SCCS-617
to be published in
Parallel Computing: Paradigms and Applications
Chapman and Hall Publishers, London, U.K. 1994. in press

Opportunities for HPCC Use in Industry:
Opportunities for a New Software Industry in HPCC

Geoffrey Fox and Kim Mills
Northeast Parallel Architectures Center
Syracuse University
111 College Place, Syracuse, NY 13244-41000
gcf@npac.syr.edu   kim@npac.syr.edu

1 Introduction

1.1 Opportunities for HPCC

Current trends strongly suggest that all computers, from personal computers to supercomputers will become parallel computers. During the 1990's, we expect a thousand fold increase in computing performance. Combined with a similar level increase in performance of communication technologies, high-performance computing and communications (HPCC) will have wide ranging impacts on industry, education, and society in general. Given the rise of information services and knowledge based industries, a competitive advantage will be gained by industries that most effectively exploit HPCC in global economic markets. Computation will emerge as a fundamental methodology in science and engineering education. The opening up of the consumer or home market, initially for entertainment, will be quickly followed by shopping, personalized information services, and access to community services. All of these developments will be based on widespread use of HPCC technologies.

Small software businesses, and small business units in larger corporations will play a key role in developing these markets. Small businesses are in the best position to quickly move in response to market opportunities, and form dynamic partnerships to combine different types of business and technological expertise. Additionally, small businesses are not burdened by the social obstacles to technological innovation that large corporations are currently struggling with.

The U.S. National HPCC effort is represented by a set of approximately fifty Grand Challenge science problems. In addition to advancing the mission of the federal agencies sponsoring this program, and the state of science in the field of physics, biology, and chemistry, and related fields such as atmospheric science (Grand Challenges, 1993), one can view the HPCC initiative as an effort to speed the process of technological innovation. Data parallelism provides a universal form of parallelism, and can be implemented on both SIMD and MIMD architectures. Hardware systems are now relatively mature, and the HPCC program has successfully accelerated the development of new software technologies such as parallel compilers, scientific libraries, system software, and programming tools. HPCC has matured into a proven technology (Fox, Messina, and Walker, 1994, but its success is mainly limited to applications in the research laboratory.

1.2 Obstacles

When will HPCC become adopted as a mainstream technology? For most new technologies, a ten fold improvement in performance is needed to entice users to make the transition. Important differences between industrial and academic research codes, and software development issues present additional obstacles to industry adoption of HPCC.

In many industrial applications, the porting of many millions of lines of existing code from sequential to parallel platforms is an insurmountable difficulty. Larger codes, greater complexity, and longer life cycles make it much more difficult to convert industry codes than academic research codes (Fox, 1992, Mills and Fox, 1992). Systems and application software is not as mature on HPCC systems as on more conventional computing systems. An often repeated complaint by industry is that the lack of software tools and interfaces makes it too difficult to develop codes on HPCC systems. For information based industries, human capital is a critical asset, and the threat of wasting expertise during a software implementation phase is considered too risky.

The scientific community has focused on high performance levels, while industry places much more importance on reliability and ease of use. Perceived risk by industry executives of adopting "exotic" HPCC technologies, a general recommendation not to put "mission-critical" industry applications on HPCC systems, and pressure to maximize short term or end of the quarter earnings (CAPPS, 1993) are examples of the obstacles to HPCC application development in industry.

While we acknowledge that important obstacles remain to be resolved, we also expect that widespread industry application of HPCC will soon occur. We use the term industry applications of HPCC to represent broad application of these technologies in diverse sectors such as manufacturing and health care, as well as education, consumer, and entertainment markets, since it is industry that is most likely to deliver technologies in these areas. Aside from issues such as performance vs. reliability, or ease of use, which are essential to the success of HPCC technologies in industry, there is a larger more pervasive issue holding back adoption of HPCC.

1.3 Information Processing Applications

Scientific simulation is the central focus of the national HPCC community. Enabled by Teraflop computers, the Grand Challenge program has produced important basic research results, and pushed development of new technologies. Results of Grand Challenge applications are potentially important to industry, but there are no Grand Challenge applications in areas of large commercial markets such as economic modeling, scheduling, manufacturing, education, entertainment, and decision support. There is also no federal HPCC funding for technology transfer to industry.

We believe the most important opportunity for HPCC technologies lies in information processing and not simulation applications. This conclusion is based on our interaction with industry over the past three years which we summarize in the form of a survey and describe in section 3 below.

The importance of information processing applications in industry is becoming a central issue in many national HPCC meetings such as the industry panel discussions at the Pittsburgh HPCC Workshop (Workshop on Grand Challenges, 1993), the first conference on Commercial Applications of Parallel Processing Systems (CAPPS, 1993) the National Information Infrastructure workshops at Supercomputing '93 (NIIT Panel, 1993) and the industry/HPCC consortia meetings (RCI, 1993). This issue is now clearly understood with the 1994 HPCC federal plan (High Performance Computing and Communications, 1994) emphasizing simulation in the so called Grand Challenge but Information intensive applications and the NII (National Information Infrastructure) in a new set of National Challenges.

1.4 Technology Transfer Issues

For HPCC to deliver on its promise, it must be widely adopted and used in industry, and for this to happen, HPCC research and development must be more strongly integrated with the needs of industry. Over the past three years, the U.S. National HPCC effort has emphasized basic research or production of new knowledge, again largely in applications of scientific simulation.

Policy experts point out that U.S. government stimulation of technical innovation has been most successful in space and defense, but is largely unsuccessful in civilian applications. Research results alone are not useful to industry, and we cannot simply assume that good technologies sell themselves. Rather, we must take steps to ensure that the links between HPCC technology producers, transfer agents, and users occur (Pinelli et al., 1992). We do not elaborate on this process here but have designed InfoMall, our outreach program for delivering HPCC technologies to industry, using a proactive approach. In a set of steps, our InfoMall technology transfer process is designed to link: HPCC vendors; industry, government, and university research institutions; and potential HPCC technology consumers such as manufacturers, school districts, or software companies developing products for homes and small businesses (Fox et al. (a), 1993, Mills and Fox, 1994).

1.5 Information Classification

This paper describes a classification of industry applications carried out over the past three years as part of our New York State funded industry outreach program (Fox, 1992) We define four classes of applications based on types of information processing applications: information “production” or simulation; information analysis or data mining; information access and dissemination; and information integration or decision support.

We conclude that a core set of HPCC technologies---parallel architectures, software, and networks--- combined with wide range of enabling technologies from partial differential equation solvers and mesh generators, to video compression and parallel databases, will form the basis of the National Information Infrastructure of the future. While scientific simulation will continue to be essential, information dissemination and analysis applications will become the most important near term, and information integration applications will become the most important long term industry applications of HPCC technologies.

2 The National Information Infrastructure (NII) and New Opportunities for HPCC in Industry

2.1 Information Industries

Access to information, and the ability to organize and integrate information is playing an increasingly important role in education, business, world affairs, and society in general. This trend will result in applications such as global networks of 800 telephone numbers, instantaneous credit card verification systems, extremely large real-time databases, sophisticated scientific simulations in the classroom, decision support in the global economic war, early warning of political instabilities, and integration of Medicaid databases to streamline processes and uncover inefficiencies.

Information based industries are rapidly expanding, growing two and a half times faster than the goods economy. Raw materials and labor now account for less than 25% of the cost of goods sold, while the cost of knowledge and information based services contributes to more than 50% of the costs of goods sold for *manufacturing* companies (Johnson, 1993). A competitive advantage will most certainly be captured by industries that can take advantage of HPCC technologies. The winners will be the ones who organize knowledge networks, and build the expertise for harnessing the expertise of knowledge workers rather than managing physical resources or product flows as in a good based industry (Ives and Jarvenpaa, 1993).

2.2 The NII and What it Brings

The NII will give us a 1,000 fold increase in communication performance, which is the difference between a 10 kilobit/second modem, and a 10 megabit/second fiber link. This increase in communication performance will enable new information access and dissemination industries. The major NII highway will likely be implemented as ATM (optical fiber) based technology supporting communication rates at a large but unknown number of gigabits/second performance. Very high speeds are required to support traffic of many small messages generated by a large community of users, just as large vehicular highway systems are required to support a large number of individual vehicles. ATM switches will connect the major NII trunk lines to cable networks reaching the home or office end users. This system must deliver two-way communication with data into the home at rates of approximately 10 megabits/second (compressed video rates), with much lower rates from the home back out to the information server to support interactive services.

Combining a thousand fold increase in communications performance, with a similar increase in computing performance, will enable a transition in the consumer or home information technology market from an analog TV and settop unit to a digital, interactive, home computer, with high definition display, CD ROMS, and virtual reality peripherals. Mass production of powerful home computers (which will soon become parallel computers) and virtual reality interfaces will make them inexpensive, commodity market products.

The home entertainment market (approximately $8B in video games, $24 in home video) is considered by many to be the most important initial market, and essential to funding the connection between the NII highway and the "last mile" to the home. Many observers point out this home entertainment is the only "killer application" on the horizon, and that additional applications remain uncertain (NIIT Panel, 1993). We expect that the home entertainment market will open vast new business opportunities for information services, HPCC vendors, and small software businesses. With every home and business having the equivalent of today's $20,000 workstation, with suitably powerful storage and interface devices, new HPCC based information products such as electronic banking and investment, customized information services, realistic three-dimensional imaging for education and entertainment (or their combination --"edutainment") will become available.

A critical technological issue we still face is how to integrate the various system components required to deliver interactive, multimedia, information sources as a marketable price. InfoVision, described in section 4.1 below, is NPAC's main application development project and is designed to address issues of multimedia evaluation, integration, and performance in an HPCC environment (Fox et al. (b), 1993) .

2.3 HPCC Enabling Technologies

The HPCC Grand Challenge program, which is largely focused on scientific simulation, is built upon and responsible for furthering the development of a core set of HPCC enabling technologies. These technologies are certainly familiar to the HPCC community, and include message passing libraries, high performance compilers, parallel operating systems, and numerical optimization software (Table 1.1).

Information processing applications will similarly be built upon a set of core HPCC enabling technologies. These technologies are becoming more widely known to the HPCC community, but have yet to be extensively tested, developed, and integrated as systems in an HPCC environment. Enabling technologies include parallel databases, high-speed networks, metacomputer integration software, multimedia support, and compression technology (Table 1.2). Our InfoVision program (section 4.1 below) creates a testbed environment for evaluation, development, and integration of these enabling technologies for information processing applications.

2.4 Industry Survey of Applications and Required Technologies

Over the past three years, a focus of our industry outreach program has been to survey potential opportunities of HPCC in industry, and to understand application problems in terms of their implementation on parallel computers. Understanding the spatial and temporal structure of an application problem reveals the parallelism inherent to the problem, and is the basis of implementing industry applications on parallel hardware/software architectures (Fox, 1992).

The trends in information industries cited above, and observations about the potential of information technologies in a range of applications prompted us to re-examine our initial industry survey of opportunities for HPCC in industry. Our original survey grouped industrial applications into 18 general areas, including for example computational fluid dynamics, electromagnetic simulation, particle transport problems, transaction processing, command and control and information processing, financial modeling, graphics, and flight simulation. For each application area, we defined example problems, and summarize the nature of parallelism inherent to the problem, and match this to appropriate machine architectures and software models. Although a number of information processing applications were included, our initial survey strongly reflected the emphasis on simulation problems in the HPCC community, a trend which continues to the present day. In (Fox, 1992) we made the observation that information processing applications are the most important opportunity for HPCC application development in industry.

We now redefine our industry survey of HPCC applications according to four types of information processing applications.

3 Classification of Industry Applications Based on Information

We begin with a very broad definition of information, including both CNN Headline news, and insights gotten from new computational physics models. We define four classes of industry applications of HPCC based on information. These categories, listed in Table 2 HPCC Applications in Information Processing, include: information "production" or simulation; information analysis or "data mining"; information access and dissemination, and information integration. For each class, we list in Table 2 application examples, comments on the application problem, and characterize the parallelism present in the problem and the natural hardware/software architectures for supporting the problem.

3.1 Information Production

Information "production" or simulation has traditionally been the main focus of the national HPCC community. Computational fluid dynamics, structural analysis, and electromagnetic simulation (items 1,2,3 in Table 2.1), are examples of simulation problems with relatively large federal research programs funded by U.S. agencies such as the National Aeronautics and Space Administration, and the Department of Energy.

Scientific simulation applications have advanced the development of sophisticated parallel software technologies such as finite element, and finite difference partial differential equation solvers, and turbulence and mesh generators. Although this class of applications has pushed development of new technology, market opportunities are limited. An obstacle to HPCC industry applications is that many industries that use simulation technologies, such as the aerospace industry, are in a period of contraction rather than growth. This makes transfer of new technology to industry extremely difficult. Simulation applications in industries experiencing growth, and therefore more amenable to funding new technological approaches, include mortgage pricing and stock option pricing in the financial industry (item 9, Table 2.1), and systems simulations in defense (SIMNET), education (such as the personal computer game SIMCITY), and multimedia/virtual reality in the entertainment industry (item 13, Table 2.1).

An economically important information production application we have worked with is network simulation in the electric power industry (item 10, Table 2.1). The problem is to define a dynamically changing optimum network of power generators in a transmission grid as power consumption fluctuates, and available power generators go on or off-line. As we list in Table 2.1, and describe in section 4.3, the computational problem to solve in this application requires a parallel sparse matrix solver. The temporal structure of the computation is loosely synchronous, meaning that the decomposed network of nodes must intermittently synchronize with each other. This problem naturally maps to a MIMD architecture, and is best expressed in message passing Fortran (or C).

3.2 Information Analysis

The purpose of information analysis is to enhance the value of data by reducing uncertainty. In commercial applications, this allows companies to more efficiently find oil, and market new financial products to profitable groups in the population. Increased performance in information technologies allows increased productivity by shortening decision cycles and allowing larger and more sophisticated analyses. HPCC information technologies are being applied in new businesses, as well as changing the way business is being done.

"Data mining" or the extraction of information from very large databases is a well established practice in the petroleum industry (item 14, Table 2.2). As in other "event" analysis problems such as high energy physics, seismic data records present a large number of embarrassingly parallel events to analyze, and may be implemented on both SIMD and MIMD architectures.

Newer data mining applications that were not possible only two to three years ago include credit card, securities, and health care payment databases (item 16, Table 2.2). Owners of these databases are concerned with ways to maximize access to the database, structure information in ways that create new meaning, and developing strategies for fast response to unique customer requests or new products (Ives and McKeown, 1993). These applications use linkage analysis of the database to find correlations between records. For example, a bank marketing credit cards wants to find new customers with a similar profile to existing high profit card holders, as well as market new products to their best existing customers. Finding correlations requires high-performance information systems as databases are large (tens of millions of records), and near real-time analysis is important in the decision cycle of analysts and marketing teams. These applications are well suited to SIMD architectures where the records are identical, or can be implemented as loosely synchronous collections of record domains on MIMD architectures. Other similar commercial applications include finding efficiencies in insurance policy databases to market products, and inefficiencies in health care payment databases to detect fraud.

Powerful, new, HPCC based information technologies such as high-speed networks, parallel databases, and hierarchical storage technologies provides tools for doing business in new ways, but also change the way business is done. Information and technologies tend to be most useful when they are right at hand. A breakthrough of the 1980s was the shift between large mainframes providing impressive increases in power and productivity for a few, to desktop technology providing distributed, inexpensive, and widely accessible computing power. Decentralized technologies push information down to the end user, such as the financial analyst in a bank, who can then make more immediate, creative, fact-based marketing decisions.

This "sense and respond" approach to marketing information based products (e.g., credit cards, insurance policies) based on managing and deploying knowledge and information is quite different from a "make and sell" approach used in the industrial age (Haeckel and Nolan, 1993) . Changes in the business operation are required in order to move from centrally controlled databases on mainframes to distributed databases controlled by the business end user. Successful organizations of the future will be "network-centric" (CAPPS, 1993) . To provide the necessary flexibility for using new information technologies, such as building powerful databases that span departments and divisions, organization will resemble networks rather than hierarchies (Ives and Jarvenpaa, 1993). InfoMall, our technology transfer program for delivering HPCC technologies to software developers is also organized as a network. We use the name InfoMall to communicate the powerful analogy of a retail shopping mall to explain our approach to linking emerging technologies, “virtual corporations,” business support, and new markets (Figure 1).

Market segmentation (item 17, Table 2.2) is an information analysis application aimed at tailoring traditional mass marketing approaches down to the level of the individual, producing “mass customization” for products ranging from banking services, news media, and retail consumer products. These problems required sorting and classification algorithms, and tend to be loosely synchronous in structure. Implementation on both SIMD and MIMD architectures is possible, with MIMD perhaps better suited to sorting tasks.

3.3 Information Access and Dissemination

Information access and dissemination covers a wide range of applications generally requiring information on demand capability. We have defined InfoVision, a major initiative at NPAC to develop information on demand technologies and applications. InfoVision stands for four important categories---database Information, Video, Imagery, and Simulation on demand (see section 4.1). The InfoVision activities are "new" applications and do not require porting of existing software. Thus HPCC technologies will not face some of the social and technical problems seen in other areas. The cost-performance advantage of parallel computing should dominate purchase decisions in these naturally parallel applications.

Information access and dissemination includes text on demand or document retrieval from educational clearinghouses by school teachers, and accessing the world’s news media and government reports by political analysts to predict political instabilities. Three dimensional geographic information systems linked with multimedia map databases are a form of simulation on demand, and could be used to deliver educational, tourism, or city planning information to students, the public, or government officials. Video on demand will soon become a large, new commercial market. Information access and dissemination applications are enabled by the National Information Infrastructure, which in the near future will link compute servers, and distributed databases over high-speed network links to the home and office settop box (soon to become a powerful PC).

Data base information on demand includes categories of information such as transaction data, and full text sources. Transaction processing of automatic teller machine records (item 18, Table 2.3) has real-time importance, and is directly linked to the data base mining application for credit card analysis (item 16, Table 2.2) described above. Transactions are short database records, are an embarrassingly parallel processing problem, and are best suited for implementation on a MIMD database system. Real-time analysis is essential to account verification at the point of sale. Transaction records include amount of purchase, as well as time of purchase, product category, and geographic location. Combining time series transaction data with additional customer demographic data adds "value" to the database and becomes information integration.

The Educational Resources Information Center (ERIC) is a U.S. Department of Education funded national information system for K-12 educators (item 20, Table 2.3). Sixteen national ERIC Clearinghouses provide teachers, parents, librarians, and administrators access to full text ERIC Digests and Help Sheets, training and reference materials for Internet use, and links to other ERIC subject-specialized clearinghouses and educational resources such as classroom teaching modules and educational experiments. Queries for information present an embarrassingly parallel processing application well suited for MIMD databases. The current collection of text based information is approximately 15 gigabytes and will soon become accessible over the Internet, and provide full text search capability. Digital text databases will soon evolve into network accessible, multimedia databases.

The Global affairs Institute of the Maxwell School at Syracuse University is currently working with the United Nations Office of the Secretary General to develop an information and decision support system to enhance the Secretary's ability to provide early warning of refugee and displaced person flows (item 20, Table 2.3). The United Nations is developing models to provide an early warning capability to be used in promoting preventive diplomacy in areas which might otherwise require deploying UN forces to resolve conflict. Accurately predicting where political hotspots could turn to conflict is currently based on teams of analysts monitoring daily newspapers in several languages. Newspapers or reports in electronic form are available in English, French, Spanish, and German from major cities of the world, including the New York Times and various private and quasi-public agencies (e.g., Deadline Data) from bureaus around the world. Early warning models require full text on demand political database for policy analysts to use in identifying historical trends, filtering incoming records, and building sophisticated queries to drive stochastic models to indicate sites of potential conflict where policy expertise and diplomatic efforts must be focused.

The combination of virtual reality display technology, and HPCC computer simulations allow one to offer sophisticated views of the real world with important educational value. We are developing a number of simulation on demand projects (item 23, Table 2.3). In "New York State--The Interactive Journey," we will support interactive real-time tours of New York State based on LANDSAT satellite images, and parallel rendering software. We will represent the landscape of New York in three dimensions, and support the links necessary to allow teachers and students to add related databases. Students on the "interactive journey" will be able to navigate through New York State geography, and related databases (e.g., archeological, geological, environmental, business, demographic). These asynchronous, hypermedia links will require MIMD database support. In a second project we will collect a set of "Grand Challenge" science simulations and re-package them as educational software programs running high-performance computers at NPAC, and distribute them to schools over NYNET, a regional high-speed network. Students and teachers will be able to direct these simulations, modify laws of physics, and see the effects in applications such as collisions of "black holes" in space, or the behavior of tornadoes.

We believe that information access and dissemination will become the most important near term industry application of HPCC. Industry shake-ups, and billion dollar scale investments by the telecommunication and software industry in information sources, initially focused on capturing entertainment markets such as movie libraries and video games, serve to illustrate this opportunity.

3.4 Information Integration

Information integration combines information production, analysis, and access in a system of systems. One might view information integration as the dual-use equivalent of command, control, communications, and intelligence for business, manufacturing, and the consumer. By analogy to the aviation term “flying by wire,” where the pilot is not flying a plane but an informational representative of the plane, the business leader or home shopper might “manage by wire” (Haeckel and Nolan, 1993) or use decision support technologies to manage, optimize, and make tradeoffs between informational representatives of complex manufacturing processes or a global consumer market. Information integration represents the largest long term market for HPCC technologies.

Based on a system of systems, information integration naturally overlaps with the other information categories we describe--information production, analysis, and access and dissemination. For example, the credit card data analysis application described above as an information analysis application, fits within a larger information framework for connecting (between corporate database and business units), sharing (between analysts and marketing experts), and structuring information (for specific product offerings).

Business decision support (item 26, Table 2.4) and political decision support (item 27, Table 2.4) are similar in that both require integration of various types of information processing activities, and provide intelligence for sensing "what's going on out there." This knowledge at a system level is the basis of strategies for capturing (or retaining existing) high-margin customers in market segments that appear monolithic to competitors, as well as providing early warning to United Nations preventive diplomacy efforts designed to head off conflict in political hotspots.

4 Information Processing Applications: InfoVision, MADIC, and Electrical Power Networks

We expect information processing applications to become the important opportunity for HPCC in industry. In this section, we describe three application projects that NPAC is currently developing. InfoVision is a set of projects concerning information on demand technologies and applications. InfoVision broadly defines information access and dissemination in our classification of information processing applications. MADIC is a multidisciplinary analysis and design industry consortium for developing pre-competitive HPCC software. MADIC is based on information production (simulation), integrated databases, and decision support, exemplifying information integration. Transient stability analysis in power networks is an NPAC project to optimize the distribution of power in an electrical transmission system. Network simulation is an example of an information production application, with an integration component--the results of the simulation are closely linked with management decisions for deploying power generators in a transmission grid.

4.1 InfoVision

HPCC will play a pervasive role in the information-rich world of the future. HPCC hardware will implement both the networks and the enormous information repositories that will function as servers on the National Information Infrastructure (NII). We term this NII server role as InfoVision, for Information, Video, Imagery and Simulation On demand. Alternately, we can think of InfoVision as high performance multimedia, scalable HPCC algorithms and software used to manage the network packets, and implement new applications running on InfoVision servers.

Dedicated gigabit links between supercomputers will be only one use of the NII. In addition, the NII will support many millions of "personalized" ten megabit/sec streams (this is approximate bandwidth needed to support transport of compressed video) supplying information on demand. InfoVision servers for video on demand will be classical parallel computers---whether clusters of workstations or integrated MPP's. We will have distributed clients with intelligent set top boxes and gigaflop video games in every home, school and business. But we will need production InfoVision servers in large data centers. HPCC techniques will be needed for the necessary hierarchical data storage, browsing, knowledge agents, text and video image content identification, and compression algorithms. (See Table 1.2) The HPCC community needs to examine message passing standards and see that they not only span all homogeneous parallel machines but also the InfoVision metacomputer (100 million clients demanding real-time response from some 10,000 InfoVision superservers).

Currently, InfoVision is an InfoMall project which involves state-of-the-art demonstrations on NYNET, a regional ATM-based gigabit network in New York State, of distributed interactive intelligent information systems. InfoVision is a set of dual-use distributed multimedia projects which prototype near term civilian and military applications of HPCC. These demonstrations are rich enough to explore most of the essential issues for high-performance distributed computing information systems---now called National Challenges---which will be implemented on the NII in the near future. These demonstrations need various key enabling technologies, and InfoVision supports the InfoTech (the technology development and software capitalization component of InfoMall) activities needed to develop, test and evaluate these technologies (Fox et al., (b). 1993).

Video on demand is an immediate market opportunity which we see as ideally suited by InfoVision. To reach this market, we expect a hierarchical information system (Figure 2) will be required and see an opportunity for networks of InfoVision servers. In a video on demand architecture, we might see local servers connected by 1-20 megabit/second cable connections linking neighborhoods of perhaps 1,000 active customers. Regional servers might then support clusters of fiber connected local servers. Local servers (64 processor MPP) will require modest database and simulation capability, and provide perhaps 100 movies (1 gigabyte per 90 minutes of compressed video). Regional servers (1024 processor MPP) will require high-performance database, and modest search capability (movies are much larger "messages" than for example a personalized news source), and provide perhaps 1,000 movies. In addition to regional servers supporting clusters of local servers, specialized servers with high end database (e.g., full text search) and simulation engines will likely be required by businesses, school districts, universities, industry).

As we point out in section (2.2) on the contribution of the NII to HPCC, the ability to reach the end user in the home and office is essential to building the HPCC technology infrastructure for the information processing applications we envision. A one-thousand fold increase in communications performance combined with a similar increase in computing power will put the equivalent of a supercomputer of five years ago in the home. The home computer of the near future will integrate the TV, the settop unit, and many peripherals. This system will be connected to a cable link and supported by an InfoVision server on the NII. Figure 3 illustrates this configuration which combines a data poor but computing rich home and business environment with a data rich but relatively computing poor local InfoVision server.

Several InfoMall partners are needed to form the virtual organizations to both implement the demonstrations with component technologies and system integration as well as to explore the InfoTech activities needed by and opened up by the demonstrations. The partners include Abrams Gentile Entertainment Inc., Columbia University, Cornell University, Digital Equipment Corporation, Eastman Kodak, IBM, Maspar, NYNEX, Oracle, the U.S. Air Force Rome Laboratory, Ultra Inc., and Syracuse University. At Syracuse University, we have involved several organizations including NPAC, Electrical and Computer Engineering, Computer Applications and Software Engineering (CASE) Center, Computer and Information Science, the School of Information Studies, the Newhouse School of Communications, the School of Education, and the Maxwell School of Citizenship and Public Affairs.

As part of our InfoVision program, we are currently developing a focused set of application demonstrations which include simulation (using weather simulations and 3D terrain modeling of New York State), text (using full text databases from Syracuse University's Maxwell School and the AskERIC educational resource in the School of Information Studies), imagery (using Kodak's GIODE PhotoCD technology), and video (using real-time educational news clips from CNN as one possible initial example database) on demand. We are using these demonstrations to develop the necessary component technologies, perform the overall system integration, then test and evaluate these information on demand technologies (Fox et al. (b), 1993).

4.2 MADIC

HPCC capability is essential to integrating manufacturing, maintenance, and design of a highly complex manufacturing product such as an airplane. Coupling multidiscplinary system components, for example, using the flex in an airplane wing in the design of the airplane control system, presents an opportunity to optimize decisions made early in the design process which typically fix the final cost.

The Multidisciplinary Analysis and Design Industrial Consortium (MADIC) project addresses applications of high-performance computing and communications technologies to the analysis and design of large-scale, complex products, such as aircraft and automobiles. Complex systems typically require consideration of multiple engineering and scientific disciplines, including fluid dynamics, structural dynamics, heat transfer, combustion, and manufacturing consideration. HPCC technologies are expected to provide a new generation of design tools that make use of large-scale, high-resolution simulations executing simultaneously in a distributed, heterogeneous computing environment.

MADIC members collaborate on projects of mutual interest, with teams of members formed to address specific projects. Current projects include evaluation of NASA software for use by industry. Current industrial members of MADIC include: Boeing, Ford Motors, General Electric Aircraft Engines, General Electric Corporate Research & Development, General Motors, Grumman, Lockheed, McDonnell-Douglas, Northrup, Rockwell, United Technologies Research Center, and Vought.

Current industrial practice typically requires 12-18 months to develop new designs for large-scale, complex products, such as aircraft and automobiles. Global competitive pressures demand reductions in time-to-market as well as improved design quality for dual-use (commercial and military) applications. This project is intended to reduce significantly the time required to develop new designs of large-scale systems, while simultaneously improving the quality and reducing the cost of new products.

MADIC operations are coordinated through the Center for Research in Parallel Computation (CRPC), which has its headquarters at Rice University and also includes the Northeast Parallel Architectures Center (NPAC) at Syracuse University. MADIC projects are intended to develop, validate and deploy an integrated analysis and design environment operating on a heterogeneous network that includes workstations, vector supercomputers and massively-parallel processors. A proposal for the entire United States Multidisciplinary Analysis and Design Environments (USMADE) was prepared but rejected in the recent Technology Reinvestment Program. USMADE includes modules for geometry definition, grid generation, physical simulation (fluid dynamics, structural dynamics, electromagnetics), process simulation (manufacturing, maintenance), optimization and visualization. Work on selected aspects relevant to the USMADE project is currently underway with funding from other sources. NASA is funding a project to evaluate physical simulation software for use in industrial applications. In this project, industry teams are applying selected NASA codes to problems of current interest, including wings, wing-body interactions, nozzle flows, internal flows (turbines and compressors) and flexible wings. As part of this project, NPAC researchers are evaluating the suitability of the NASA codes to be implemented in High-Performance Fortran. CRPC researchers at Rice University are evaluating the use of the automatic differentiation system ADIFOR.

4.3 Transient Stability Analysis for Power Systems

One major NPAC project involves the utility industry, where we are working with Niagara Mohawk Power Corporation on simulations of their electrical transmission system. This can be done in real time using modern parallel machines with around 64 nodes. We are creating prototype software that Niagara Mohawk will evaluate. Future steps involve the utility industry research consortium Electric Power Research Institute, other utilities and the companies that currently build the sequential power grid software now used by electrical utilities. This application is straightforward to parallelize as it involves forming a matrix (with embarrassingly parallel matrix element generation) followed by a sparse matrix solve (IEEE, 1992) (Koester, Ranka, and Fox 1993). The latter uses a direct method on a roughly 10000 X 10000 matrix. Developing an efficient parallel sparse matrix solver is the most difficult problem for this application. Utilities must balance the cost of producing power against the cost of failures. Thus an accurate simulation of transient stability analysis---the response of the network to a disturbance---allows better cost---security tradeoffs and hence cheaper power. Power is very expensive and even modest savings from real-time simulation would warrant purchase of HPCC supercomputers by the utilities.

5 Summary and Conclusions

The National Information Infrastructure of the future will be built upon core HPCC technologies such as parallel architectures, software, and high-speed networks. In addition to the key enabling technologies associated with high-performance scientific computing, such as image processing algorithms and linear algebra libraries, information processing applications will require development and integration of new, enabling HPCC technologies such as collaboration services, and multimedia databases. These new technologies will find application in information industries such as home entertainment and agile manufacturing.

While simulation will remain an important HPCC application, the real opportunity for developing significant industrial applications of HPCC technologies lies in information processing. Information industries such as home shopping, electronic banking, distance education, and televirtuality are growing at a much faster rate than the product industries such as aerospace, or other manufactured goods. The current focus on scientific simulation in the national HPCC community must change in response to this trend.

Based on our survey of industrial applications of HPCC compiled over the past three years, we described four classes of information processing---production, analysis, access and distribution, and integration. We concluded that information analysis (e.g., data mining credit card holders' purchase patterns) is the most important near term opportunity, and information integration (e.g., decision support in the global economic war) is the most important long term opportunity for HPCC applications in industry.

This scenario provides great opportunities for new software entrepreneurs, a repeat of the opportunities opened up by the PC. We have set up a new networked virtual organization InfoMall to further this vision.

List of Figures and Tables

Figures

Tables

6 References

CAPPS (1993) Commercial Applications in Parallel Processing (CAPPS). Microelectronics and Computer Technology Corporation. Austin, Texas. October 19-20.

Choudhary, A., Fox, G., Hiranandani, S., Kennedy, K., Koelbel, C., Ranka, S., and Saltz, J. (1991) A Classification of Irregular Loosely Synchronous Problems and Their Support in Scalable Parallel Software Systems, in *Proceedings from Darpa Software Technology Conference*, 1992, pp. 138--149. Syracuse University Technical Report SCCS-255, June, 1991.

Fox, G. (1992) 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).

Fox, G., Bogucz, E., Jones, D., Mills, K., and Podgorny, M. (1993 (a)) 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).

Fox, G, Hariri, S., Chen, R., Mills, K., and Podgorny, M. (1993 (b)) InfoVision--Information, Video, Imagery, and Simulation On Demand. SCCS--575. (Available from NPAC, Syracuse University, Syracuse NY 13210-4100).

Fox., G., Messina, P., and Williams, R. (1994) *Parallel Computing Works!*. Morgan Kaufmann, San Mateo, CA.

Grand Challenges (1993) High Performance Computing and Communications, The FY 1993 US Research and Development Program, Walter Massey, Chair. A Report by the Committee on Physical, Mathematical, and Engineering Sciences, Federal Coordinating Council for Science, Engineering, and Technology Committee, Office of Science and Technology Policy. (68 p.)

High Performance Fortran Forum (1993). High performance Fortran language specification, High Performance Fortran Forum, May 3, 1993, Version 1.0. 184 pp. HPFF92a. Copyright Rice University, Houston, Texas. (Available by anonymous ftp at minerva/npac.syr.edu HPFF/hpf-v10-final.ps)

Haeckel, S., and Nolan, R. (1993) The Role of Technology in an Information Age. in *The Knowledge Economy*. The Nature of Information in the 21st Century. Institute for Information Studies. 1993-1994.

IEEE Committee Report: Parallel Processing in Power Systems Computations (1992) IEEE Transactions on Power Systems, **7**(2), 629.

Ives, B. and Jarvenpaa, S. (1993) Competing With Information: Empowering Knowledge Networks with Information Technology. in *The Knowledge Economy*. The Nature of Information in the 21st Century. Institute for Information Studies. 1993-1994.

Ives, B., and McKeown, M. (1993) The Promise of a New World Information Order. in *The Knowledge Economy*. The Nature of Information in the 21st Century. Institute for Information Studies. 1993-1994.

Johnson, N. (1993) The Nature of Information. in *The Knowledge Economy*. The Nature of Information in the 21st Century. Institute for Information Studies. 1993-1994.

Koester, D., Ranka, S., and Fox, G. (1993) Parallel Block-Diagonal-Bordered Sparse Linear Solvers for Electrical Power System Applications, *Proceedings of the Scalable Parallel Libraries Conference*, Mississippi State University, October 1993. NPAC technical report SCCS-552.

Mills, K., and Fox, G. (1992) HPCC Applications Development and Technology Transfer to Industry. *Proceedings of the New Frontiers: A Workshop on Future Directions of Massivley Parallel Processing*. IEEE Computer Society and National Aeronautics and Space Administration. IEEE Computer Society Press. p. 58-65.

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

National Information Infrastructure Testbed Panel (NIIT) (1993) Portland, Oregon. November 15-19, 1993. *Supercomputing '93*. IEEE Computer Society Press, Los Alamitos, CA.

High Performance Computing and Communications: Toward a National Information Infrastructure. (1994) A report by the Committee on Physical, Mathematical, and Engineering Sciences; Federal Coordination Council for Science, Engineering and Technology; Office of Science and Technology Policy.

Pinelli, T., Barclay, R., Bishop, A., and Kennedy, J. (1992) Information Technology and Aerospace Knowledge Diffusion. *electronic networking: research, applications and policy*. **2**:2 31-50.

RCI North American Annual Member Executive Conferences on Technology Transfer. (1993) RCI, Ltd. Worldwide Headquarters, 1301 East 79th Street, Suite 200, Minneapolis, MN 55425.

Workshop on Grand Challenge Applications and Software Technology. (1993) Industry Panel Workshop. May 4-6, 1993. Hyatt Regency Hotel, Pittsburgh, PA.

Table 1.1 Core Enabling HPCC Technologies for Information Production (Simulation)

- PVM, Express, Linda, MPI
- ISIS (Cornell)
- High Performance Fortan (HPF) Compiler
- High Performance C, C++ Compiler
- HPF Extensions - PARTI
- Parallel / Distributed Computing Runtime Tools
- ADIFOR (Differentiate Fortran Code)
- AVS and Extensions
- High Performance Fortran Interpreter
- Image Processing
- Parallel Debugger
- Parallel Performance Visualization
- Parallel Operating Systems
  - I/O
  - Scheduling
- Virtual Reality
- Event Driven Simulator
- Mesh Generation
- SCALAPACK
- Sparse Matrix Solvers Templates and Libraries
- Particle Dynamics Kernels Templates and Libraries
- Optimization Methodology and Templates
  - linear programming
  - non-linear programming
- Scheduling (neural-net, parallel) Templates

Table 1.2 Core Enabling HPCC Technologies Information Analysis, Access, Integration

- Parallel (Relational) Database e.g. Oracle 7.0
- Object database
- High Speed Networks
- Multilevel Mass Storage
- Integration Software ("glue")
- Integration of Parallel and Distributed Computing
- Multimedia Support
    - Video Browsing
    - Image Content
    - Full Text Search
    - Real time I/O (disk ---> network)
- ATM Network Protocols and Management
- Compression
- Parallel Rendering
- Linkage Analysis (between records of database)
- Sorting (large databases)
- Collaboration Services
    - Multi user video conferencing
    - Electronic whiteboards, etc.
- Security and Privacy
- Usage and Charging Algorithms
- Televirtuality
    - The world as a metacomputer
    - Naming
- Human-Computer Interfaces
    - Mosaic
- Image Processing
    - Terrain Rendering
    - Kodak Photo-CD
- Geographical Information Systems
    - Spatial databases

Table 2.1 HPCC Industrial Applications: Simulation or Information Production

Item	Application Area and Examples	Problem Comments	Problem Architecture	Machine and Software
1	Computational Fluid Dynamics	PDE, FEM Turbulence Mesh Generation	Loosely Synchronous Adaptive Mesh is Asynchronous	SIMD, MIMD for irregular, adaptive HPF(+) Unclear for adaptive irregular mesh
2	Structural Dynamics	PDE, FEM Dominated by Vendor Codes (e.g. NASTRAN)	Loosely Synchronous mesh as in 1	MIMD as Complex geometry HPF(+)
3	•Electromagnetic Simulation •Antenna Design •Stealth Vehicles •Noise in high frequency circuits •Mobile Phones	PDE Moment method (matrix inversion) dominates ______________ Later FEM, FD?	Synchronous ______________ Loosely Synchronous	SIMD HPF ______________ SIMD, MIMD, HPF(+)
4	<u>Scheduling</u> •Manufacturing •Transportation (Dairy delivery to military deployment) •University Classes •Airline Scheduling of crew, planes in static or dynamic (midwest snowstorm) cases	Expert systems and/or Neural Networks, Simulated Annealing Linear Programming (hard sparse matrix)	Asynchronous Synchronous Loosely Synchronous	MIMD (unclear speedup) Asyncsoft SIMD HPF MIMD
5	Environmental Modeling - Earth/Ocean/ Atmospheric Simulation	PDE, FD, FEM Sensitivity to data	Loosely Synchronous	SIMD, MIMD for irregular, adaptive HPF(+) Unclear for adaptive irregular mesh

PDE	Partial Differential Equation
FEM	Finite Element Method
FD	Finite Difference
ED	Event Driven Simulation
TS	Time Stepped Simulation
VR	Virtual Reality
HPF	High Performance Fortran (High Performance Fortran Forum, 1993)
HPF+	Natural Extensions of HPF (Choudhary et al., 1991)
MPF	Fortran plus message passing for loosely synchronous software
Asyncsoft	Parallel Software System for (particular) class of asynchronous problems
CFD	Computational Fluid Dynamics

Note on Language: HPF, MPF are illustrative for Fortran: one can use parallel C, C++ or any similar extensions of data parallel or message passing languages

Table 2.1 HPCC Industrial Applications: Simulation or Information Production (continued)

Item	Application Area and Examples	Problem Comments	Problem Architecture	Machine and Software
6	Environmental Modeling - Complex systems e.g. lead concentration in blood	Empirical models Monte Carlo and Histograms	Embarrassingly Parallel plus global reductions	Some SIMD MIMD more natural
7	Basic Chemistry •Chemical Potentials •Elemental Reaction Dynamics	Calculate Matrix elements Matrix eigenvalue Multiplication, inversion	Embarrassingly Parallel Synchronous	MIMD (maybe SIMD) HPF
8	Molecular Dynamics	Particle Dynamics with irregular cutoff forces Fast Multipole methods Mix of PDE and Particles in PIC	Loosely Synchronous	HPF(+) or MPF for fast multipole
9	Economic Modeling •Real Time Optimization •Mortgage backed securities •Option Pricing	Individual (Monte Carlo) Full simulations of portfolios	Synchronous, Embarrassingly parallel Metaproblems	SIMD, HPF MIMD, SIMD Integration software
10	Network Simulations	Sparse matrices; Zero structure defined by network connectivity	Loosely Synchronous	MIMD MPF

PDE Partial Differential Equation
FEM Finite Element Method
FD Finite Difference
ED Event Driven Simulation
TS Time Stepped Simulation
VR Virtual Reality
HPF High Performance Fortran (High Performance Fortran Forum, 1993)
HPF+ Natural Extensions of HPF (Choudhary et al., 1991)
MPF Fortran plus message passing for loosely synchronous software
Asyncsoft Parallel Software System for (particular) class of asynchronous problems
CFD Computational Fluid Dynamics

Note on Language: HPF, MPF are illustrative for Fortran: one can use parallel C, C++ or any similar extensions of data parallel or message passing languages

Table 2.1 HPCC Industrial Applications: Simulation or Information Production (continued)

Item	Application Area and Examples	Problem Comments	Problem Architecture	Machine and Software
11	Particle Transport Problems	Monte Carlo methods as in neutron transport for explosion Simulations	Embarrassingly Parallel (Asynchronous)	MIMD HPF
12	Graphics (rendering) Hollywood Virtual Reality	Several operational Parallel Ray Tracers Distributed model hard	Embarrassingly Parallel (Asynchronous)	MIMD HPF Asyncsoft for distributed database
13	Integrated Complex Systems Simulations •Defense (SIMNET, Flight Simulators) •Education (SIMCITY) •Multimedia/VR in Entertainment •Multiuser virtual worlds •Chemical & Nuclear Plants	Event driven (ED) and Time Stepped (TS) Simulations. Virtual Reality Interfaces. Database backends. Interactive	Metaproblem Fully asynchronous if ED Loosely Synchronous for TS	Timewarp or other ED Asyncsoft HPF(+) Integration Software Database

PDE	Partial Differential Equation
FEM	Finite Element Method
FD	Finite Difference
ED	Event Driven Simulation
TS	Time Stepped Simulation
VR	Virtual Reality
HPF	High Performance Fortran (High Performance Fortran Forum, 1993)
HPF+	Natural Extensions of HPF (Choudhary et al., 1991)
MPF	Fortran plus message passing for loosely synchronous software
Asyncsoft	Parallel Software System for (particular) class of asynchronous problems
CFD	Computational Fluid Dynamics

Note on Language: HPF, MPF are illustrative for Fortran: one can use parallel C, C++ or any similar extensions of data parallel or message passing languages

Table 2.2 HPCC Industrial Applications: Information Analysis - "Data Mining"

Item	Application Area and Examples	Problem Comments	Problem Architecture	Machine and Software
14	Seismic and Environmental data analysis	No oil in NY State. Parallel Computer already important	Embarrassingly parallel as in many "event" analysis problems (high energy physics, astronomy, etc.)	SIMD, maybe MIMD needed HPF
15	•Image Processing •Medical Instruments •EOS (Mission to Planet Earth) •Defense Surveillance •Computer Vision	Commercial Applications of Defense Technology. Component of many Information Integration Applications e.g. Computer Vision in Robotics	Metaproblem Synchronous (low level) Loosely Synchronous (medium level) Asynchronous (expert system)	Metacomputer SIMD (low level) MIMD (medium/high level) HPF(+) Software Integration Asyncsoft Database
16	•Health Fraud Inefficiency •Securities Fraud •Credit Card Fraud	Linkage Analysis of database records for correlations	Synchronous if records "identical" otherwise Loosely Synchronous	SIMD MIMD
17	Market Segmentation	Sort and Classify records to determine customer preference by region (city --> house)	Loosely Synchronous	Aspects could be SIMD MIMD better for sorting?

PDE	Partial Differential Equation
FEM	Finite Element Method
FD	Finite Difference
ED	Event Driven Simulation
TS	Time Stepped Simulation
VR	Virtual Reality
HPF	High Performance Fortran (High Performance Fortran Forum, 1993)
HPF+	Natural Extensions of HPF (Choudhary et al., 1991)
MPF	Fortran plus message passing for loosely synchronous software
Asyncsoft	Parallel Software System for (particular) class of asynchronous problems
CFD	Computational Fluid Dynamics

Note on Language: HPF, MPF are illustrative for Fortran: one can use parallel C, C++ or any similar extensions of data parallel or message passing languages

Table 2.3 HPCC Industrial Applications: Information Access and Dissemination: InfoVision - Information, Video, Imagery and Simulation on Demand

Item	Application Area and Examples	Problem Comments	Problem Architecture	Machine and Software
18	Transaction Processing •ATM (automatic teller machine)	Database-most transactions short. As add "value" this becomes Information Integration	Embarrassingly Parallel	MIMD Database
19	Collaboratory •Telemedicine	Research Center or doctor(s) - patient interaction without regard to physical location	Asynchronous	High Speed Network
20	Text on Demand •Digital (existing) libraries •ERIC Education database, •United Nations - Worldwide newspapers	Multimedia database Full text search	Embarrassingly Parallel	MIMD Database
21	Video on Demand •Movies, News (CNN Newsource & Newsroom), •Current cable, •United Nations - Policy Support	Multimedia Database Interactive VCR, Video Browsing, Link of video and text database	Embarrassingly Parallel	MIMD Database Compression (SIMD) Video Editing Software
22	Imagery on Demand Kodak GIODE •"clip art" on demand •Medical images •Satelite images	Multimedia database Image Understanding for Content searching and (terrain) medical feature identification	Metaproblem Embarrassingly Parallel plus Loosely Synchronous Image Understanding	MIMD but much SIMD image analysis
23	Simulation on Demand •Education, Tourism, City planning, •Defense mission planning	Multimedia map database Generalized flight simulator Geographical Information System	Synchronous terrain rendering with Asynchronous Hypermedia	SIMD terrain engine (parallel rendering) MIMD database

Table 2.4 HPCC Industrial Applications: Information Integration

	These involve combinations of Information Production, Analysis and Access thus need software and machine architecture issues given for these "subproblems" • Systems of Systems
24	Command and Control • Battle Management, Command, Control, Communication, Intelligence and Surveillance (BMC^3IS) • Military Decision Support
25	SIMNET - Military Simulation with computers and people in the loop
26	Business decision support
27	Political decision support United Nationals uses video and multilingual newspaper (Maxwell School at Syracuse University)
28	Robotics
29	Electronic banking
30	Electronic shopping
31	Agile Manufacturing - Multidisciplinary Design - Concurrent Engineering •MADIC Industrial Consortium

Figure 1. InfoMall: A Scalable Organization for Development of HPCC Software and Systems

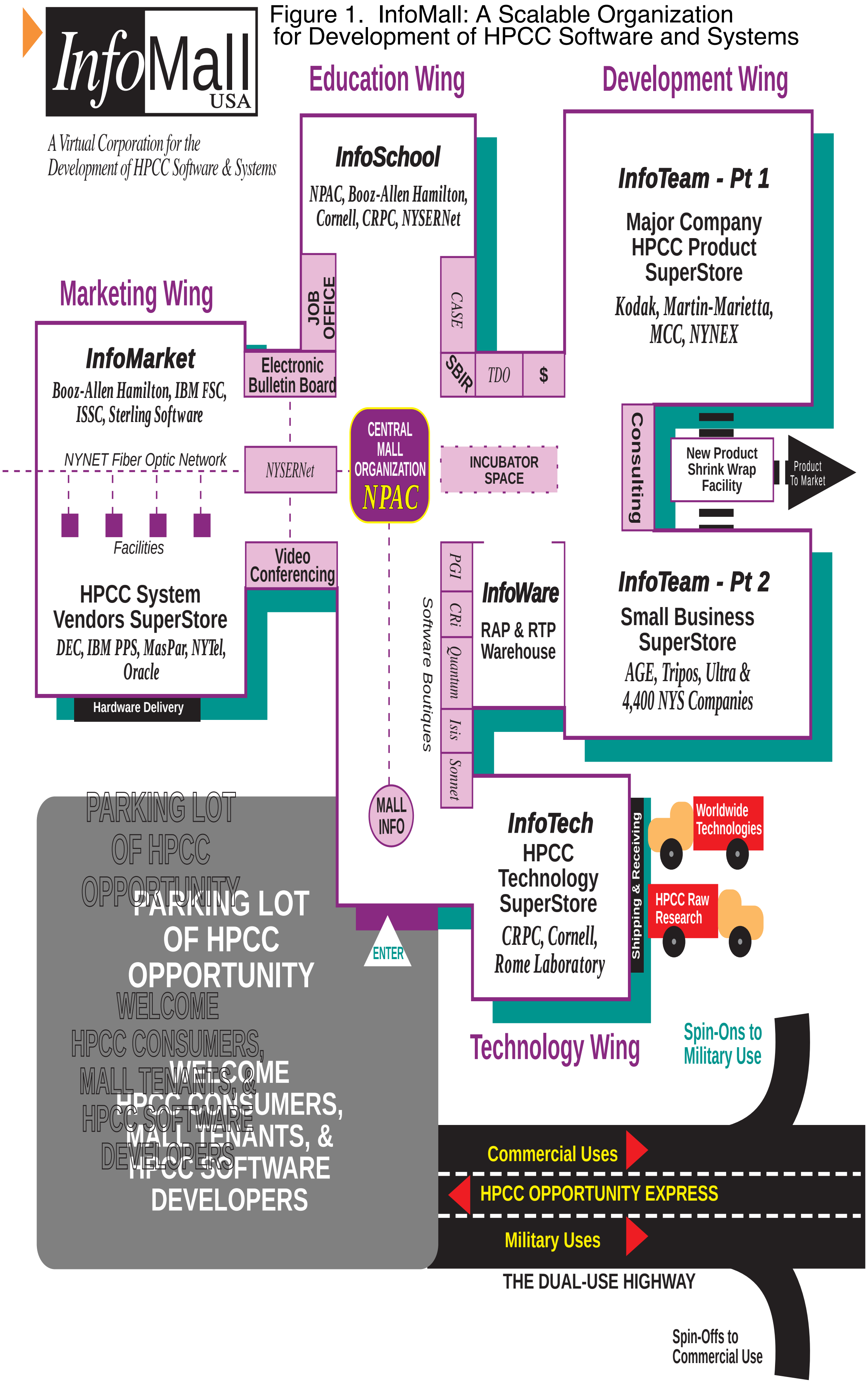

Figure 2. A Possible Hierarchical Information System for Video on Demand

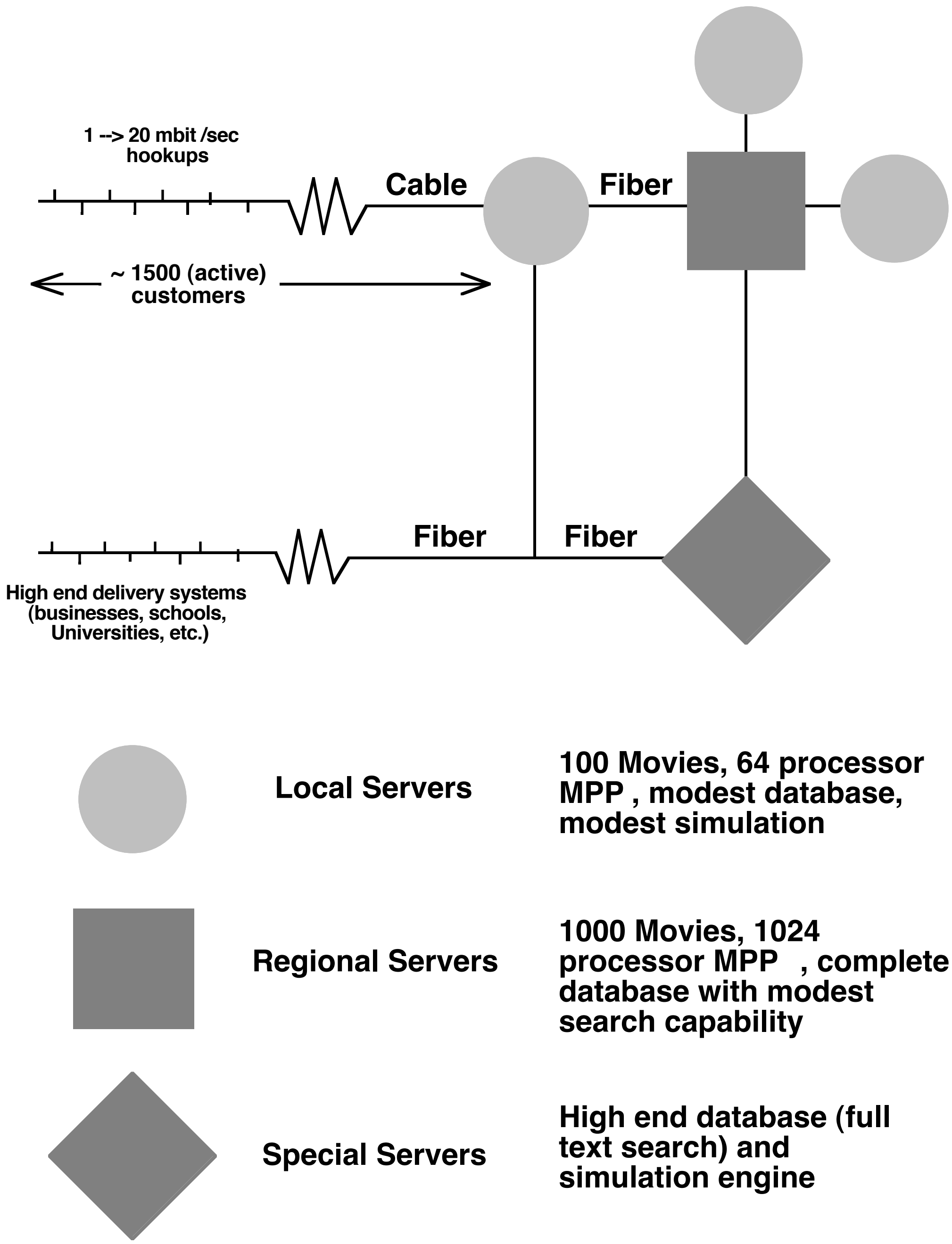

Figure 3. InfoVision Delivery Over the NII