Involvement of Industry in the National High Performance Computing and Communication Enterprise

Geoffrey C. Fox
Syracuse University
Northeast Parallel Architectures Center
111 College Place
Syracuse, New York 13244
gcf@npac.syr.edu

Abstract

We discuss aspects of a national computer science agenda for High Performance Computing and Communications (HPCC). We agree with the general direction and emphasis of the current program. In particular, the strong experimental component and linkage of applications with computer science should be continued. We recommend accelerating the emphasis on "national challenges" with more applications and technologies from the information, as compared to simulation areas. We suggest modifying the grand challenge concept to complement the current teaming of particular computer science and applications researchers. We would emphasize better linking of each application group to the entire (inter) national computer science activity. We express this in terms of a virtual corporation metaphor. The same approach can be used to involve industry in HPCC for both the consumers of HPCC technology (application industries) and producers—Independent Software Vendors (ISV) and the hardware system companies. We illustrate this approach with InfoMall, a HPCC technology transfer program funded by New York State. The federal program should have greater incentives for the involvement of both ISV's and their products.

1 Introduction

This commentary on the National High Performance Computing and Communications Initiative, and the associated National Information Infrastructure does not set forth a complete new program. This is not necessary—the current federal program is "almost right on." Rather we present a set of remarks based on the involvement of applications and industry in the program. We will consider both the industries that use HPCC as well as those that (will) produce the new HPCC software and system products that will drive the integration of HPCC into the mainline computer industry.

2 Opportunities for HPCC in Industry

While at Caltech, one of my major activities involved studying the parallel software and algorithms needed by large scale scientific and engineering computation [Angus:90a], [Fox:88a], [Fox:94a]. One of my interests in moving to Syracuse was to extend this approach, but focus on industrial and not academic applications. This activity, now called InfoMall [Fox:93c], [Mills:93a], [Mills:94a] is supported by New York State and aims to accelerate the introduction of High Performance Computing and Communication technology into the State's industry. It has three major components. Firstly, it involves those industries that uses computers (i.e., the applications of HPCC, which are described later in this section). Secondly, InfoMall seeks to create (enhance) an HPCC software industry in the State, and Section 3 describes issues connected with this activity. The final component of InfoMall is devoted to training and is especially centered in the mid-Hudson (Kingston, Poughkeepsie, and Fishkill) area of New York where some 10,000 former IBM engineers are looking for new job opportunities—InfoMall aids those interested in HPCC careers. Such pools of engineers exist in many parts of the country, a byproduct of the major restructuring in aerospace, manufacturing, computer, and other industries. This national resource indicates that we have the people to implement the integration of HPCC into the country's industry. We "only" need to identify the viable opportunities and the necessary funds (venture capital) to implement them.

Now, we discuss which opportunities are in fact viable—this returns to the first component of InfoMall—what are the relatively near term industrial applications of HPCC. The first phase of our New York State project involved a significant survey reported in Tables 1 to 6, which divides possible applications into four major areas, and then in more refined fashion, 33 classes. The good news is that as summarized in Table 2 and Column 2 of Tables 1 to 4, we already have the necessary, if not the most productive, software technology needed to implement all of these applications. Further, extensive computer science research tested in grand challenge and other academic and research implementations, has given us excellent parallel algorithms. Again, a good argument can be made in every entry in Tables 1 to 5, that HPCC will make a difference and improve the performance and/or competitiveness of the associated industry. Nevertheless, only in a few cases will the return on investment justify the work involved in implementing the HPCC application. This conclusion is explained in [Fox:92e] [Mills:94b], [Fox:94b], [Fox:94c], and here we can only summarize and illustrate this conclusions. Note that the four application classes in Table 1 are:

(A) Table 1,2: Simulation or Information Production

(B) Table 3: Data or Information Analysis

(C) Table 4: Access to and Dissemination of Information

(D) Table 5: Integrated Information Services including Components of the Earlier Categories (A), (B), and (C).

Originally, I had envisioned a major InfoMall activity in the Simulation (A), as this was the natural counterpart to my personal academic activity at Caltech, as well the broad national endeavor in the Federal Grand Challenge Initiative. However, the clear near-term value of HPCC simulations in the research applications is not so obvious in industry. We can only illustrate this conclusion with one example here. Consider the application of HPCC to manufacturing—an area of exceptional importance and promise. One can use HPCC for one or more of the components of manufacturing—such as the first three classes of Table 1—fluid flow, structural and electromagnetic simulation. However, these are in general quite small parts of the whole manufacturing and design process. Thus, we see Amdahl's law applied to manufacturing—if simulation is (for example) 10% of the design cycle, then parallelizing this one part can give at best a speedup of 1.1 We are still dominated by the 90% remaining "sequential" part of the process to which we have not applied HPCC. Thus, the best opportunity for HPCC lies in its integration into the

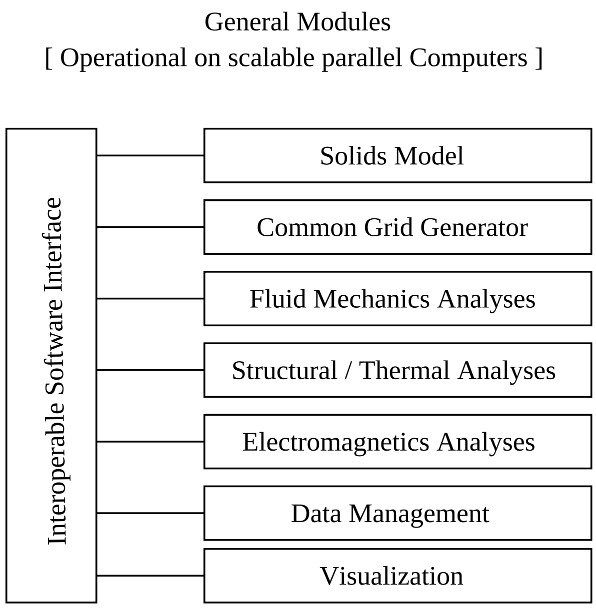

Figure 1: Software Structure for Multidisciplinary Analysis and Design

entire manufacturing operations. This will enable agile manufacturing and as components of this, concurrent engineering and multidisciplinary analysis and design. Just considering the latter, this is as shown in Figure 1 a challenging but in principle, perfectly feasible integration task. However, the real world makes it "very hard". Thus, components of Figure 1 are existing one million codes (such as NASTRAN) which not only have to be parallelized, but also integrated with other major software packages. Perhaps more important difficulties are nontechnical issues. Most manufacturing companies face restructuring, cutbacks and/or fierce global competition. Agile manufacturing implies not only technology investment and (risky) development, but remaking the whole engineering process. Thus, we cannot address manufacturing just as an HPCC technology problem, but many national infrastructure and education issues are involved. I see this area as an exciting long-term area for HPCC, which need major government investment as industry will find it hard to justify the investment based on a typical one to three year return.

Returning to the survey of Tables 1 to 6, we found (in New York State) much more near-term promise in the information related applications (B), (C), and (D). There are several reasons for this. The simplest reason is phase space—information processing always has been the dominant use of computers in business and this is likely to be accentuated and not decreased in the future. Further, one "only" needs to parallelize a few software pieces (e.g., implement a parallel database) to enable a large set of applications. Although very challenging, this task is well underway commercially with, for instance, Oracle 7.1 supporting parallel query. In contrast, simulation has no single "killer application" and each piece of software is typically only used in specialized fashion. The National Information Infrastructure (NII) applications in (C), parallel servers on high-speed networks, have another advantageous feature—namely they are "new", that is, involve developing software from scratch and not as in many simulation applications, porting existing sequential codes. It is clearly easier to justify the investment in parallel machines for new software systems than for the cases when porting of large existing systems is involved.

This is a rambling story and incomplete to boot—what do we deduce from it? I concluded that information processing is the most promising role for the HPCC industry. Correspondingly, I have redirected our New York State funded activities in this area. Our project must focus on relatively near term opportunities to satisfy its (funded) economic development goals of job creation. However, the federal government needs to support both simulation and information processing related technologies—HPCC simulation will be critical in the long run and not an easy investment for industry. Thus, I support the broadening of the Grand Challenges to include a new set of National Challenges focussed on information based applications. This should not only feature NII applications (C) (and parts of (D)), but also the basic information analysis cases in Table 3. In fact, as the NII does not yet exist except as the comparatively low bandwidth internet, industrial interest will initially center on those cases in Tables 3 and 5, which do not require the pervasive multimegabit/second network for everybody, everywhere and at every time promised by the NII.

3 Technology Transfer and the Role of Grand Challenge Teams

The approximately 50 Grand Challenge teams have been correctly set up as collaborations between computer scientists and application types. This can help the particular computer scientists if the application uses, evaluates and motivates further development of their personal technology. I have seen this is in many grand challenges and have concerns because this is important, but only half the story. Thus, we also need to understand and develop the best computer technologies in the world for this application. From this point of view, we have the application and not computer science driving the technology. This identifies another role for the computer science members of the grand challenge team. They should act as a window (educational/knowledge source) into the best HPCC technology that one can bring to bear on the application. One needs applications to act as testbeds for new HPCC technologies. However, in the grand challenges, I suggest that an application pull is most appropriate, and that in general other (funding) mechanisms be used for the application testbeds (computer science pull) which are also definitely needed, especially at the initial stages of technology development.

We have embodied this concept into the model for technology transfer in InfoMall, which is illustrated from two points of view in Figures 2 and 3. We have a world wide customer (application) pull (on right side of figures) that feeds from a world wide pool of HPCC technologies (left side of figures). These are fed through the "window" discussed above into a software capitalization (InfoTech) process that produces usable prototype systems, which can be commercialized at lower risk than raw technologies. Key to this InfoMall engine is the entrepreneur or ISV (Independent Software Vendor) who takes InfoTech products, and with traditional economic development support, produces the HPCC products brought to the customer by the marketing and system integration components of InfoMall. InfoMall offers to ISV's the special high technology support (Table 8) needed for an entrepreneur in HPCC. This includes both access to facilities and educational resources, as mentioned earlier in context of our mid-Hudson project.

We believe that major efforts should be made to increase the role of all parts of industry in the HPCC program. This should include the users with realistic estimates of the time scale on which HPCC will be useful technology—as illustrated in Section 2, this will certainly vary from case to case. However, the ISV's are also critical for there can be no HPCC hardware or systems business, and no significant user base, without ISV's linking them. We see at least three ways how the HPCC program could encourage the HPCC software industry.

1. Encourage federal proposals that use commercial rather than home grown software systems. I am reminded of a story of one computer vendor who complained that computer scientists could not buy their system because the commercial systems software was too good—one could only get federal funding to build software for systems where the existing commercial offerings were poor. Instead, we should encourage research that builds on top of existing commercial software. We should encourage the latter to offer open interfaces enabling this strategy.

2. Change the funding of MPP to explicitly encourage purchase of software, as well as hardware—recently, I was told that HPCC systems software companies were nonviable. One couldn't sell HPCC software

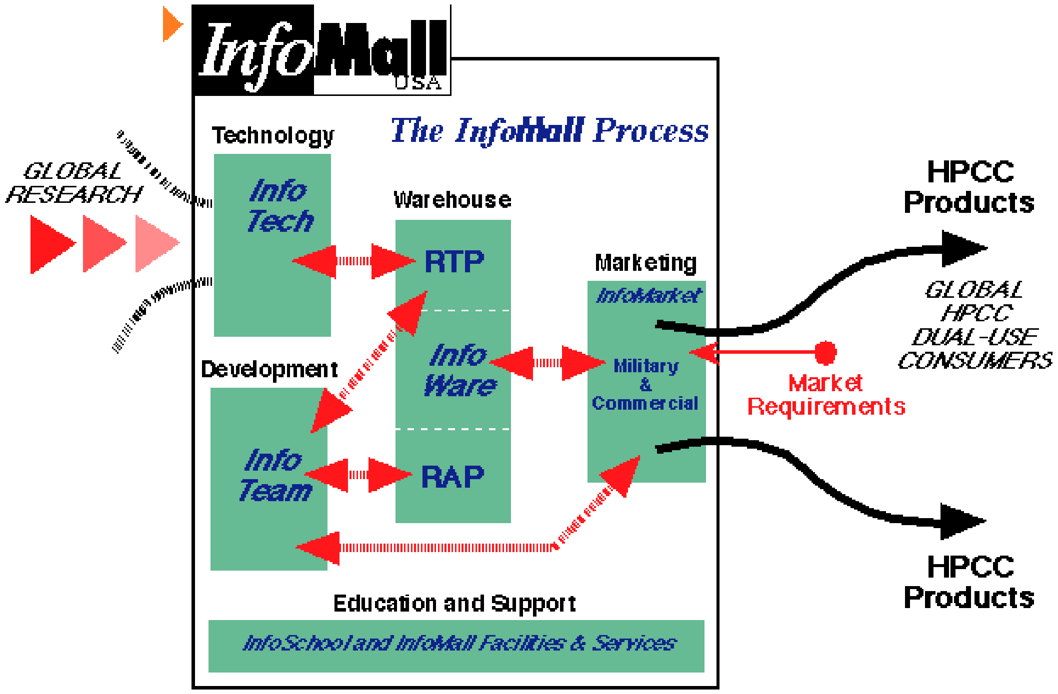

Figure 2: InfoMall Process for Technology Development. (RAP) Reusable Application Pieces (RTP) Reusable Technology Pieces

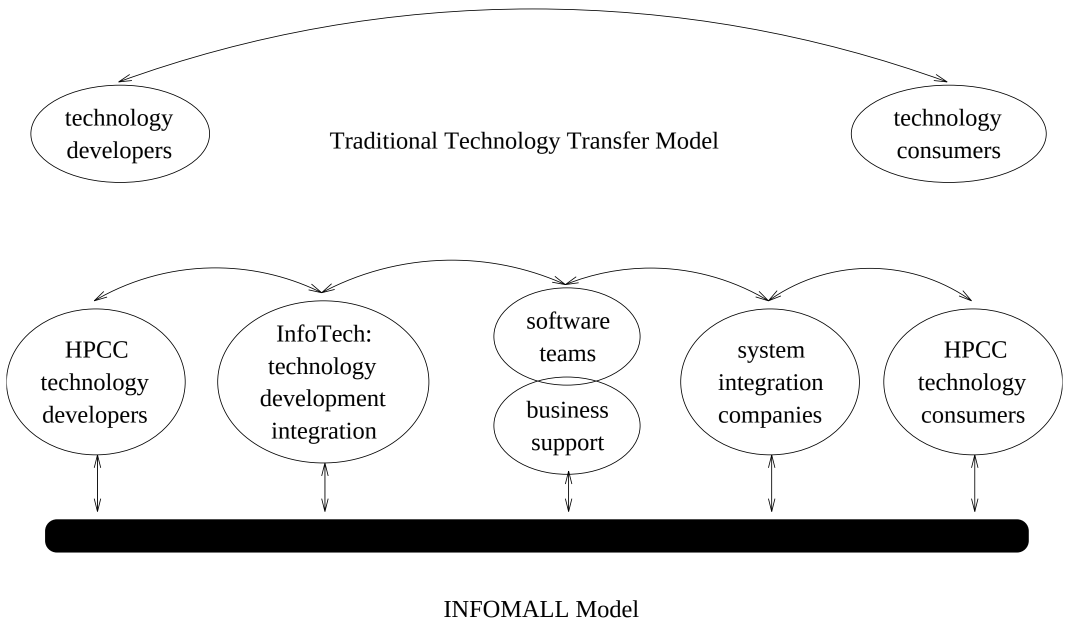

Figure 3: InfoMall bridges the gulf between technology developers and consumers in a number of steps. Traditional technology transfer models make one big leap.

Table 1: **Simulation Opportunities for HPCC in Industry**

Application Areas and Comments	**Machine and Software**
(1) **Computational Fluid Dynamics** PDE, FEM Turbulence Mesh Generation	SIMD MIMD for irregular, adaptive **HPF** and **ADAPTIVE**
(2) **Structural Dynamics** PDE, FEM Dominated by Vendor Codes	MIMD as Complex geometry **HPF** and **ADAPTIVE**
(3) **Electromagnetic Simulation** PDE Moment method (matrix inversion dominates)	SIMD **HPF**
Later FEM, FD? Fast Multipole	SIMD, MIMD, **HPF** and **ADAPTIVE**
(4) **Scheduling** Expert systems and/or	MIMD (unclear speedup) **Asynchronous**
Neural Networks, Simulated Annealing	SIMD **HPF**
Linear Programming (hard sparse matrix)	MIMD ADAPTIVE
(5) **Environmental Modeling** (Ocean, Earth, Atmosphere model) PDE, FD, FEM Sensitivity to data	SIMD MIMD for irregular, adaptive mesh **HPF** and **ADAPTIVE**
(6) **Environmental Modeling** (complex systems) Empirical models Monte Carlo and Histograms	Some SIMD MIMD more natural **HPF**
(7) **Basic Chemistry** Calculate Matrix elements Matrix Eigenvalue Multiplication, Inversion	MIMD (maybe SIMD) **HPF**

Table 2: **Simulation Opportunities for HPCC in Industry (contd.)**

Application Areas and Comments	**Machine and Software**
(8) **Molecular Dynamics** Particle Dynamics with irregular cutoff forces Fast Multipole methods Mix of PDE and Particles in PIC or DSMC	**HPF** and **ADAPTIVE**
(9) **Economic Modeling** Individual (Monte Carlo) Full simulations of portfolios	SIMD, **HPF** MIMD, SIMD **Integration**
(10) **Network Simulations** Sparse matrices; Zero structure defined by network connectivity	MIMD, **HPF** for matrix elements. Library matrix solver is **ADAPTIVE**
(11) **Particle Transport Problems** Monte Carlo methods as in neutron transport for explosion Simulations	MIMD **HPF**
(12) **Graphics** (rendering) Several operation Parallel Ray Tracers Distributed model hard	MIMD **ADAPTIVE**
(13) **Integrated Complex Systems Simulations** Event driven (ED) and Time Stepped (TS) Simulations Virtual Reality Interfaces Database backends Interactive	Timewarp or other ED **Asynchronous** **Integration** Software Database **HPF** and **ADAPTIVE** for TS simulation

Table 3: **Information Analysis—"Data Mining"—Opportunities for HPCC in Industry**

Application Areas Comments	Machine and Software
(14) **Seismic and Environmental** No oil in NY State Parallel Computer already important	SIMD, maybe MIMD needed **HPF**
(15) **Image Processing** • Commercial Applications of Defense Technology • Component of many Information Integration Applications, e.g., Computer Vision in Robotics	Metacomputer for full problem Low Level SIMD, **HPF** Medium/High level MIMD **ADAPTIVE** **Integration** Software **Asynchronous** Database
(16) **Statistical Analysis** Packages (libraries) Optimization Histograms (see category 4)	**HPF** adequate for many libraries
(17) **Healthcare Fraud** (also credit card, securities) Linkage analysis database records for correlations	SIMD or MIMD Database Access (Parallel SQL plus category 16)
(18) **Market Segmentation** Sort and Classify records to determine customer preference by region (city → house)	**ADAPTIVE** Some cases are SIMD Parallel SQL plus category 16

Table 4: **Information Access: InfoVision—Information, Video, Imagery and Simulation on Demand—High Performance Multimedia Computing and Communications**

(19) Transaction Processing Database-most transactions short. As add "value" this becomes information Integration	MIMD Database
(20) Collaboratory Research Center or doctor(s)—patient interaction without regard to physical location	High Speed Network
(21) Text on Demand Multimedia database Full text search	MIMD Database
(22) Video on Demand Multimedia Database Interactive VCR, Video Browsing, Link of video and text database	MIMD Database Compression (SIMD) Video Editing Software
(23) Imagery on Demand Multimedia database Image Understanding for content searching and feature identification	MIMD but much SIMD image analysis
(24) Simulation on Demand Multimedia map flight simulator Geographical Information System	SIMD terrain engine (parallel rendering) MIMD database **Integration** software

Table 5: **Information Integration Opportunities for HPCC**

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

Table 6: **Abbreviations used in Tables 1–4**

PDE	Partial Differential Equation
FEM	Finite Element Method
FD	Finite Difference
ED	Event Driven Simulation
TS	Time Stepped Simulation
CFD	Computational Fluid Dynamics
VR	Virtual Reality
HPF	High Performance Fortran
Adaptive	Software for Irregular Loosely Synchronous Problems handled by pC++, HPF extensions, Message Passing
Asynchronous	Software for asynchronous problems
Integration Software	Software to integrate components of metaproblems

unless it was bundled with an expensive piece of hardware. That's why we have so many MPP companies.....

3. Examine federal funding of software and structure it so that it encourages ISV's. This would include funding explicit software capitalization (InfoTech) activities. Secondly, one needs to ensure that federal projects enhance rather than compete with commercial activities. We need open (free) interfaces and not necessarily free software.

The above picture is that of a virtual corporation with products built of interoperable multiuse modules—each module in principle built by a separate organization, which are linked together in InfoMall to give a complete product. The HPCC national endeavor offers the possibility of a far larger and more exciting and important virtual corporation. This will only be possible when greater encouragement is given for software and user industries to participate in and benefit from HPCC.

Table 7: Runtime Software Support for Tables 1 to 5

<table>
<tr><th>STATIC</th><th>ADAPTIVE</th><th>ASYNCHRONOUS</th><th>INTEGRATION</th></tr>
<tr><td>HPF</td><td></td><td rowspan="2"></td><td rowspan="2">used in metaproblem modules</td></tr>
<tr><td colspan="2">HPF+, pC++</td></tr>
<tr><td colspan="4">HPF+ plus FortranM plus AVS, etc?</td></tr>
<tr><td colspan="4">HPC++
(invoking HPF+ modules for performance) supporting higher level systems</td></tr>
<tr><td>Regular Data Parallel or Embarrassingly Parallel Problems

e.g., finite difference

i.e., static (compile time identified) analysis</td><td>Irregular Adaptive Data Parallel Problems

e.g., data mining, multigrid

i.e., collective (correlated) irregularity</td><td rowspan="2">Asynchronous Problems

e.g., Event driven simulations, Expert systems, Transaction processing

i.e., no exploitable correlation in adaptive structure</td><td rowspan="2">Metaproblems Integrating Static, Adaptive, and Asynchronous modules

e.g., Command and Control, Ocean-Atmosphere integrated climate models, multidisciplinary analysis and design</td></tr>
<tr><td colspan="2">● Nearly ALL scientific and engineering simulations
● Gives massive parallelism in many metaproblems incl. lots of NII applications</td></tr>
</table>

Imprecise Mapping of Runtime Language in Tables 1–7
into Problem Class and Language Terms

STATIC	Synchronous and Embarrassingly Parallel Problems—current HPF (High Performance Fortran)
ADAPTIVE	Loosely Synchronous but not Synchronous—future capabilities of extended High Performance Fortran (HPF+) but can be supported well in message passing
ASYNCHRONOUS	Asynchronous
INTEGRATION	Metaproblems
	AVS works well but also can be integrated into languages such as HPC++, Fortran-M

Table 8: **What Does HPCC Small Software Systems Business Need?**

General: Economic Development Services

- Conventional start-up assistance
  - Incubator space
  - Marketing and administrative help
  - Tax breaks ...
- Funding
  - Venture and other private capital

InfoMall Services

- Funding (continued)
  - Support in SBIR Grants
  - Support in Other Federal Grants
- Enabling Technologies
  - Technologies needed for specific product
  - Technologies/Interfaces used in other products with which this must integrate
- Access to HPCC facilities
  - Distributed workstations
  - High speed networks
  - Central state of the art parallel computers
  - Visualization support
- Support for Virtual Corporation of several small companies
  - Distributed Collaboration Technologies

References

[Angus:90a] Angus, I. G., Fox, G. C., Kim, J. S., and Walker, D. W. *Solving Problems on Concurrent Processors: Software for Concurrent Processors*, volume 2. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1990.

[Fox:88a] Fox, G. C., Johnson, M. A., Lyzenga, G. A., Otto, S. W., Salmon, J. K., and Walker, D. W. *Solving Problems on Concurrent Processors*, volume 1. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1988.

[Fox:92e] Fox, G. C. "Parallel computing in industry—an initial survey," in *Proceedings of Fifth Australian Supercomputing Conference (supplement)*, pages 1–10. Communications Services, Melbourne, December 1992. Held at World Congress Centre, Melbourne, Australia. Syracuse University Technical Report SCCS-302b. CRPC-TR92219.

[Fox:93c] Fox, G., Bogucz, E., Jones, D., Mills, K., and Podgorny, M. "InfoMall: a scalable organization for the development of HPCC software and systems." Technical Report SCCS-531, Syracuse University, Syracuse, NY, September 1993. Unpublished.

[Fox:94a] Fox, G. C., Messina, P. C., and Williams, R. D., editors. *Parallel Computing Works!* Morgan Kaufmann Publishers, San Francisco, CA, 1994.

[Fox:94b] Fox, G., and Mills, K. *Information Processing and Opportunities for HPCN Use in Industry*, pages 1–14. Number 796 in Lecture Notes in Computer Science. Springer-Verlag, New York, April 1994. Proceedings of HPCN Europe 1994, "High Performance Computing and Networking."

[Fox:94c] Fox, G., and Mills, K. "Information processing and HPCC applications in industry," in *Proceedings of Annual 1994 Dual-use Conference*, Utica, NY, May 1994. IEEE Mohawk Valley.

[Mills:93a] Mills, K., and Fox, G. C. "HPCC applications development and technology transfer to industry," in I. D. Scherson, editor, *The New Frontiers: A Workshop on Future Directions of Massively Parallel Processing*, pages 58–65, Los Alamitos, CA, October 1993. IEEE Computer Society Press.

[Mills:94a] Mills, K., and Fox, G. "InfoMALL: an innovative strategy for high-performance computing and communications applications development," *Internet Research*, 4:31–45, 1994.

[Mills:94b] Mills, K. "Opportunities for HPCC use in industry: Opportunities for a new software industry in HPCC." Technical Report SCCS-617, Syracuse University, Syracuse, NY, March 1994. To be published in Parallel Computing: Paradigms and Applications, Chapman and Hall Publishers, London, UK.