SCCS-624
Supercomputing Symposium '94, Toronto, Canada
June 6–8, 1994

InfoVision — A Prototype of the Future NII

Kim Mills, Geoffrey Fox, Marek Podgorny, Salim Hariri, Roman Markowski

Northeast Parallel Architectures Center, Syracuse University,
Syracuse, New York 13244–4100, kim@npac.syr.edu

1 Introduction

The National Information Infrastructure (NII) is seen as essential to U.S. national competitiveness, and necessary to supporting important societal needs. The physical bitways of the future NII will support multi-function, multi-use, and multimedia services which can be assembled in flexible ways for different applications. These applications include a broad set of "National Challenge" application in civil infrastructure, digital libraries, education and lifelong learning, energy management, the environment, health care, manufacturing processes and products, national security, and public access to government information [1]. A particularly important class of activities will center on decision support. This is illustrated by crisis management in government, simulation and information on demand for education, remote support for medical treatment (telemedicine) in healthcare, and multimedia community databases for personal and local planning decisions. These diverse problems will share multi-use services on the NII, impacting business, education, consumer markets, and the community.

The technical components of the NII include scalable high performance computers, advanced high speed computer communications networks, and advanced software. These general categories encompass a diverse set of enabling technologies required for information processing applications. Just as the Grand Challenge scientific simulations have been built upon high performance compilers, scientific subroutine libraries and other technologies, a similar set of enabling technologies will be required to build National Challenge information processing applications. We refer to these enabling technologies as InfoVision technologies for supporting multimedia information applications in an High Performance Computing and Communications (HPCC) environment.

We refer to this general environment as HPMMCC – High Performance Multimedia Computing and Communications. In this context, we are currently evaluating and demonstrating information on demand technologies and applications for a variety of information sources. The purpose of this paper is to describe InfoVision, a microcosm of the NII under development at the Northeast Parallel Architectures Center (NPAC). We apply an applications oriented approach to integrating enabling technologies, and building prototypes of information on demand systems, which when linked with the National Challenge areas, serve as prototypes of the future NII. Our recent focus has been on video and simulation on demand services enabled by HPMMCC. Our prototype development laboratory is based on high performance parallel and distributed computing systems, local area ATM networks, and NYNET— a high speed, wide area ATM network currently linking Rome Laboratory, Syracuse and Cornell, with a state wide extension expected soon.

2 High Performance Multimedia Computing and Communications

InfoMall

InfoMall is NPAC's industry outreach program, which is supported by New York State to apply high technology to economic development. InfoMall is an NPAC led partnership of approximately 50 industry, university, government laboratory, and small business software developers, entrepreneurs, and researchers. InfoMall is designed to exploit the core technologies of the NII— parallel computing, network communication, and software— and jump start the HPCC software industry. InfoMall emphasizes dynamic partnerships, efficiently gathering the best available technologies, evaluating, developing, and integrating enabling technologies, and building new HPCC systems and software products [5,9].

InfoVision

Based on our survey of industry over the period 1990–1993 [10], we identified information processing as the most promising opportunity for applying HPCC in industry. We define information processing very broadly, with four important classes: information production (e.g., simulation), information analysis (e.g., data mining), information access and dissemination (the focus of InfoVision), and information integration (e.g., decision support in business). This classification of information processing applications provides a context for InfoVision, and is described in more detail elsewhere [2,4].

InfoVision stands for **I**nformation, **V**ideo, **I**magery, and **S**imulation **on demand**, and is the central focus of our InfoMall program. InfoVision is wide in scope and necessarily encompasses a diverse set of technologies [8]. These core enabling InfoVision technologies include: parallel databases, high speed networks, multilevel mass storage, integration software, integration of parallel and distributed computing, multimedia support, ATM network protocols and management, compression, parallel rendering, collaboration services, and user interfaces. The basic activities of InfoVision include benchmarking hardware, software, and approaches (algorithms); technology evaluation and development; support of commercial enterprises; providing early production facilities; and demonstrating prototype systems to evaluate technologies and their integration.

We are building InfoVision to develop information access and dissemination prototype systems for a variety of information sources, including database information, text, digital images, video, and the output of simulation models. As a multimedia information server in an HPCC environment, InfoVision will be in position to deliver information on demand for industry, small business, education, the military, and the consumer. Currently, NPAC has 18 InfoVision demonstrations in place (Table 1). These demonstrations draw upon our InfoMall partners, and require a wide variety of base technologies, illustrating the InfoMall process outlined above. These demonstrations illustrate database information (e.g., DR-LINK linguistic based retrieval of text, ReFlex commercial multimedia systems), video (e.g., CNN video Newsroom), simulation (e.g., financial models, tornado model, and image (e.g., parallel JPEG image compression) on demand (see Appendix 1 Summary of InfoMall Projects).

We propose a similar approach to developing the NII. NII software should be built in a modular fashion to allow optimization for particular applications, and enable the involvement of small businesses. This is the InfoMall model of virtual corporations, with different system components developed by different organizations, within a coordinated framework. The systems must scalable so that they can be demonstrated on today's massively parallel systems, as NPAC is developing for NYNET, but scaled in the future to the much larger systems which will be deployed on the NII. Supported clients must include personal computers as well as more powerful workstations.

Table 1 NPAC Activities in High-Performance Multimedia Computing and Communications (HPMMCC)

InfoMall Projects	**Information on Demand Multimedia Applications**				**Enabling Technologies**					**Bench-marking**
	Text	Audio Video	Imagery	Simulation	Parallel Database	Net-working	Computer Architec-tures	User Interface	Pro-gram-ming Envir-onments	
SC'94 Multimedia Demo	X	X	X	X	X	X	X	X	X	X
Syracuse Language Systems	X					X				
ReFlex	X	X	X			X				
HTML Database	X	X			X	X	X	X		X
Orangesource	X	X				X				
DR-LINK	X									X
AskERIC	X	X	X		X	X		X		X
National Software Exchange	X								X	X
Comp. Science Education	X							X		
Comm-unityNet	X					X		X		
CNN VOD		X				X		X	X	X
Parallel Video Compression			X			X	X		X	X
Kodak Picture Exchange			X			X	X	X	X	
Tornado Simulation				X			X	X	X	X
Financial Simluation				X			X	X	X	X
Target Tracker				X		X	X	X	X	X
EM Simulation				X		X	X	X	X	X
Geographic Info. Systems				X		X	X	X	X	X

3 Current Video and Simulation on Demand Demonstrations

From the set of 18 InfoMall projects listed in Table 1, we focused our initial efforts on video and simulation on demand demonstrations, and use these projects to describe the current state of our InfoVision demonstrations, and our plans for the future.

Video on Demand

Our current Video on Demand (VOD) laboratory is based on a cluster of Sun and Silicon Graphics workstations linked via an ATM (Asynchronous Transfer Mode) local area network, supported by a FORE ASX-100 switch, providing direct ATM interfaces (bandwidth of several tens of Megabits/second). This switch is equipped with 8 TAXI (Transparent Asynchronous Transmitter/Receiver Interface) ports (140 Megabits/second each), and serves as a gateway between the LAN ATM cluster, and NPAC's parallel and distributed computational facilities. In addition, the switch has two Sonet OC/3 (SONET Synchronous Optical Network Optical Carrier) ports providing a 155 Megabits/second connection to the NYNET wide area ATM network.

Our current video on demand laboratory is based on two different hardware configurations: Sun IPX workstations (SunOS 4.1.3) with Parallax Video cards, and Silicon Graphics Indy workstations (IRIX 5.2) with Galileo cards. The Parallax card is a graphics video card for Sun workstations providing JPEG hardware compression for capture and display. The Galileo card provides video input, output, and effects for the SGI Indy workstation (compression is done in software). It is used to create professional video productions, and supports NTSC & PAL standards, s-video and composite formats.

The user interface is based on NCSA Mosaic client/server software. The compressed video data is stored on the SGI Challenge file server, and is delivered to the digital video browser via NFS protocol. The browser, written at NPAC with the use of the XView toolkit, offers full random access to the video material. The SGI solution is entirely software-based. We implemented a digital video browser using the SGI movie library and the MVC1 (Motion Video Compressor 1) compression scheme. MVC1 is a Silicon Graphics proprietary algorithm used by the MovieMaker/Player tools. MVC1 is a fairly lossy algorithm that does not produce compression ratios as high as JPEG, but is well suited to movies.

We used CNN (Cable New Network) Newsource as the content of the video material. We captured and digitized a video tape, and implemented the search and browsing capabilities of the Mosaic interface, providing hyperlinks between the video clips and text indices in HTML documents. In both solutions, we used Internet Protocol over ATM transport and a frame rate of 15 frames/second. One minute of a video clip requires approximately 18.5 MB of disk space (Sun, JPEG compression) and 34 MB (SGI, MVC1 compression)

Figure 1 (top) illustrates our current CNN Video on Demand demonstration. A digital video browser started from a Mosaic-based user interface allows simultaneous display of multiple video clips on a Silicon Graphics Indy display.

Over the next 12–24 months, we will build on our existing workstation based prototype described above to construct a video service testbed and evaluate and develop video input, compression, storage, delivery, and search schemes in an HPCC environment. Of commercial importance is the ability to deliver multiple video streams to multiple users over a wide area.

To carry out this plan, we propose a system architecture for video on demand which consists of the following components: multilevel storage, video server hardware and software, data input, video network services, user interface, and system integration (Figure 2). Our focus is on the hardware and software of the VOD server, network services, and system integration. We will expand our prototype VOD server based on workstation technologies and linked by an ATM LAN to develop HPMMCC systems based on leading MPP architectures and the NYNET regional ATM gigabit network. We intend to take advantage of national development projects and technologies such as the DoE National Storage Lab initiative, NCSA Mosaic network browsing software, and state of the art satellite video receiver and capture facilities.

As illustrated in Figure 2, a number of base technologies are associated with each of the VOD system components. For example, the server technologies include hardware and software based video compression codecs. Storage schemes for providing fast access disk storage for real-time VOD applications include streamlined disk devices, RAID systems, parallel database support, and tertiary storage facilities. A critical issue to industry application of a movie on demand server is an evaluation of the cost performance of parallel, distributed, and heterogeneous server architectures. A backend storage facility for VOD applications requires a storage hierarchy including main memory cache, attached disk storage, archival storage, and network accessible storage. This hierarchy of storage levels must be managed to provide fast access to video in high demand stored in cache and on disk, as well as access to large collections of source material on both local and remote network servers (e.g., a Syracuse server might store the equivalent of 1,000 movies, while a Hollywood server might store the equivalent of 100,000 movies).

Simulation on Demand

As a parallel computing center and part of the national HPCC community, NPAC has a great deal of experience in scientific simulation applications. Our current simulation on demand laboratory is based on a variety of Grand Challenge type scientific simulation models running on a variety of high performance parallel and distributed systems.

At NPAC, these high performance computing systems include a Connection Machine-5 (32 node), an IBM SP1 (16 nodes), and Intel iPSC (16 nodes), an NCUBE2 (64 nodes with 200 GB disk), two DECmpp (8K and 16K systems), a Digital Alpha cluster (8 node), and IBM RS/6000 cluster (8 nodes). The output of these simulation models is currently displayed via an NCSA Mosaic network interface running in the LAN ATM environment.

As part of our InfoVision simulation on demand demonstrations, we implemented a Grand Challenge tornado prediction model [7], chemistry transport in the atmosphere (NASA Goddard HPCC program), electromagnetic scattering simulation [6], and stock option pricing model [11,12]. These applications are extremely demanding in terms of computational resources and require parallel processing facilities. The scientific computing issues in each of these applications are relatively well understood. The issue of relevance to InfoVision, and simulation on demand as a class of information processing applications concerns delivering the output of these simulation models running on high performance computers over networks to users in a wide geographic area.

Our current demonstrations are based on a NCSA Mosaic interface running on workstations in NPAC's ATM LAN (the video on demand laboratory described above). By clicking on one of the hyperlinks on the Mosaic client page, one can start real-time processes on network connected (currently on a LAN) parallel computers including the CM–5, DECmpp or IBM SP1. For example, we implemented a parallel version of a Grand Challenge tornado prediction model. The results of real-time calculations are incorporated into an AVS graphic user interface to provide three dimensional rendering and interactive model control. We plan to deliver the output of this model on-demand over networks to schools giving the unique opportunity for students to manipulate a the behavior of a tornado in an interactive session. Figure 1 (bottom) illustrates simulation on demand using an interactive remote visualization environment for electromagnetic scattering and tornado simulations on a high performance computing system.

Figure 1. Video on Demand (top) shows a digital video browser displaying multiple video clips.
Simulation on Demand (bottom) shows an interactive visualization of electomagnetic and tornado simulation.

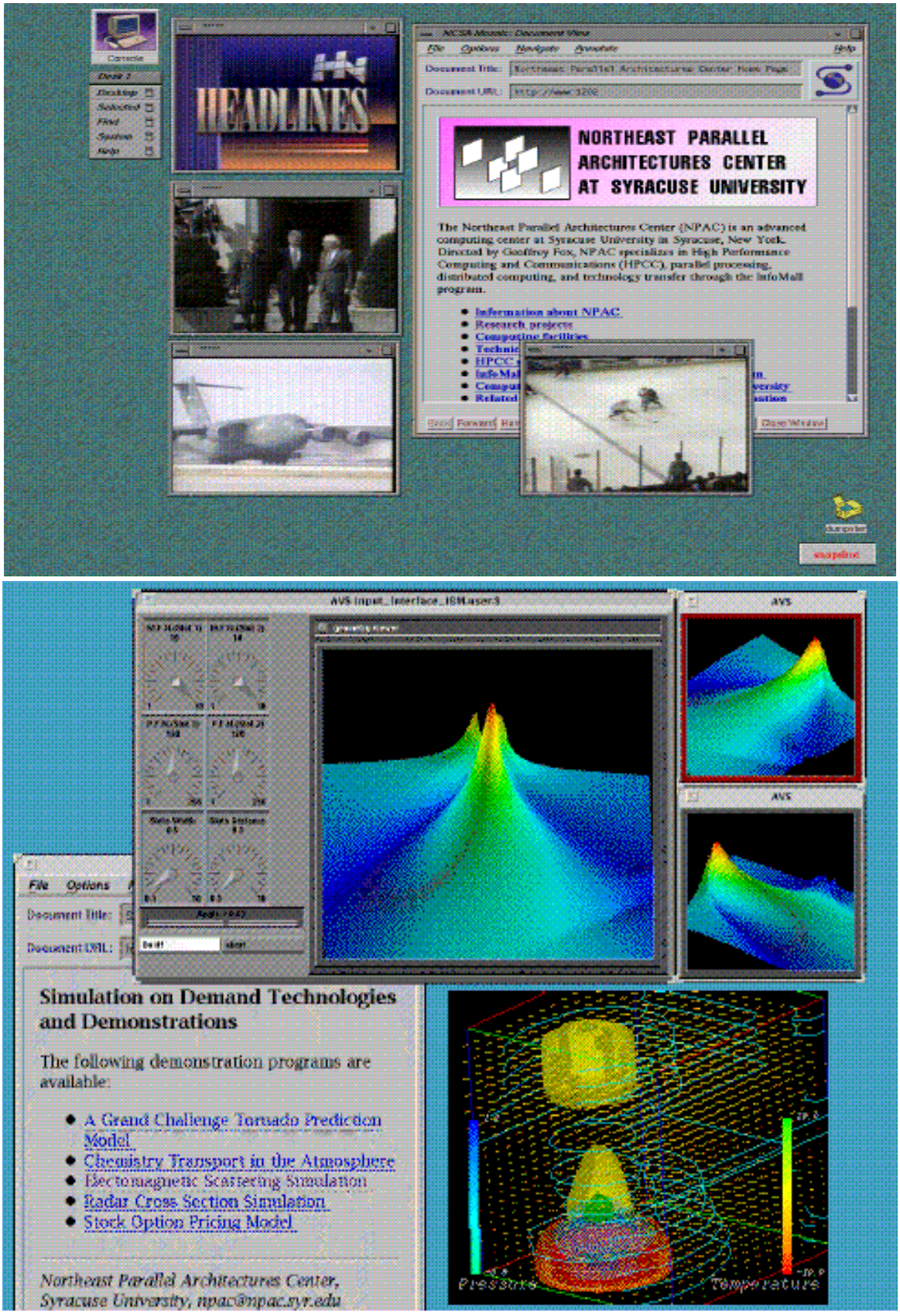
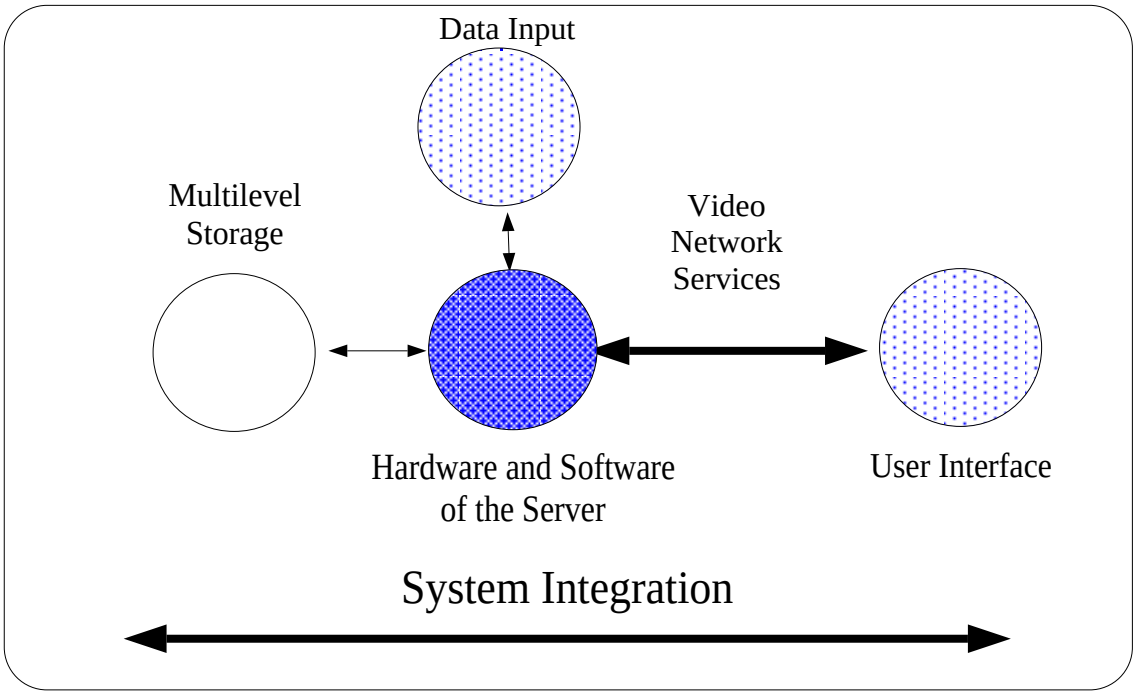


Figure 2. A Possible Video on Demand System Architecture and Associated Base Technologies



Multilevel Storage

Video
- storage
- indexing
- browsing

Network Services
- message passing

Database Technologies
- file systems
- memory management
- multimedia storage
- retrieval
- intelligent agents

Server Technologies

Video
- compression
- storage
- indexing
- browsing
- distribution

Network Services
- message passing
- hardware cards

Database Technologies
- file systems
- memory management
- parallel database
- multimedia storage
- indexing
- retrieval
- intelligent agents

Computer Architectures
- parallel
- distributed
- heterogeneous

Benchmark Evaluation

Video Network Services

Video
- browsing
- distribution

Network Services
- data transport protocols
- message passing
- hardware cards

Database
- parallel database
- multimedia storage
- retrieval
- intelligent agents

4 Summary and Conclusions

From our experience in HPCC technology transfer to industry, we consider information processing applications as the most important market opportunity for HPCC technologies, and InfoVision an example

of the ultimate client server application. Assuming a total of 100 million clients, each of which could be a PC class client, video game, or settop unit running at 100 megaflops in every home and office, would require perhaps 1,000 to 10,000 high performance multimedia servers (providing 1 teraop/second performance and costing on the order of $10M each). This configuration of information servers will be one to two orders of magnitude larger than current supercomputing installations. This large scale client/server application will also help create commodity markets for workstations, compact disk players, settop units, and virtual reality interfaces. A home technology mass market will drive down the cost of higher end versions of the same technologies used in the small business environment.

Existing media, entertainment, telephone, cable, computing, and video game industries provide the necessary technologies and business partnerships to establish digital information services for what many observers expect will be the initial offering of the NII— entertainment, news, and interactive television (e.g., shopping). These applications represent extremely large markets, justifying the capital investment by private industry, as U.S. federal policy requires, to put the National Information Infrastructure in place. Once this occurs, we expect great opportunities for business, research, education, and community use of the NII. All of this will stimulate an HPCC software industry which InfoMall is designed to support.

We propose InfoVision, our information technology testbed development environment as a prototype of the NII. InfoVision is based on our approximately 50 business, technology, and research partners, a variety of multimedia information sources delivered in an HPCC environment, and a diverse set of base technologies. We are developing InfoVision in a modular design with multiple development partners, and propose the same approach to developing NII software. This design allows the best available technologies to be easily incorporated, and facilitates the participation of innovative small businesses.

InfoVision specifically supports the issue of information access and dissemination. We foresee a future model of data rich, network based information servers supporting relatively data poor but computing rich home and office environment by supplying hundreds of thousands of small messages at bandwidths of approximately 10 Megabit/second over the NII. This model suggests a role for centralized computing with information rather than compute cycles as the centralized resource.

The component technologies of high performance multimedia computing and communications systems are all available. Our current task to integrate, evaluate, benchmark these prototype systems based on real applications that initially address entertainment and individualized information services, but will soon expand to provide exciting new opportunities in business, education [3], and community affairs.

5 References

[1] 1994. High Performance Computing and Communications: Toward a National Information Infrastructure. A Report by the Committee on Physical, Mathematical, and Engineering Sciences. Federal Coordinating Council for Science, Engineering, and Technology. Office of Science and Technology Policy. Sally E. Howe, editor.

[2] 1994. G. Fox and K. Mills. Opportunities for HPCC Use in Industry: Opportunities for a New Software Industry. chapter to be published in Parallel Programming: Paradigms and Applications. Chapman and Hall Publishers, London. U.K. A. Zomayae ed. SCCS-617.

[3] 1994. G. Fox and K. Mills. High Performance Computing and Communications–Yet another revolution in education. 4th Annual 1994 IEEE Mohawk Valley Section Dual-Use Technologies and Applications Conference.

[4] 1994. G. Fox and K. Mills. Information Processing and Opportunities for HPCC Use in Industry. The International Conference and Exhibition on High Performance Computing and Networking HPCN Europe'94. Munich, Germany. April 18-20. 1994. SCCS-610.

[5] 1994. K. Mills and G. Fox. InfoMall: An Innovative Strategy for High-Performance Computing and Communications Applications Development. to be published in electronic networking: research, applications, and policy.

[6] 1993. G. Cheng, Y. Lu, G. Fox, K. Mills, T. Haupt. An Interactive Remote Visualization Environment for an Electromagnetic Scattering Simulation on a High Performance Computing System. Supercomputing '93. pp. 317-426. SCCS-467.

[7] 1993. K. Droegemeier, M. Xue, K. Johnson, M. O'Keefe, A. Sawdey, G. Sabot, S.Wholey, K. Mills, and N.T. Lin. Design and Implementation of a Scalable Parallel Storm-Scale Numerical Weather Prediction Model. to appear in High Performance Computing. Addison-Wesley Publishers. November, 1993. 47 pp.

[8] 1993. G. Fox, S. Hariri, R. Chen, K. Mills, M. Podgorny. InfoVision: Information, Video, Imagery, and Simulation On Demand. Syracuse Center for Computational Science Technical Report-575.

[9] 1993. Fox, G., Bogucz, E., Jones, D., Mills, K., and Podgorny, M. InfoMall Program Plan: A Scalable Organization for the Development of HPCC Software and Systems. SCCS–531. (Available from NPAC, Syracuse University, Syracuse NY 13210-4100).

[10] 1992. Fox, G. Parallel Computing in Industry: An Initial Survey. in Proceedings Fifth Australian Supercomputing Conference. World Congress Center. Melbourne, Australia. SCCS–302, Center for Research on Parallel Computation-TR92219. (Available from NPAC, Syracuse University, Syracuse NY 13210-4100).

[11] 1992. K. Mills, G. Cheng, M. Vinson, S. Ranka, and G. Fox. Software Issues and Performance of a Parallel Model for Stock Option Pricing. The Fifth Australian Supercomputing Conference, December 7-9, 1992, Royal Melbourne Institute of Technology and The University of Melbourne. SCCS-273b.

[12] 1992. K. Mills, M. Vinson, and G. Cheng. A Large Scale Comparison of Option Pricing Models with Historical Market Data. The Fourth Symposium on the Frontiers of Massively Parallel Computation, October 19-21, 1992, McClean, Virginia, IEEE Computer Society and NASA Goddard Space Flight Center. SCCS-260.

1 Summary of InfoMall Projects

SC'94 Multimedia Demonstration: is a demonstration proposed for Supercomputing '94 designed to show text, video, imagery, and simulation information on demand demonstrations in a high performance computing and communications environment. NPAC will lead this demonstration, but will need the collaboration of Rome Laboratory, NYNEX, and a number of other InfoMall partners.

Syracuse Language Systems: is a multimedia, PC based learning environment for teaching children new languages (French, Spanish, German, English). In our InfoMall project, we store the multimedia database supporting this system on an NPAC InfoVision server, and distribute the data to multiple school sites over a wide area ATM network.

ReFlex: is a Syracuse based computer animation company and InfoMall small business partner. ReFlex developed a multimedia community yellow page directory which is distributed in kiosks located in public sites. Future plans include linking to InfoVision servers at NPAC to deliver parallel rendering and video information over a network.

World Wide HTML Database: NPAC proposes to gather network distributed HTML documents to develop a centralized database for a set of data mining experiments.

Orangesource: is a university campus wide multimedia information system based on Macintosh servers supporting kiosk sites.

DR-LINK: is an ongoing research project in Syracuse University School of Information Studies. DR-LINK is an ARPA funded text retrieval system based on natural language processing. This system currently runs on a workstation. Future plans are to port this NPAC high performance computing and communications environment.

AskERIC: is a U.S. Department of Education funded information clearinghouse accessible over Internet and used by K-12 teachers around the nation.

National Software Exchange: is a NASA funded project to gather software templates from the HPCC community and create a "consumer reports" type evaluation for feedback to users.

Computational Science Education: is an NPAC led project to integrate computational science training within traditional academic disciplines. A set of network accessible training material is available from NPAC information on demand servers.

CommunityNet: a Central New York community information system is being developed on the FreeNet software model. NPAC is participating in this development effort and will integrate InfoVision demonstration projects as the network evolves.

CNN Video on Demand: this InfoVision video on demand project is currently based on a local area ATM network and workstation platforms. The Rome Laboratory VOD project will further develop this demonstration for an HPCC environment.

Parallel Video Compression: this project is based on parallel software for JPEG compression and runs portably on MPP systems.

Kodak Picture Exchange: this is a commercial software system currently marketed by Kodak to the stock photography market. Currently, digital images are stored in a central database and delivered over a phone line to a Macintosh client. Icon and keyword matching are used to query the image database.

Tornado simulation: is a Grand Challenge scientific simulation developed at the University of Oklahoma and "parallelized" at NPAC. We intend to use this model in entertainment and educational applications.

Financial simulation: is a set of stock option pricing models running in a heterogeneous computing environment including workstations, SIMD, and MIMD parallel machines.

Target tracker: is a component of a military command, control, and communications system for real time data fusion and decision support. This demonstration run in a distributed computing environment.

Electromagnetic Simulation:this project uses parallel and distributed computing environments to simulate and display aperature electromagnetic scattering, a related project predicts the radar cross section around an arbitrary three dimensional object.

Geographic information systems: this project will combine three- dimensional spatial databases, parallel rendering software, and links to multimedia databases.