remote monitoring. • 8 Gbps for moving checkpoint files from one site to another
• 5 years	• GTC on NERSC and ORNL Supercomputers	• copy checkpoint files from ORNL/NERSC	• 96 Gbps	• 80 Gbps
• 5+ years	• GTC, M3D, Other FSP related codes at supercomputing sites across the U.S.	• Coupled simulations (e.g. kinetic-MHD coupled codes)	• 160 Gbps	• 160 Gbps

3.7 Fusion Research at Tech-X Corporation

Background

Tech-X participates in 4 fusion SciDAC projects:

(1) Framework Application for Core-Edge Transport Simulations (FACETS) - full-scale reactor modeling for the U.S. fusion program and ITER, with the emphasis of integrating core, edge and wall phenomena (Lead Principal Investigator: John Cary).

(2) Center of Extended Magnetohydrodynamic Modeling – modeling MHD phenomena effecting stability of existing and future fusion devices (Tech-X PI Scott Kruger).

(3) Center for Simulation of Wave Plasma Interactions– modeling electromagnetic wave processes in fusion plasmas (Tech-X PI David Smithe).

(4) Simulation of Wave Interactions with Magnetohydrodynamics (SWIM) – coupling RF and MHD modeling (Tech-X PI Scott Kruger).

Tech-X also has four active SBIR Phase 2 projects and two active SBIR Phase 1 projects. Total approximate OFES funding of Tech-X is approximately $2M/year.

Current Local Area Network Requirements and Science Process

The science process consists of developing new computational software capability through development, executing the resulting software applications to obtain data, reducing the data to manageable size, and then performing statistical analysis on that data or visualizing it. Given that Tech-X designs its software to run on (Unix-based, i.e., Linux, OS X, AIX) machines from desktops to supercomputers, the development process can nearly all occur on Unix-based desktops. Of course, development can require use of the larger clusters at Tech-X in order to do performance testing or to debug issues that arise with parallelism. But this aspect of the process has minimal networking requirements, as data transfers are minimal, especially as most scientists work with tools, like ASCII text editors and debuggers that require minimal data transfer. The same is true for data reduction and statistical analysis, which can easily be done through modest network connectivity.

The need for connectivity arises with the use of graphical interface tools, such as the TotalView debugger, and visualization, where the latter can involve either transferring data to local workstations for visual analysis or use of X-Window based tools for remote visualization or the use of client-server tools, such as VisIT. For the Tech-X 100 Mbps local area network, all of these uses are reasonable except possibly for data transfer and X-based transfer. With local high-end simulations generating tens to 100 data files, each of 16 GB, it requires 20 minutes to transfer a file locally. With researchers willing to tolerate about 1 minute of transfer time (still an hour for all the data of an entire simulation) this limits local simulations to sub GB data files, which correspond to 30 Mcell simulations.

These considerations lead to local networking needs of the order of Gbps. Thus, to meet these needs, Tech-X would need to upgrade its local area network by a factor of 10, which it could do by going to 1000BT Ethernet.

Another way of meeting the need is to further develop client-server visualization tools, like VisIT. This is a great tool with a good architecture, but resources are needed to add the data reading “plugins” to Visit to allow it to read all of the formats that used by scientists at Tech-X. More cost effective may be providing software layers that ensure that data is written out in a way that a standard plugin can read it in, and then developing that plugin. This is the VizSchema approach currently being pursued by Tech-X.

Current Wide Area Network Requirements and Science Process

The above considerations all carry over to the wide area network, but that status is worse. At Tech-X, 3 T1 lines, providing 4.5 Mbps aggregated transfer rate, provide Internet access. It is through this that Tech-X scientists access several high-performance computers including Franklin, Jacquard, and Bassi at the National Energy Research Scientific Computing Center (NERSC); and Jaguar at the National Center for Computational Sciences (NCCS). Jacquard, for example, is a 64-bit Opteron cluster with 356 dual CPU, 6 GB memory, nodes. These supercomputers have performance capabilities 10-100x the clusters at Tech-X. Thus, the situation is made worse at both ends. The "last-mile" connectivity at Tech-X is a factor of 20 smaller than the local, and the amount of data being generated is a factor of 10-100 larger.

Consequently, scientists at Tech-X have had to adopt a different work methodology. Data generated at remote resources is post-processed and transformed to images and movies remotely. These images and movies are then brought up locally or displayed on a web site. Some sub-selected data of interest is brought up locally for local visualization and analysis. There is simply no way that full datasets can be transferred back to our home facility, as the last-mile problem causes the transfer of even one data file to require 8 hours for transfer.

The typical amount of data generated at remote sites comes from usage of VORPAL (10-20 scientists running VORPAL producing 50 GB per time step during approximately 1 hour and having 40-100 steps on 500 processor.

The next level in connectivity for improved workflow would be to get to the level of connectivity that allows client-server visualization. We estimate that this is possible at around 100 Mbps. For data transfer of modest sized simulations we need to have about 1 Gbps connectivity.

Local Area Network Requirements – the next 5 years

In the next year, Tech-X will be increasing its local computing capability by a factor of 5. It is reasonable to expect a factor of 10 over the next 5 years. The amount of the generated data will increase dramatically once we include turbulence simulations into FACETS. Thus, local maximal requirements for timely data set transfers leads to the need for a local 10 Gbps Ethernet network.

Wide Area Network Requirements – the next 5 years

The factors needed scale with the size of the coming large-scale facilities. We already compute on 20000 cores, so with 200,000-core machines coming on line, one can imagine that for client-server capability, keeping the data remote, we need of the order of 1-10 Gbps connectivity to the major computing facilities.

Beyond 5 years – future needs and scientific direction

We have not given sufficient thought to this, but it will remain critical to keep up with software development issues. Robust, client-server visualization and analysis applications with parallel back-ends will help minimize networking needs.

Primary Collaborators of Tech-X

We collaborate mostly with LLNL, ANL, PPPL, ORNL, LBNL, NERSC, and ITER in France.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
• Near-term	• Computational projects modeling various phenomena in fusion plasmas	• Designing first principle modeling codes and incorporating some legacy codes. • Analysis and visualization of separate and partially integrated phenomena.	• 100 Mbps local • 1 Gbps desired	• 3 T1 lines with access to NERSC. • 100 Mbps connectivity desired.
• 5 years	•	• Full fusion device modeling framework accommodating parallel and legacy fusion codes. • Remote and local parallel data analysis and visualization for V&V. • Sensitivity studies of integrated scenarios.	• 10 Gbps	• ESnet access at 1 Gbps would be desirable.
• 5+ years	•	•	• >10 Gbps	• > 1 Gbps

3.8 U.S. ITER Construction

Background:

U.S. Contributions to ITER is a DOE Office of Science Major Item of Equipment (MIE) construction project consisting of procurement of hardware (including supporting R&D and design); assignment of personnel (U.S. engineers and scientists) to the ITER site in Cadarache, France and in Field Teams in the ITER parties; and cash contributions to the ITER Organization for the U.S. share of common expenses such as personnel infrastructure, assembly, and installation. All U.S. ITER project activities will be managed by the U.S. ITER Project Office.

Currently, ITER is a construction project and will be until approximately 2018. During this time, the main need for U.S. ITER network capability will be for construction related activities. The activities are primarily CAD related, collaboration related, and CODAC (controls, data acquisition, and communications) development related. These needs will be provided by the DOE network infrastructure and other ITER domestic agency network infrastructures outside of the United States.

Current Local Area Network Requirements and Science Process

ITER is currently in a construction phase and is scheduled to be until 2018. The local area network requirements at the U.S. ITER Project Office (USIPO) consist of the ITER Collaborative Network (ICN), general office needs, and videoconferencing. Internally at ORNL, the USIPO uses gigabit Ethernet and connects to the rest of ORNL with gigabit Ethernet as well. This speed supports our needs for three dimensional computer aided design modeling tools (Dassault Catia & Enovia), multipoint video conferencing, and all of the general USIPO office needs.

The ITER Collaborative Network (ICN) is a private network between ITER IO and ITER domestic agencies (DAs) that wish to participate in a highly integrated manner. The ICN is used as a mechanism to create a virtual ITER network at each site. The first use of the ICN is for CAD data caching and replication. The ITER Collaborative Network (ICN) currently supports the CAD data caching to help increase data access speeds for the CATIA models. The next step is to develop the CAD data replication capability. Future uses of the ICN will help support development such as CODAC development and testing for remote collaboration.

Current Wide Area Network Requirements and Science Process

The main wide area network requirements at this time are for the ITER Collaborative Network (ICN) and videoconferencing. The ITER agreed ICN minimum network connection speed is 20 Mbps using standard TCP/IP protocols within each DAs subnet. We have already achieved over 30 Mbps to Cadarache France and our goal is >100 Mbps. This has been accomplished through the use of a SkyX protocol accelerator device.

The videoconferencing needs for ITER include the ability to have multiple multipoint videoconferences with at least 6 other sites at 384 Kbps or above. The locations for these multipoint connections would normally be in North America, Europe, or Asia.

Local Area Network Requirements – the next 5 years

Over the next 5 years the gigabit network connections for the USIPO and partner sites will be sufficient to support the USIPO needs.

Since ITER is an international collaboration, the biggest needs or possible roadblocks ahead could be security. We need the ability to connect and collaborate effectively with the ITER domestic agencies while maintaining good security.

Wide Area Network Requirements – the next 5 years

Over the next 5 years the use of the ICN will grow to include not only the CAD designers located at ORNL but also people at PPPL, Sandia, LANL, Savannah River, and vendors. The current plan is to allow these sites to connect to the USIPO ICN via VPN. This will require ~5 VPN tunnels and ~20 desktop tunnel connections to ORNL.

Within the next 5 years CODAC development will begin to see progress and therefore the testing of remote operation capability will commence. The ICN will be the protected

network to begin the development of the CODAC components on and the CODAC remote operation capability.

Over the next five years it is expected the amount and speed of videoconferencing will escalate. Some factors that will cause this are personal videoconferencing, high definition videoconferencing, and a larger number of participants from laboratories, vendors, and other domestic agencies.

Beyond 5 years – future needs and scientific direction

Until the end of the ITER construction phase, the ICN, VC, and general collaboration needs of the U.S. ITER Project Office will continue and capacities needed above. Security will continue to be an issue.

As ITER becomes an operating facility there are expectations that ITER could provide throughput requirements of 10 Gbps after 2018. Since ITER CODAC and diagnostics systems are only in conceptual design at this point, it is not possible to accurately predict the network requirements this far in advance.

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
• Near-term	• ICN • VC	•	• 155 Mbps • 3 Mbps	• 155 Mbps • 3 Mbps • Total ~160 Mbps
• 5 years	• ICN • VC • CODAC and Remote Operation Development	•	• 155 Mbps • 10 Mbps • 30 Mbps	• 155 Mbps • 10 Mbps • 30 Mbps • Total ~200 Mbps
• 5+ years	• ICN • VC • CODAC and Remote Operation Development	•	• 155 Mbps • 15 Mbps • 50 Mbps	• 155 Mbps • 20 Mbps • 60 Mbps • Total ~250 Mbps

3.9 GTC User Group at University of California, Irvine

Background

UCI is involved in SciDAC GPS project. Major research topics are: gyrokinetic simulations of momentum transport (electrostatic with adiabatic electrons); trapped

electron modes (electrostatic with kinetic electrons); energetic particle physics (electromagnetic, multispecies).

Current Local Area Network Requirements and Science Process

Currently there is 100 MB/sec LAN available at UCI. For large simulation data sets the transfer rate of 1 Gbps is anticipated.

Current Wide Area Network Requirements and Science Process

At present time most of the data processing and analysis is performed on local cluster at UCI. Typically, it is about 10-100 GB/day data transfer from ORNL or NERSC. Part of the data analysis, which requires more computational resources and larger data sets is performed remotely, without being transferred to UCI.

Local Area Network Requirements – the next 5 years

The UCI LAN is expected to reach bandwidth of 1 Gbps in a couple of years.

Wide Area Network Requirements – the next 5 years

3D+time fluid data transfer between ORNL/NERSC and UCI (up to ~ 1TB/day);

5D particle data transfer between ORNL/NERSC and UC Davis/LLNL (~10TB data sets).

Core-edge coupling simulations (~1-10 Gbps)

Beyond 5 years – future needs and scientific direction

Experimental data transfer from ITER and other large machines would require significant increase of WAN bandwidth

Summary Table

Time Frame	Science Instruments and Facilities	Process of Science	Anticipated Requirements	
			Local Area Network Bandwidth and Services	Wide Area Network Bandwidth and Services
Near-term	• ORNL (Jaguar), NERSC (Franklin) for running GTC	• Delta-f PIC simulations of momentum transport, CTEM and EP physics	• 1 Gbps	• 1 Gbps for transferring simulation data
5 years	•	• Full-f, core-edge coupling	• 10 Gbps	• 10 Gbps
5+ years	• ITER	• Experimental Data	•	•

3.10 University of Maryland Fusion Projects

Background

There is approximately $2M/year of effort funded by OFES at the University of Maryland (UM) in the departments of Physics, Engineering, Mathematics, Computer Science, and the Center for Multiscale Plasma Dynamics. The dominant network usage is associated with running simulations at national supercomputer centers at NERSC and ORNL. UM is an Internet2 site, and most of the scientists in the plasma theory program have 1 Gbps connections all the way to the desktop. UM scientists not heavy network users, due to our policy to mainly leave the data where it is computed, and to focus on moving highly reduced quantities of direct scientific value around the country. We do value a high-availability, high-reliability network, so that we can do occasional development remotely (interactively).

Current Local Area Network Requirements and Science Process

We are well situated, but primarily because we have defined our science process to fit within our network and computing resources. Locally, we have fast connections and a unified campus-wide file system. The Computer Science Dept at Maryland is strong, and we share a building with them, which means that we get upgrades earlier than even the rest of campus.

Current Wide Area Network Requirements and Science Process

Again, we have defined our science process to fit the available resources. This drives our "compute and reduce there model". We rarely move large datasets around.

We also support remote users on computers at Maryland. We attempted to use Globus and later the FusionGrid for authentication and so on, but have temporarily abandoned that development effort because our funding is coming from other directions. We are active TeraGrid members as well.

Local Area Network Requirements – the next 5 years

We are moving to a 10 Gbps LAN. We are considering whether to build a small supercomputer, which would drive modest local network needs. Compared to experimental data acquisition and processing needs (which are in groups other than ours), though, we do not see ourselves as drivers of the local LAN upgrade schedule.

Wide Area Network Requirements – the next 5 years

We have a strong need for good bandwidth to the national labs, to support occasional interactive development (say, with Totalview). We have no plans to explore 'streaming data' from simulations, or moving restart files around. Like other simulation groups, we are using the large computers at ORNL and NERSC to do high-end simulations (with more than 10,000 processors per job). However, we do not rely on the network for heavy data flows with any regularity.

Beyond 5 years – future needs and scientific direction

We hope to be a part of whatever infrastructure develops around the ITER project, and to supply codes, analysis, and CPU cycles to simulation projects in the fusion program.

4 Major International Collaborations

Background

International collaboration has been a key feature of magnetic fusion energy research since declassification in 1958. Over the last 25 years, formal multi-lateral and bilateral agreements have created, in effect, a single, loosely coordinated research enterprise. The fusion community, which traditionally included the US, Western Europe (+Australia), Russia (USSR) and Japan, has expanded in recent years to include eastern Europe, Korea, China, and India. Planning and program advisory committees typically have cross membership, particularly among the most active nations (US, EU, JP). Preparation for ITER has further strengthened cooperative research, especially through the ITPA (International Tokamak Physics Activity). Driven by improvements and broad deployment of network technology, the changes in modalities for collaborative research have been dramatic with remote access to data and remote participation in planning and execution of experiments now routine. However, despite technological advances, challenges remain; high-performance wide-area networking has not reached the endpoints for all of our collaborators. Moreover, collaborations which cross major administrative domains must cope with different choices for standards as well as different policies for privacy, data access, remote participation and remote control.

Ongoing and planned international collaborations between major facilities are described below:

AUG: ASDEX-Upgrade is a mid-sized divertor tokamak located at the Max-Planck-Institut für Plasmaphysik (IPP) in Garching, Germany. The primary mission for the machine has been support for ITER design and operation, with focus on integrated, high-performance scenarios, the plasma boundary and first wall issues. There are major collaborations in place with U.S. facilities, including on C-Mod (H-mode pedestal physics, ICRF heating, metallic first walls and steady-state scenario development); DIII-D (divertor and pedestal physics; ECRF heating and current drive and steady-state scenario development); NSTX (diagnostics development and turbulence studies). Important collaborations on theory and modeling are also in place with many U.S. groups.

JET: The Joint European Torus, which is under the European Fusion Development Agreement (EFDA), is located at the Culham Science Centre, in Abingdon, United Kingdom. It is the largest tokamak currently in operation in the world. Major collaborations in place with U.S. facilities; include C-Mod (H-mode pedestal physics, SOL transport, self-generated core rotation, TAE physics and disruption mitigation); DIII-D (H-mode pedestal physics, especially ELM suppression, neoclassical tearing modes, resistive wall modes and rotation, steady state scenario development); NSTX

(Alfven eigenmodes physics, neoclassical tearing modes and resistive wall mode research).

ITER: ITER is a collaboration among seven parties (Europe, Japan, USA, China, South Korea, Russia and India) to build the world's first reactor scale fusion device located in Cadarache, France. The ITER Project expects to finish major construction in 2018 and to operate for 20 years. The project is presently scheduled to begin DT fusion experiments in 2022. Collaboration during the construction phase discussed in another chapter of this report; the research phase discussed below under "beyond 5 years"

KSTAR: KSTAR is an all-superconducting tokamak experiment located at Daejeon, Korea. It will operate with hydrogen and deuterium and expects first plasma is anticipated to occur by the middle of 2008. KSTAR size, operation capabilities and mission objectives for the initial operating period will be comparable to the present DIII-D tokamak. The main research objective of KSTAR is to demonstrate steady-state high-performance advanced tokamak scenarios. PPPL (PCS, diagnostics, ICRF), ORNL (fueling), DIII-D (PCS, data analysis, ECH), MIT (Long-pulse data system), Columbia U. (data analysis).

JT60-U: The U.S. collaboration with the Japan Atomic Energy Agency (JAEA), formerly known as the Japan Energy Research Institute (JAERI), has a rich history extending over the previous 30 years. The present machine, JT60-U, is operated by the JAEA at the Naka Fusion Institute. Presently, this machine is scheduled to cease operation towards the end of 2008. Collaborations on JT60-U will then move to data analysis exclusively.

EAST: EAST (Experimental Advanced Superconducting Tokamak), located at the Chinese Academy of Sciences, Institute of Plasma Physics (ASIPP), Hefei, China, is the world's first operating tokamak with all superconducting coils. The collaboration with scientists from the United States was instrumental in their successful first plasma September 2006. EAST is somewhat smaller than DIII-D but with a higher magnetic field so the plasma performance of both devices should be similar. Its mission is to investigate the physics and technology in support of ITER and steady-state advanced tokamak concepts. Major collaboration with U.S. facilities include, DIII-D (digital plasma control, diagnostics, advanced tokamak physics, operations support), PPPL (diagnostics, PCS), Columbia University (data analysis), MIT (long-pulse data system development) and the Fusion Research Center at the University of Texas (diagnostics, data analysis, theory).

SST-1: SST-1 (Steady State Tokamak) is located at the Institute for Plasma Research (IPR), in Bhat, India. It is the smallest of all the new superconducting tokamaks with a plasma major radius of 1.1 m, minor radius of 0.2 m, and plasma current of 220-330 kA. First plasma is anticipated to occur in 2009. The main object of SST-1 is to study the steady state operation of advanced physics plasmas. At this time facilities collaboration are with DIII-D in the area of physics, plasma operation, theory, and ECE diagnostics. It is anticipated that this collaboration will grow to encompass other groups within the United States.

LHD: LHD is a large (R = 3.9 m, a = 0.6 m, B = 3 T) superconducting stellarator device that began operating in 1998 at the National Institute of Fusion Science, Toki, Japan. There are active U.S. collaborations on this device.

Current Wide Area Network Requirements and Science Process

The wide area network obviously plays a critical role in the ability of U.S. scientists to participate remotely in experimental operations on any of the international machines discussed above. Network use includes data transfer as well as specialized services like a credential repository for secure authentication. Overall, the experimental operation of these international devices is very similar to those in the U.S. with scientists involved in planning, conducting and analyzing experiments as part of an international team.

Experimental planning typically involves data access, visualization, data analysis, and interactive discussions amongst the distributed scientific team. For such discussions, Access Grid, VRVS/EVO, H.323 videoconferencing, and Skype have all been utilized. It should be noted that for some foreign collaborations (e.g. EAST), the ability to use traditional phone lines for conversations are not an option due to the prohibitive expense. Which technology that is used often depends on the technical capability of scientists at each end and on their experience. Recently, there has been a trend to use H.323 for more formal larger meetings and these are facilitated by Multipoint Control Units (MCU) to bridge together numerous participants. Data analysis and visualization is typically done in one of two ways; either the scientist logs onto a remote machine and utilizes the foreign laboratories existing tools or they use their own machine and tools and remotely retrieve the data. The widespread use of MDSplus makes the latter technique easier and more time efficient yet this is not possible at all locations due to the fact that not all sites have adopted MDSplus as an interface standard to provide remote access.

Remote participation in international experiments has the same time critical component as does participating in experiments on U.S. machines. The techniques mentioned above are all used simultaneously to support an operating tokamak placing even higher demands on the wide area network, especially predictable latency. In addition to what was discussed above, information related to machine and experimental status needs to be available to the remote participant. The use of browser-based clients allows for easier monitoring of the entire experimental cycle. Sharing of standard control room visualizations is also being facilitated to assist the remote scientist to be better informed.

Despite improvements in intercontinental links and development of national networks, collaborators still report problems with link speed to sites in China, Korea and Japan. This information is anecdotal rather than systematic and is usually brought to our attention when U.S. scientists travel abroad. This suggests that expectations by researchers at some foreign labs are still relatively low. It is not clear if the problem is with the connection from lab to national backbone or with the local area network at these labs.

Further development of tools, services and middleware would be particularly useful for support of international collaboration. The issues are similar to those needed for domestic collaboration but with added difficulty due to differences in technology,

standards and policies in the various political entities involved. Needed capabilities include:

1. Federated security: Technical and policy advancements to allow sharing of authentication credentials and authorization rights would ease the burdens on individual collaborating scientists. This sort of development is crucial for more complex interactions, for example where a researcher at one site accesses data from a second and computes on that data at a third site. (The fusion collaboratory deployed this capability for data analysis within the U.S. domain.)
2. Caching: Smart and transparent caching will become increasingly important as data sets grow. By the time ITER is in operation, this capability will be essential. Good performance for interactive computing and visualization will require optimization of caching and distributed computing. At the same time, complexity needs to be hidden from end users.
3. Document and application sharing: Improved tools for sharing displays, documents and applications are already urgently needed. Cognizance of different technology standards and policies will be important.
4. Network monitoring: We need to be monitoring the network backbone as well as end-to-end connections. Tools for testing and visualizing the state and performance of the network should be readily accessible.

Better voice and video conferencing tools will be essential to support these international collaborations. These include:

- Directory services
- Centrally administered conferences (call out)
- VoIP/SIP collaboration tools
- Screen sharing presentation tools
- Recording and Playback
- Instant messaging - collaboration and meeting setup
- Higher quality multipoint video
- Presence / availability information
- Better integration of above elements
- Integration with authorization tools

Better video conferencing support is also needed, including:

- Quickly determine the state collaboration services
- Ability to communicate this state to meeting participants

Wide Area Network Requirements and Science Process – the next 5 years

EAST: Over the next five years the operation of EAST will continue to expand both in the amount of data taken (the number of diagnostics will increase) as well as the amount of time that the machine is operated. The superconducting nature of the EAST tokamak allows for 24 hour a day operation for weeks at a time. There have been discussions between the U.S. and China for the U.S scientists to become actively involved in EAST's third shift operation (daytime in the U.S.). If this is pursued, then there is the possibility

of a greater increase in the breadth and scope of this collaboration as well as an increase in the amount of network traffic.

KSTAR: Physics research on KSTAR is anticipated to begin in the middle of 2008. In a similar fashion to EAST, as KSTAR continues to operate over the next five years more data will be available to remote participants and there will be greater opportunity to participate in experiments. In contrast to working on EAST, there has been no discussion regarding third shift operation of KSTAR. Therefore, we conclude for the time being, that the network requirements from the U.S. to EAST will exceed those of the U.S. to KSTAR.

Beyond 5 years – future needs and scientific direction

W7X: Located at the Max-Planck Institut für Plasmaphysik in Greifswald, Germany, W7X is a large (R = 5.5 m, a = 0.53 m, B = 3 T) superconducting modular stellarator device that is scheduled to begin operating in 2014. W7X will test the principle of "quasi-omnigeneity " for 3D shaped plasmas. The U.S. has an active stellarator program centered at PPPL and ORNL. The recent cancellation of a major U.S. stellarator will likely increase the importance of collaborations on this device.

JT60-SA: The JT-60SA ("Super Advanced") is a large, breakeven-class, superconducting magnet tokamak proposed to replace the JT-60U device at Naka, Japan. This program represents a coordinated effort between the EFDA and JAEA. Although there is a rich history of collaboration between the U.S. and Japan the extent of the U.S. involvement in this experiment is not clear at the present time.

ITER research phase: Though the ITER experiment will not start up for roughly ten years, detailed planning has begun for the research program and for the data and communications systems needed to support that program. Estimates on data volume are based on extrapolation from the current generation of experiments. A (hopefully) more accurate bottoms-up estimate will be carried out as work progresses on all ITER subsystems. Using a variety of methods, the current best guess is that ITER will acquire 1-10 TB per shot; 1-10 PB per year and will aggregate in the neighborhood of 100 PB over its lifetime. The requirements for off-site access have not been established, but the project is committed to full remote exploitation of the facility. Based on extrapolation from current practice, the project might be required to export 10-100 TB per day, during operation, with data rates in the neighborhood of 0.3-3.0 GB/sec (2.4 – 24 Gbps). At the same time, a steady level of traffic for monitoring and control will be expected. However, this should be less than 10% of the numbers quoted above. In all cases, some form of intelligent caching is assumed so that large data sets are sent only once over intercontinental links. With reasonable effort, the projected data volumes could be accommodated today, so they are not expected to present particular difficulties in ten years time, assuming adequate resources are applied.

On the other hand, coordinating research in such a vast collaboration will likely be a formidable challenge. Differences in research priorities, time zones, languages and cultures will all present obstacles. The sort of ad hoc, interpersonal communications which are essential for the smooth functioning of any research team, will need to be

expanded tremendously in scope. The hope is to develop and prototype tools using the current generation of experiments and to export the technology and expertise to ITER.

USNGE: (U.S. Next Generation Experiment) Discussions have begun on new facilities for the U.S. program. These would operate contemporaneously with ITER and fill gaps in our knowledge left by that project and the balance of the world program. The stated aim would be to put the U.S. in a position, technically, to build a demonstration fusion reactor following successful completion of ITER – if a political decision is made to do so.

Bandwidth Summary for International Collaborations

- near term (AUG, JET, EAST, LHD, JT60-U, KSTAR): 100 Mbps
- 5 years: (EASR, KSTAR, JET): 1Gbps
- 5+ years (ITER): > 10 Gbps

5 Findings

The following issues were reported and discussed at the workshop.

Connectivity Issues

There is a need to establish better connectivity between the U.S. fusion community and the Asian fusion experiments EAST (China) and KSTAR (South Korea). There may be significant issues in getting high performance to these sites, EAST in particular. ESnet is currently engaged in discussions with R&E networking partners with the goal of expanding connectivity to Asia – the needs of the Fusion community have been added to the discussion.

As in other areas of science, connectivity to the commercial Internet is increasingly important as more staff and scientists telecommute.

End-to-End Performance

A number of sites reported problems with end-to-end performance of data transfers. Much of the problem is that the current state-of-the-art tools and techniques are only partially deployed. This is due at least in part to a combination of the following: lack of knowledge about the new tools, lack of support and maintenance for some of the middleware and data transfer tools, and lack of an easy to use interface to some of the high-performance tools. Even for some tools that are now supported by the supercomputer centers (e.g. GridFTP), there is no clear long-term plan for funding support for development and maintenance of the code base into the future. Although long-term funding and support issues for software tools exist, there is a clear need for consistent deployment and regular testing of modern data transfer tools and hosts, as well as better tuning of the computers used in data transfer systems. In particular, the supercomputer centers should deploy and regularly test a common set of tools that provide a robust, easy to use, high-performance data transfer capability to enable users to easily move data sets between DOE supercomputing resources.

End-to-end issues are currently affecting the way in which some scientists analyze the data sets produced by large simulation runs. For example, if the entire data set cannot be moved from the supercomputer center to a local resource for some reason (e.g. data transfer performance problems, filesystem performance limitations, data set too large, etc), scientists must move a portion of their data to their local site and conduct their local analysis on this subset of the full simulation data. Because these issues are different for different sites, there is some discrepancy in the requirements numbers given in several case studies for codes such as GTC. A model is currently being developed that takes into account the relative network/data transfer performance, filesystem performance, computational capability and filesystem size profiles for data set analysis.

There is also a need for the consistent deployment and support of interoperable workflow tools to facilitate the movement of data sets between resources. Predictability and reliability are often more important than performance in the realm of data transfer tools. This is true both in the fusion community and in other scientific disciplines as well. Typically scientists are working on several tasks concurrently, and it is a boon to productivity to be able to set up a data transfer and allow the middleware to move the data, interrupting the scientist when the task has been completed – this allows the researcher to focus on other tasks during the transfer instead of "babysitting" the transfer. On the other hand, if the data transfer tools are not reliable and require constant monitoring, data transfer becomes an impediment to productivity. Unfortunately, the data transfer infrastructure at the major computational resources is not typically reliable enough to operate unattended.

Collaboration Services

Further development of tools, services and middleware is needed for support of international collaboration. The issues are similar to those needed for domestic collaboration but with added difficulty due to differences in technology, standards and policies in the various political entities involved. Needed capabilities include federated security, smart and transparent data caching systems, improved tools for sharing displays, documents and applications, and tools for testing and visualizing the state and performance of the network.

In addition to these new tools and services, there is a need for better diagnostic tools for video conferencing that would allow users to know things such as expected outage duration when the system is down – this would allow other avenues to be pursued when communication is time-critical.

Data Flow Model

When transferring data between sites for local analysis, there are complex issues to consider. In particular, since the full-size data set from a large simulation run might make the movement of the full data set prohibitive (either because of network bandwidth constraints, disk space constraints, remote site computing power constraints, or other constraints), simulation data sets often need to be reduced in size or sub-sampled before they are moved. However, the relationship between the various constraints of compute power, disk size, disk performance and wide area network bandwidth are not well defined. This can make it difficult for FES scientists to accurately forecast network

requirements, since the practical utility of a given amount of network bandwidth can be governed by other factors. At the moment, the FES community does not appear to have a comprehensive end-to-end model that captures these interactions.

Fusion Simulation Project (FSP)

The conceptual design phase of the proposed FSP is expected to start in FY 2009 and will be completed in 2 years. The FSP will result in the development of coupled models, and the increased coupling of models and simulations running on computational resources with running experiments at the experimental facilities. The design of the FSP might significantly change the data movement needs for some aspects of the DOE Fusion program.

ITER

The ITER computing model is not yet defined, but in general it is expected that computing and data will be co-located. However, there is also the expectation that data produced at ITER will be moved to remote sites (e.g. the supercomputer centers) for analysis. The ITER computing model will be better defined in 3 years, and the networking requirements for ITER science will be determined at the next ESnet requirements workshop for FES. There is some risk in postponing this planning, as it can take several years to procure new high-bandwidth transatlantic links if this is determined to be necessary.

6 Requirements Summary and Conclusions

A number of common themes emerged from the case studies and workshop discussions. One is that Fusion science, like many other disciplines, is becoming more and more distributed and collaborative in nature. Another common theme is that sophisticated collaboration tools are becoming more and more important, but are currently lacking. In addition, it is clear that ITER and the Fusion Simulation Project (FSP) will have significant impact on the network requirements of the FES community, but ITER and FSP are not yet well-defined enough to provide concrete estimates of their future network requirements.

The experimental fusion programs (DIII-D at General Atomics, Alcator C-Mod at MIT, Princeton Plasma Physics Laboratory) operate in a similar mode. The experiments are highly collaborative, with tight timelines for data analysis by a large team of local and remote collaborators. These place a significant demand on collaboration services such as video conferencing, and there is a need for enhanced videoconferencing services that is unmet by current technologies (neither the current commercial videoconferencing tools nor the current research pipeline address the situation of remote collaboration in a tokamak control room). In terms of bandwidth requirements, most sites that conduct data-intensive activities (the Tokamaks at GA and MIT, the supercomputer centers at NERSC and ORNL) show a need for on the order of 10 Gbps of network bandwidth for FES-related work within 5 years. PPPL reported a need for 8 times that (80 Gbps) in that time frame. Estimates for the 5-10 year time period are up to 160 Gbps for large simulations. Bandwidth requirements for ITER range from 10 to 80 Gbps.

In terms of science process and collaboration structure, it is clear that the Fusion Simulation Project (FSP) has the potential to significantly impact the data movement patterns and therefore the network requirements for U.S. fusion science. As the FSP is defined over the next two years, these changes will become clearer. Also, there is a clear and present unmet need for better network connectivity between U.S. FES sites and two Asian fusion experiments – the EAST Tokamak in China and the KSTAR Tokamak in South Korea.

In addition, several participants discussed the need for intelligent workflow or other back-end data processing tools that would allow for on-the-fly data reduction, subsetting, or other processing before streaming the data to a remote site. In many cases, scientists do not have local resources of sufficient scale to handle the full data set from a big simulation run, but there is a need to process a portion of that data set on a local resource.

Action Items

The action items for ESnet that came out of this workshop include:

- Work to help increase performance between FES user facilities and the Asian fusion experiments:
    - EAST at the Chinese Academy of Sciences Institute for Plasma Physics
    - KSTAR at the National Fusion Research Institute in South Korea
- Track ITER data distribution and computing models
- Track FSP data distribution and computing models
- Continue development and deployment of the ESnet On-demand Secure Circuits and Advance Reservation System (OSCARS - http://www.es.net/oscars/)
- Continue to add content to the web site http://fasterdata.es.net and continue to help users with end-to-end data transport issues

7 Acknowledgements

This work would not have been possible without the contributions and participation of those who provided information and attended the workshop. ESnet would also like to thank the FES and ASCR program offices for their help in organizing the workshop and providing insight into the science supported by the FES program. In addition, the LBNL conference support and logistics staff was very helpful.

ESnet is funded by the U.S. Dept. of Energy, Office of Science, Advanced Scientific Computing Research (ASCR) program. Vince Dattoria is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This work was supported by the Directors of the Office of Science, Office of Advanced Scientific Computing Research, Facilities Division, and the Office of Fusion Energy Sciences.