Industrie Canada Industry Canada

# A STUDY ON HIGH-PERFORMANCE COMPUTING IN CANADA

## *FINAL REPORT*



CITI

Centre d'innovation
en technologies de l'information

Centre for Information
Technology Innovation

Canada

Industry Canada

# A STUDY ON HIGH-PERFORMANCE COMPUTING IN CANADA

***FINAL REPORT***



Prepared for the

Centre for Information Technology Innovation (CITI)

By

Ron Crossan
(R.H. Crossan Consulting)

and

Bruce Attfield
(iai Consulting Ltd.)



Laval
February, 1995

This report was prepared in connection with work carried out by the Centre for Information Technology Innovation (CITI) of Industry Canada. The opinions expressed in the report are those of the authors.


Catalogue no. Co28-1/122-1995E
ISBN 0-662-23444-8

# TABLE OF CONTENTS

Page

## TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

**TABLE OF CONTENTS**
(continued)

Page

## INTRODUCTION TO THE CANADIAN HIGH-PERFORMANCE COMPUTING STUDY

In July 1994, the Centre for Information Technology Innovation (CITI), an applied research laboratory of Industry Canada, commissioned a study to examine the state of high-performance computing (HPC) in Canada and to develop recommendations on a distinctive Canadian direction. This study was motivated by the realization of CITI directors that HPC provides a major stimulus to industrial competitiveness in the G-7 nations that have embraced the technology. Since there has been little apparent HPC activity in Canada, this study was intended to report on the actual situation and to develop recommendations for the dissemination of technology to the industries that could benefit most. Also, it was expected that detailed comparisons of the Canadian HPC environment to that of the U.S., Europe and Japan, where applications are far more advanced, would help identify the important opportunities open to Canada.

Following several months of research by designated expert consultants, a detailed preliminary report was issued on October 15, 1994, summarizing the HPC environment in Canada and making detailed comparisons with the situation in the U.S., Europe and Japan. This report was distributed for further study to an advisory committee of 18 major stakeholders, representing a well-balanced cross section of sectorial interest groups from all regions of Canada. The Committee met for a full day on January 19, 1995, to discuss the relevant issues and to arrive at a consensus of opinion on the recommendations for a Canadian strategy. These issues and recommendations are summarized in a companion report entitled *A Study on High Performance Computing in Canada: Executive Summary.* This final report provides all the detailed study information that was presented in the preliminary version, along with updates resulting from the feedback received since the release of the preliminary document.

This project was conducted by Ron Crossan of the Montreal firm R. H. Crossan Consulting, in association with Bruce Attfield of iai Consulting Ltd. in Ottawa. Other technical experts and specialists were used occasionally to complete the collection and analysis of the required information. We would especially like to recognize the contributions of Jim Stacey of

The Collaboratory Inc. and Elizabeth Pearce, President of Super*Can, the Canadian Association for High Performance Computing and Communications.

# 1 INTRODUCTION TO THE FINAL REPORT

## 1.1 Subject Matter

High-performance computing (HPC) is a field of activity that introduces computational science as a third scientific methodology, complementing traditional theory and experimentation. By exploiting the capabilities of a special class of very high-performance computer systems, mathematical models are developed to examine physical phenomena that are too difficult or too tedious to explore by means of the other two methodologies.

Academic researchers use these techniques to examine the physics of very small-scale phenomena, for example the molecular reactions that reveal interesting properties of new materials, genetic structures or drug interactions with biochemical organisms; or, at the other extreme, the researchers may study problems in astrophysics that are impossible to explore by any other means.

The findings and techniques of academic research are then passed on to industry, which does applied research to make their organizations more competitive, through better product or process designs, with shortened market delivery times. For example, an aerospace manufacturer may use HPC to design an aircraft by simulating a computer model, as a more cost-effective method than using a wind tunnel.

Government applies the same kind of modelling techniques to develop new policies to improve the quality of life, for example the use of climate models to better understand the long-term effects of noxious gas emissions into the atmosphere.

This report will also examine a new emerging use of HPC technology for commercial and government applications that involve massive databases of information. These applications need high-performance computers and networks to provide rapid interactive access and decision-support facilities. Many of these applications are incorporating multimedia information, which increases the size and complexity of the problems.

Most industrialized countries have important government programs in place to support the infrastructure requirements of high-performance computing and communications. The U.S., Europe and Japan each have multi-billion-dollar programs in place. This study examines whether it is appropriate to have a national strategy to support and develop this activity in Canada and will recommend specific actions that would establish a Canadian HPC program.

## 1.2 Scope of the Study

The Study on High-Performance Computing in Canada included a detailed assessment of the current Canadian situation and a comparison of the Canadian environment with that of the U.S., Europe and Japan. It examined the technology in place or planned the usage and application of existing facilities; the benefits to industry, government and the academic community; the sources of financing; the nature of collaborative efforts; the strengths and weaknesses of current policies and programs; and lastly, the identification of requirements, trends and success factors.

## 1.3 Methodology of the Study

The study commenced in early July with a review of the substantial amount of available Canadian and international documentation, previous study reports, journals, technical proceedings and other sources of accumulated knowledge. A list of reference material is provided in Appendix A.

The study then proceeded with an interview process carried out by fax, E-mail, telephone and personal visits, involving over 150 respondents. The purpose of this process was to confirm or validate the available information, request any important updates, collect more application success stories, and solicit opinions and recommendations regarding Canadian directions and strategies.

## 1.4 Report Structure

The following briefly describes the structure of the remainder of this report.

*Section 2 Introduction to High-Performance Computing (HPC)*

This section introduces the subject matter of HPC, describes the nature of traditional and emerging technologies within the study area, and introduces the types of applications that are characteristic of HPC, along with a general statement of the benefits of adopting HPC.

*Section 3 Review of the Current Canadian Situation*

This section reviews the HPC situation in Canada. Provided are a summary of the facilities in place; a review of the network requirements and availability; a discussion of the kinds of HPC applications in use or under development; and a report on actual or perceived strengths, weaknesses, major issues, benefits and trends. Also included is an examination of the Canadian HPC software-development industry as it is today and an indication of some of the future opportunities.

*Section 4 Comparison with Canada's Economic Partners*

*Section 4.1 U.S.A.*

The U.S. is indisputably the dominant supplier and the largest user of all varieties of HPC technologies, including hardware, software and networks. It is important for Canadians to understand the ways the U.S. has applied this technology, especially for improving industrial competitiveness. As Americans have done leading-edge research in almost every conceivable application area, we have much to learn from their successes and failures as we embark on any new programs in this country. It is also useful to study the nature of the government/university/industry collaboration, as these models provide lessons for Canada.

The U.S. government is the world's largest promoter of and contributor to HPC. The federal HPCC (High Performance Computing and Communications) program alone invests approximately $1 billion U.S. per year into HPCC R&D and infrastructure. As the federal government of Canada has no national HPC program and has only made a few relatively small ad hoc investments in this field over the past decade, it is important to examine the leveraging effect of the U.S. investment on such a grand scale.

### *Section 4.2 Europe*

Europe was an early adopter of HPC technology, but has abandoned any hopes of becoming an important supplier of the hardware technology itself. It does, however, have aggressive plans to become a world-class supplier of HPC software, especially with regard to applications for emerging parallel computer systems.

As Canadians, we can observe developments in Europe by individual countries that are closer in size to Canada than the U.S. However, what we see emerging rapidly in Europe are collaborative programs that include all members of the European Commission, providing combined resources to compete most effectively with the U.S. and Japan. The new European Commission HPCN (High Performance Computing and Networks) program is forecasting annual investments of $1 billion ECUs by 1995.

### *Section 4.3 Japan*

Japan is an interesting case study in the effective use of HPC to support the nation's ambitious R&D programs that have fuelled impressive economic growth over the past decade. However, because of differences in culture, language, political structure and business practices, it is less clear how much we could learn and apply to Canada. Certainly it makes sense to investigate more thoroughly how it is that Japan has integrated HPC technology into industry to an even greater extent than the U.S.

Appendix B lists the 176 largest HPC sites world-wide.

*Conclusions and Recommendations*

Those who read the *Preliminary Report* will note that this final version does not contain the final sections on issues, conclusions and recommendations. These items have now been reviewed by the HPC Advisory Committee and the summary of the deliberations and the final recommendations are summarized in the *Executive Report.*

A list of Advisory Committee Members is provided in Appendix C.

# 2 INTRODUCTION TO HIGH-PERFORMANCE COMPUTING (HPC)

## 2.1 Definition

High-performance computing (HPC), may be simply defined as the use of the fastest, largest and most advanced computers currently in existence. In the 1970s and 1980s, what we now call HPC was generally referred to as "supercomputing", as the only technology at that time that could solve many of the large problems encountered in science and engineering was a vector supercomputer (described in subsection 2.2.1 below). While vector supercomputers will continue to be one of the essential HPC technologies, at least until the end of this decade, several new technologies have emerged that offer the promise of superior performance at a lower cost by overcoming some of the inherent limitations in vector architecture.

To refine the definition of HPC, it is necessary to make a few introductory comments on the newer technologies that have come into play of HPC. First, desktop workstations have advanced so rapidly that they are now capable of solving some of the problems that previously required supercomputer access. Furthermore, such workstations are now the "window" for access to all high-performance systems, as a user will prepare his/her work and then "visualize" the end results at his/her workstation, regardless of where the problem is directed for high-performance processing. Secondly, since most workstations generally remain idle a good part of the time, techniques have been developed to network multiple workstations in "clusters" so that a single large problem can be partitioned in smaller segments that run concurrently across multiple machines, taking advantage of the idle time in the network. Thirdly, we have seen the emergence of "massively parallel processors", which give the promise of orders-of-magnitude price-performance improvements over the traditional vector supercomputers. There are also systems generally referred to as "multiprocessor servers", which provide another way of combining workstation elements into a single system to solve larger problems. For the next several years, these different technologies and approaches will be very complementary, as at this stage, no one technology can adequately solve all classes of problems. This report will also examine what is called the "metacentre", which in simple terms is a high-speed means of networking of all these heterogeneous technologies so that components of a large problem can be directed to the system that is most effective at processing each component. Lastly, HPC now includes the software and the networks that make HPC hardware accessible to the users.

What these various HPC systems have in common is that they are at the cutting edge of information technology (IT). Hence, HPC is important to the IT industry in general because it is at the frontier of advanced information technology. This alone makes HPC essential to any industrialized nation, because of the importance of the IT industry internationally. In addition, the advanced technology that is being developed in the HPC programs around the world is providing the information highway designers with base technology. HPC is most important, however, because of the extraordinary things it allows for and facilitates in other industries and in other areas of science and technology. It allows for the rapid development of new and improved products and services which would be impractical or impossible otherwise, and it opens the way to a whole new paradigm for scientific investigation.

As HPC is very loosely defined, it is important to describe the boundaries of the subject matter that have been established for this study. To some individuals, high-performance computing is restricted to very high-end computing technologies, such as vector processors and massively parallel computers, typically costing several million dollars each. This is somewhat the attitude in the U.S., where such technologies are prevalent. In Canada, where merely knowledge of these high-end systems is sparse, some individuals believe that the term should encompass everything down to low-cost desktop workstations and PCs. For the purposes of this study, it was decided to exclude this low end of the spectrum, recognizing that the low-end systems are ubiquitous, affordable and well understood, and that they do not provide the performance levels necessary to solve the grand challenges of science and engineering at an internationally competitive level. It is only by gaining access to the higher end of this technology that Canadian industry will improve its competitive stature in the global economy. Lastly, it is only the higher end of the technology spectrum that requires government strategy and support programs. Consequently, the scope of this study has been restricted to HPC systems that go beyond the performance levels generally found on a user's desktop, by accessing facilities such as vector supercomputers, parallel processors, multiprocessing servers and clustered workstations.

## 2.2 Mature and Emerging HPC Technologies

### 2.2.1 Vector Supercomputers

The first vector processor was developed and introduced into the market in the U.S. by CRAY in 1976. From that date until the beginning of the 1990s, supercomputing was synonymous with vector processing; the term HPC was not introduced until there were other contending systems available to warrant a more generic term for very high-performance systems.

A vector processor was the first computer to deviate substantially from the traditional Von Neumann architecture, whereby each instruction is executed in sequence, with one instruction commencing only when the previous instruction is completed. This traditional type of computing is also referred to as "scalar processing". A vector machine takes advantage of the fact that most scientific and engineering problems are expressed as systems of equations that are solved by a computer model using matrix computations. Each row or column of the matrix is a string of numbers called a "vector", and the computational process involves basic operations such as additions or multiplications of these vectors. In a vector processor, one instruction is issued to perform a complete vector operation, as opposed to one instruction per vector element in traditional computers.

This vector operation is then "pipelined", a process that can best be described using the analogy of an assembly line. In an assembly line, it takes a relatively long time to produce the first item, as it must pass through each stage of processing. However, once the assembly line is full, products comes off the end of the line at a very rapid rate compared to the time it took to pass through the line. The pipeline of a vector processor parallels the assembly line in this analogy.

Vector processing is not the only technique employed to speed up a vector supercomputer. First, these systems have always been built with expensive customized chips that have the industry's fastest clock cycle, such that more instructions can be executed per second, even in scalar operations (not all operations are vector). Today's systems typically operate with clock cycles around 3 nanoseconds, which is several times faster than any other technology.

Secondly, these systems operate with very fast main memories, since memory access speeds are often the bottleneck in high-performance computing. Special techniques are used to provide very high bandwidth for transferring data between memory and the processors. Furthermore, the systems are available with very large main memories, which permit very large arrays of data to reside in memory while processing, to avoid the relatively slow "swapping" of data to and from disk storage devices that is required by smaller memory systems. There is normally also an option of secondary memory, which provides storage of even larger arrays in high-speed memory to further avoid the much slower disk-access speeds.

Even the mass storage systems (e.g. disk access) are higher-performance, as there will always be occasions when a problem will not entirely fit into these large main and secondary memories. The higher performance is realized by higher-speed channels and the use of parallel channels to move the data at higher speeds.

Lastly, the system performance has been further enhanced by allowing for more operations to proceed in parallel. The U.S. manufacturers have addressed this challenge by offering several processors (currently up to 32) in one system. These processors can either be assigned separate jobs to increase the throughput of the system, or they permit true parallel processing by allowing one job to be segmented, with each component running simultaneously on a separate processor. The Japanese manufacturers originally offered parallelism on vector machines by providing multiple pipelines that could all operate simultaneously, but they too are now offering multiple processors. This hybrid type of system is often called "vector-parallel-processing" or "parallel-vector processing". Traditionally, vector-parallel processors have been designed in such a manner that all processors share access to one central memory. This type of parallel processing has severe limitations as to the number of processing elements, when compared to massively parallel processors (MPP systems), described below. As the number of processors is increased, there is increasing contention for central memory access, thus creating a situation of diminishing returns as more processors are added. The Japanese are now offering distributed memory versions of vector-parallel systems to circumvent this restriction on number of processors.

In summary, these vector supercomputers have been designed as very balanced systems, speeding up all components of processing to avoid single bottleneck points that could slow down the whole process. These systems have had almost twenty years of development, resulting in well balanced systems improvements in ease of use, and the availability of large libraries of application programs. However, this type of architecture has its limitations and may have severe restrictions for future growth relative to other approaches described below.

### 2.2.2 Massively Parallel Processors (MPPs)

In the description above, it was mentioned that partitioning a problem into multiple components segments and executing pieces in parallel on separate processing elements was an effective way to speed up processing. However, there are two restricting limitations with traditional vector machines if an attempt is made to extend the concept to very large numbers of processors. First, each processor is very expensive, so a very large number of processing elements would result in an exceedingly expensive configuration. However, the fundamental limitation of the traditional supercomputer is that it uses "shared memory", i.e. all processors share one physical central memory, which creates memory-contention problems when we build systems with too many processors (currently the limit is 32). The system bogs down when too many processors are simultaneously trying to access one common memory.

At the risk of generalizing, it can be stated that MPP systems are differentiated from traditional supercomputers by resolving these two limiting factors in the following manner. The MPP systems circumvent the cost problem by using inexpensive off-the-shelf micro-processor chips, usually the same chips that are mass-produced at low cost for workstations. Although even two or three years ago these chips were very slow in relation to the custom chips used in vector processors, the gap is rapidly closing, as the rate of improvement of MPP (or workstation) chips is doubling approximately every 18 months an improvement rate significantly higher than that of the vector-technology chips. The memory contention problem is normally resolved by replacing shared memory by distributed memory, whereby each processor has its own local memory. There is no precise definition of MPPs, as the field is too new and there are too many different architectural approaches. However, most experts would agree that to qualify as an MPP, the system should provide scalability to at least 1,000 processors and physical memory must be distributed (even though some developers make this appear like shared memory for ease of programming).

The net result of these two factors provides the promise of much higher performance at much lower cost than any other approach that is feasible today. There is currently a race to build a "teraflop" machine by 1996-1997. By definition, this type of unit will execute one trillion arithmetic operations per second, approximately 40 times faster than today's fastest shared-memory vector machines. Visionaries are even meeting to design the "petaflop" machine, 1,000 times faster again, expected to be available somewhere around the year 2010. Traditional vector machines can not compete in this race because of the shared memory problem and because of the severe technical difficulties in significantly speeding up the clock speeds of the few processing elements in the system.

MPP is an emerging technology, unlike the mature vector technology. Consequently, there are still many ongoing experiments to arrive at the best architectural design, relatively little software exists and the systems are not yet robust. However, experts world-wide realize that this is the most promising technology for much higher performance at affordable rates. Far more research dollars are being spent on developing this technology than on any previous HPC architecture. The manufacturers of vector technology are now shifting most of their R&D to MPP development.

MPP systems are very complex with respect to the manner in which programs must be partitioned and the techniques used for managing the concurrent execution of the separate processes. The real challenge is to develop software that automates this process as much as possible and hides the complexity from the user. There are many conflicting approaches to solving this challenge and it will be some time before standard techniques emerge. For example, some researchers are promoting the development of new programming languages that are designed to work with the parallel systems, while others promote retaining traditional languages with the addition of parallel constructs in order to preserve the enormous investment in legacy application codes. It would appear that the latter approach will prevail, as standards have emerged that provide parallel application development using HPC's traditionally favourite language: FORTRAN.

A second level of complexity, which can have a major negative impact on the performance of an MPP system, is the necessity of individual processors to communicate with each other during execution by "message passing", for control and coordination, and data passing, when access is required to information in another processor. It is, however, beyond the scope of this introductory overview to explore this fascinating topic in any further technical depth.

It is worth noting one other trend with respect to MPP, as it represents a major deviation from what until recently has been the private domain of HPC. MPP systems are now being introduced to the commercial sectors of computing as very powerful database servers. While MPP was originally designed for large computationally intensive scientific problems, it turns out that the same systems, with some additional software, can out-perform all other systems by a significant margin in transaction processing and database searches. Using the MPP's parallel capability, different requests can be concurrently processed by the multitude of available processors. Using the MPP as a database engine is most promising for certain information highway applications where enormous databases will be accessible to the masses. The need for parallel processing of information is even far more acute in new multimedia applications because of the massive numbers of bits of information that must be stored and moved for digitized video applications.

### 2.2.3 Multiprocessor Servers

In addition to the MPP systems introduced above, there are two approaches to providing moderate parallelism, also using off-the-shelf workstation processor technology. The two approaches are generally referred to as "multiprocessor servers" (described in this subsection) and "clustered workstations" (described in subsection 2.2.4).

A server is a machine on a network which satisfies the special requirements of a group of clients, in the so-called client-server model. In the HPC world, the client sits at a workstation, where he/she locally solves all of the problems that are within the capacity of this desktop unit and passes the larger tasks on to a server on the network. The server could be a local or remote vector processor or an MPP system. Alternatively, the server could be a "multiprocessing server", which is often a device that is compatible with the workstation, but contains multiple processor chips, all sharing one central memory. These shared-memory servers go by the technical name "symmetric multiprocessor servers" (SMPs). The concept is much like the shared-memory vector supercomputer, but it is limited to scalar processing and uses inexpensive workstation processor and memory chips. Because these systems use shared memory, there is again an upper limit to the number of processors that can be used, as the memory contention is once again a problem. Most of these systems currently involve 20 or fewer processors, but some vendors are trying to push the technology up to 64 processors.

The advantages of this technology are that it is relatively inexpensive, it is normally compatible with a given workstation technology and it is easier to program than MPP systems. It provides higher performance than a single workstation, but does not hold the promise of scaling to the high-performance levels of MPPs. Using this technology effectively in parallel, i.e. to partition one job to run concurrently across multiple processors, is less difficult a task than using a distributed memory MPP, but the challenge is still not negligible.

### 2.2.4 Clustered Workstations

Large engineering and scientific organizations often possess countless numbers of workstations scattered throughout their facilities. Many of these workstations will have a high percentage of idle time, either because the user is not active, or because the user is active but is performing trivial tasks such as E-mail, which leaves many of the processing cycles unused. It be wonderful if a large HPC user could simply go over the network and appropriate all these wasted cycles for a large job. This is precisely the purpose of clustered workstations.

A set of clustered workstations is conceptually similar to an MPP system. In other words, a user attempts to partition his program over multiple processors, each of which has its own memory (i.e. the distributed memory described as an attribute of MPPs). Consequently, the user faces all the same challenges of MPP complexity with one additional concern: the network that connects individual workstations is a local area network (LAN), which is very slow compared to the high-speed internal network that connects processors in an MPP. In parallel executions, there is normally substantial communication between processors for transferring data and managing synchronization.

Despite the challenges of mastering efficient clustering, the motivation to find solutions is extremely high, as it provides a method of capturing unused cycles which are virtually free (other than the cost associated with converting programs to run in parallel). As with MPP systems, there is much research being conducted on this process, with the objective of making the process of parallelization as transparent and automatic as possible to the user.

### 2.2.5 Visualization

So far, this section has described the various techniques in use or under development to provide the highest possible processing speeds for computationally intensive scientific problems (or massive database searches in the case of one special application of MPPs). The scientific applications achieve their accuracy by simulating reality over dense meshes in time and space, with resulting voluminous amounts of output data. Often the results could fill millions of pages if they had to be provided as printed numbers, and no one would have the time to interpret what this mass of numbers means. Fortunately, over the past several years, there has been remarkable progress in terms of displaying these simulation results as pictures, a process called "visualization". Three-dimensional colour images, with optional full-motion video, can render the masses of data comprehensible to a scientist who intuitively knows what he/she is looking for.

In HPC it is not just 3-D objects that are displayed, but most importantly, physical attributes are juxtaposed on top of the geometry to display colour-coded values of parameters such as temperature, pressure and stresses, with the simulated variation of these same parameters over time.

Most visualization can be most effectively displayed on the user's workstation once the numerical results of a simulation have been saved on mass-storage devices. As HPC and network performance improves, some applications are being developed to show visualization throughout the actual simulation process on a background HPC system. This enables a scientist to adjust the parameters of his/her model during its execution, thereby probing deeper and more efficiently into the mysteries of the simulation in a dynamic fashion.

Visualization will continue to be a critical area of development for HPC environments. There will be more investigations into the best ways to communicate visual, numerical, mathematical and dynamic elements to the user. This will open the door to new perspectives on reality, allowing for a deeper understanding of the generated data, which could, for example, result in better interpretation of oil reserves by means of seismic processing or a better understanding of molecular interactions which would lead to more rapid design of effective drugs.

### 2.2.6 Comparisons of Technologies

All of the above technologies are currently in active use, each for different kinds of applications, throughout the industrialized world. None, including the vector supercomputer, has reached the end of its life cycle.

The vector supercomputer is still very much in demand by organizations that require a robust production system in order to run a wide variety of working application programs in a variety of disciplines. An excerpt from the U.S. Congressional Budget Office June 1993 Study on HPC may help to clarify the continuing important role of vector machines relative to emerging MPP systems:

> Displacing vector supercomputers from their current position is not likely to be easily or quickly accomplished. It is not yet certain that the higher theoretical peak speeds of the massively parallel supercomputers will translate into features that the marketplace will value. The difficulty in programming the massively parallel supercomputers, as well as their lower performance on many kinds of problems, may limit their share of the supercomputer market. Finally, many massively parallel supercomputers are incompatible not only among manufacturers, but in some cases also between generations.
>
> In benchmark tests conducted at NASA's Ames Research Laboratory, the actual performance of several brands of massively parallel supercomputers across a wide range of problems proved to be between 1% and 5% of theoretical peak efficiency. In most problems, the massively parallel supercomputers were able to obtain only a fraction of the speed of the older vector supercomputer.
>
> Close to 20 years' worth of software has been refined to run on that [vector] hardware. The difficulty of writing software for massively parallel supercomputers will make them slow to supplant conventional supercomputers.

MPP systems however do hold amazing potential once sufficient system and application software is available. Today, most parallel machines have been installed in research labs for the development of software, especially parallel programming languages and tools (e.g. analyzers, optimizers and debuggers). Once the fundamental development tools are generally available, then the real-world applications must be converted or re-written. Some universities and research labs have already written application codes that will significantly out-perform vector machines; at present however, these are merely isolated cases. One of the advantages of MPP systems is that small configurations, adequate for experimentation, may be acquired for very modest investments (often starting at around $200,000). Since small start-up systems are possible and since the systems are used mostly for software development and experimentation (as opposed to production runs), most systems installed today could be called moderately parallel, with fewer than 64 processing nodes.

Jack Worlton, a leading U.S. spokesman for HPC, sums up the MPP situation as follows:

> Managers of organizations dependent on computational science and engineering need to make sufficient commitments to MPP so their users can experiment with and learn about different MPP approaches—a "trialability" phase—but until the technology and industry of massive parallelism matures, it is imperative for organizations to maintain a world-class production capability in mature high-performance computers, that provide sustained performance, rich application libraries, transparent system software, generality, compatibility, high reliability and programmers skilled in their use.

In addition to the requirements for MPP systems to satisfy the demand for computationally intensive scientific problems, we have noted that there is also a spin-off market that seems to be developing for commercial database applications. This study examines the relevance of this opportunity to Canada's information highway.

While most analysts believe that MPP systems are the future of HPC, at least at the high end of the performance scale, experts disagree on how near and how parallel that future may be and whether we will get there through incremental steps or major leaps. By trying to leap-frog the development of intermediate technology (between vector and MPP), producers of MPP risk leaving software producers and users with no simple path for the evolution of their products. To some extent, it is the multiprocessing server that can provide a bridge. As described above, these servers have performance limitations, as the shared-memory architecture prevents scalability above some limit—perhaps 64 processing elements. Nevertheless, they allow for important improvements beyond the limitations of the individual workstation, and for many applications they may extend the range far enough. They have the advantage of compatibility with the workstation itself, although considerable effort may be still required to add parallelism.

No matter what happens at the high end of the HPC spectrum, there will be a growing demand for and acceptance of clustered workstations. The reason is quite simple: clustering takes advantage of idle time on workstations that have already been acquired (i.e. no incremental costs). While many software tools have entered the market to aid in the parallelization process, these systems still suffer from the same lack of system and application software, in the same manner as MPP systems. Also, this alternative will become more attractive as local area network bandwidth increases (e.g. with the acceptance of Fast Ethernet and ATM LANs), as clustering often requires very high-speed communication between the collaborating workstations in the cluster. However, in this case, any improvement is welcome, as it will involve virtually no incremental cost.

Finally, visualization has become the only practical method to view and interpret the results of computational science. Consequently, results are visualized regardless of whether the simulation is executed on the user's own workstation or on a remote supercomputer. Normally, a results file is downloaded to the user's workstation and then the user displays the results, using the workstation itself to vary the perspectives, employing techniques such as zoom, pan, slicing, rotation and rendering.

In conclusion, the point has been made that while there are several very different architectures that are currently competing for the user's HPC dollar, no one technology can adequately solve all problems. Whenever a local server can solve the user's problem in a realistic amount of time, it is normally the most cost-effective solution. Where sufficient enhancement in performance can be achieved with clustering an organization's workstations, this is almost a zero-cost booster. Where more performance is still required, the user has the choice of a wide variety of architectural approaches to vector processing, vector-parallel processing and MPPs, each of which is well suited to a certain class of problems while not necessarily to others. It would be wonderful if each user had a choice between all these alternatives for each different kind of problem or, better still, if a user could partition a single large problem and direct each segment to the type of system best capable of processing that segment. With very high-speed national broadband networks, this is possible. In effect, this is clustering on a grand scale. In fact, the U.S. HPC community is moving in this direction with a concept called the "metacentre", that will be explained at greater length later in this document. We will examine how this approach is particularly advantageous for Canadians, as it also allows for the sharing of a small number of scarce, expensive HPC resources.

## 2.3 Needs and Benefits of HPC

The needs, uses, applications and benefits of HPC should become evident as we describe the various users in Canada and internationally. This section well briefly introduce the types of problems that most benefit from HPC technology.

*Computational science* is a new scientific methodology that complements the two traditional methods, theoretical science and experimentation. The traditional methods have distinct limitations that can be circumvented by this new method. In traditional experimentation, nature is often too difficult to investigate—for example, when the phenomena under observation are too small, too far away, too fleeting, or they unfold over too long a time period. Theoreticians can provide very precise general equations, for example, the law of general relativity, but they are not able to mathematically evaluate complex real-world scenarios such as predicting the future state of the environment. Computational science involves the construction of a digital computer model which is then simulated as described below.

*Computer modelling* of a large variety of physical systems has reached the point where reality can now be simulated with a high degree of accuracy. High-performance computers enable scientists to replace the physical world with numerical models, themselves derived from a mathematical representation of the physical system being studied. These models can be coded for simulations on high-performance computers, and the results can be transformed into visual imagery which is more intuitively accessible to human beings. The scientist's challenge then becomes one of refining the model so that it corresponds as closely as necessary to the physical reality that is being modelled in the first place. These models of real physical systems (such as the atmosphere, turbulence, chaos, combustion in chemical/mechanical systems, aerospace and automotive vehicles, engines, protein molecules, factory operations or economics) can be sufficiently detailed to be used for realistic predictions. Once it has been verified that the model is accurately predicting behaviour of the phenomena being simulated, then engineers and scientist can begin to use the computer model as an alternative to time-consuming and expensive laboratory experiments.

*Simulation* by means of high-performance computer systems is increasingly replacing experiments and scale modelling in a wide range of scientific, engineering and commercial applications. In some industrial areas, such as in the design of aircraft, space vehicles, engines, ships and cars, simulation is already an inescapable necessity. In other areas, scientists and engineers are on the verge of large-scale breakthroughs in the design of drug molecules, enzymes, catalysts and new materials. The time elapsed from the conception of a new product to its introduction on the market and the cost of its design process can be dramatically reduced with the help of computer simulation. This time reduction, combined with the accuracy attained in reproducing reality, increases industrial competitiveness by reducing design costs and improving product quality.

Industry can use computer models to shorten design time and to reduce the number of prototypes that have to be built. Petroleum companies can model the earth's subsurface from seismic surveys to minimize the number of dry holes that are drilled. Environmentalists, by modelling the atmosphere and oceans, can predict the effects of pollutants and noxious gases years before experimental results provide the data. Pharmaceutical companies can bring new effective drugs to market much sooner, by using computer models to focus on research that has the most promise.

High-performance computers, which in the right circumstances can process information thousands of times faster than personal computers and workstations, have become the engines that solve these complex numerical models of physical phenomena. They have become indispensable tools for researchers and engineers in disciplines where computer modelling provides high payoffs, including geophysics, microelectronics, environmental modelling, chemistry, physics, material science, biotechnology, product design, medical imaging, remote sensing, and economic modelling.

*Optimization* is similar to simulation, insofar as it depends on the availability of accurate mathematical models of reality and powerful computers. Optimization is a rather new application for high-performance computers and it is becoming increasingly important in the operations of large industries and service-providing companies. An example of a highly challenging use of optimization involves the search for the most efficient design of an entire aircraft. This is of the utmost importance for economic designs of fuel-efficient aircraft. In many areas of business, the allocation of resources such as people, raw materials, capital or time can be among the most challenging problems. The difference between merely good solutions to these problems and a near-optimal or optimal solution may mean the saving of huge sums of money. The main application areas for optimization include the design of complex structures, investment and portfolio management, production planning and scheduling, distribution planning, vehicle routing and crew scheduling. The use of HPC coupled with improved human interaction can dramatically enhance the creative power of the human mind.

*Data mining* is a term that refers to new class of HPC applications for the extraction of information under complex query criteria from massive databases. Until recently, large database systems were the exclusive domain of large mainframe systems. However, the HPC community is now demonstrating that MPP systems, by performing database queries in parallel, out-perform mainframes by orders of magnitude, thus opening the possibility for new large-scale applications for both massive corporate and government databases.

# 3 REVIEW OF THE CANADIAN SITUATION

## 3.1 Introduction and Background

Canadian recognition of and reliance upon high-performance computing dates back to the early 1960s. At that time, three systems that were forerunners of modern supercomputers were in active use. These included systems at the Central Analysis Office of the Weather Service, in Dorval Quebec, at the Research Laboratories of Atomic Energy just west of Ottawa at Chalk River Ontario and at the University of Toronto.

Large vector supercomputers found their way onto the Canadian scene in the early 1980s with a major installation at the Canadian Meteorological Centre of the federal government's Atmospheric Environment Service (AES), otherwise known as the Canadian Weather Service. Since then, AES has maintained its Canadian leadership position in the use of this technology by implementing leading-edge U.S. and Japanese supercomputers. These systems perform analyses of the Canadian weather situation and provide essential information to the public, government and industry. They also provide climatic and atmospheric data to the research community, but are used primarily for numerical weather prediction. The facility has won Canada recognition as a world leader in this field since the early 1960s.

Vector supercomputers were also installed in the late 1980s at the University of Calgary and at the University of Toronto to support researchers from a variety of disciplines, while offering industrial and other research organizations opportunities to use the facilities as well. Both facilities were forced to close down because of funding situations that would not support maintain its position as a leader in modern technology.

Five additional supercomputer facilities were set up in the late 1980s, but for only brief periods. Two industry start-up companies in Toronto and one on the East Coast went bankrupt shortly after their supercomputer facilities opened, while the supplier of supercomputers to the

University of Western Ontario and two Calgary seismic companies ceased operations and could not support the installations.

The current facility base described below reflects the Canadian situation over the past five years.

Adequate networks to support access to high-performance computing have never been a Canadian priority, though the availability of Canada's primary research network, CA*Net, has allowed for some cross-country HPC access at low to medium speeds.

Funding for shared central facilities in high-performance computing and networking has primarily been dependent upon provincial and federal government infrastructure grants, as there has been no evidence that HPC centres can become commercially viable without such support. This approach parallels the situation in the U.S., Europe and wherever else such systems are in use, with the possible exception of Japan. The major difference for Canada appears to be the comparative level of funding per capita and the period of commitment for funding. In the U.S. and Japan, such funding currently exceeds C$5 per capita per year. In the European Union, plans suggest an investment that will exceed C$3. The known plans of the governments of Canada and its provinces are sporadic and certainly do not exceed C$0.25 per capita per year.

## 3.2 Canadian HPC Facilities

High-performance computing, as defined in Section 2.1 excludes desktop or single-user workstations, despite the continual growth in capacity of such devices.

The information presented here on Canadian HPC facilities is drawn from the prior knowledge of the consultants, as well as data gathered from the vendors and users of HPC equipment. To support the analysis, detailed questionnaires were sent or administered through interviews with a total of 112 identified users of HPC technology, representing a good cross section of the different sectors where HPC is being used today. Of the approximately

46 organizations that responded to the survey, some operated multiple systems. The information collected from the questionnaires was complemented with information gathered in previous surveys, the consultants' direct knowledge and dozens of telephone interviews to those who failed to send completed questionnaires.

Seven vendors were contacted for supplementary information regarding user sites. This excluded the two large Japanese vendors of vector systems, as the authors have close knowledge of the single, very large facilities that these companies have set up here in Canada. It also excluded two major MPP-system vendors from the U.S., each of which has one site in Canada, and one U.S. multiprocessor server vendor, which has a system installed but is no longer in business. Thus, the information gathered represents the offerings of 12 HPC system vendors (9 American, 2 Japanese and 1 Canadian). A few very small systems, dedicated to specific applications, may have been overlooked in the consultants' research.

Two HPC vendors did not responded in detail. First, Cray Research (Canada) Inc. expressed concern over the release of its client database, even though the firm was offered assurances that any information specifically identified as confidential would be protected. However, through separate sources, all the major CRAY installations have been identified. Secondly, Alex Informatique Inc. of Montreal, states in its corporate literature to have upwards of 100 installations but identified only three in the questionnaire. These three sites are included in the results presented below; there is no information generally available to confirm that any additional Alex sites indeed have capabilities comparable to the other high-performance computing systems in general use.

Despite these qualifying remarks, this review of Canadian facilities identified a significant number of HPC installations, as summarized in the table that follows. Although the numbers are significant, most of these installations are very small departmental machines. For example, most experts would agree that for a system to qualify as a supercomputer today, it must have a peak rating of several gigaflops (billions of operations per second) and sustain a reasonable percentage of this level of performance during major application execution. In the table below, there are only two such machines in Canada, compared to several hundred in the U.S., Europe and Japan. Appendix B lists the world's largest installations (176 in total) and includes only one for Canada.

The total number of HPC systems identified in Canada, of all sizes and types according to the definition provided in Section 2, is as follows:

| Vector | MPP | SMP | Cluster | Old | Total |
|---|---|---|---|---|---|
| 3 | 17 | 19 | 5 | 25 | 68 |

It is important to note that this summary excludes the Canadian security sites that employ such systems. The summary also excludes small MPP installations, where the devices are used for specific, probably dedicated, research and are therefore not easily identifiable.

An important, though almost historical, record is the number of "old" vector systems which are well outdated in accordance with today's HPC standards but may continue to be in operation. Their existence reflects the interest and perceived benefits of such systems to Canadian users in the mid-1980s, but, sadly, it also reflects the current lack of funding to renew the technology. The summaries that follow exclude these "old" facilities.

The following summaries exclude the Canadian security sites and "old" vector facilities.

The 44 more modern HPC installations identified are operated by 31 organizations. One of these is Canada's only centre that offers HPC services on a commercial basis and is described in detail in Section 3.4.

### 3.2.1 Regional Distribution of HPC Systems

The provincial distribution of HPC systems is outlined in the table that follows. No installations were identified in the Yukon or the Northwest Territories.

| BC | Alta. | Sask. | Man. | Ont. | Que. | NB | NS | PEI | Nfld. |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 0 | 0 | 12 | 17 | 0 | 3 | 0 | 1 |

### 3.2.2 Sectoral Distribution of HPC Systems

Sectoral distribution, as outlined in the following table, includes, the Government category, organizations that are primarily funded by government grants, despite their goals to supplement revenues by commercial offerings. Industrial organizations are those that operate without government or university funding.

| Government | University | Industry-Petroleum | Industry-Aerospace | Industry-Other |
|---|---|---|---|---|
| 14 | 17 | 6 | 3 | 4 |

One final classification of installed systems could be performance level. It has been difficult to obtain exact performance measurements, as there is no commonly accepted method of defining the performance level of HPC systems. Most industry surveys classify systems based on peak theoretical performance. As such surveys today start this type of classification at several gigaflops (billions of operations per second of peak theoretical performance) and must sustain a reasonable percentage this of performance during major application benchmark execution, for Canada there are definitely only two such machines, the two vector supercomputers at the Atmospheric Environment Service in Dorval, Quebec (each rated greater than 25 gigaflops). There are 10 other machines that are rated at more than 2 gigaflops but less than 5 gigaflops.

One is the vector system at the HPC Centre in Calgary and the rest are SMP and MPP systems on which the actual performance on real applications can be much less than their advertised peak performance because of the difficulty of developing software to take advantage of the parallelism.

The more meaningful evaluation of performance capability is to measure the number of sustained operations per second on real applications. In the case of the Atmospheric Environment Service in Dorval and the HPC Centre in Calgary, many of the production applications actually sustain performance at levels above 2 gigaflops. There were no identified examples of performance levels above 2 gigaflops on real applications for any other system in Canada.

### 3.2.3 Industrial Uses of HPC Facilities

The dominant industrial application for HPC in Canada at this time is clearly seismic processing. This application may be run directly by a petroleum company, but more frequently it is contracted to a seismic processing company. With one identified exception, the application software used in this field appears to have been developed in the U.S. and Europe; relatively little software has been developed here in Canada. It seems that compatible in-house installations have been implemented in Canada to counteract the costs and delays of shipping massive data for processing at off-shore parent companies over expensive networks. Many oil, gas and seismic companies that do not have their own HPC facilities are nevertheless accessing such facilities at either the HPC Centre in Calgary or at their parent company's facilities in the U.S. Many companies have a plethora of workstations, presently operating as separate independent devices, but with plans to experiment with clustering software to enable faster processing by "parallelizing" their applications over multiple workstations.

Other industrial installations support product design and manufacturing, including two in aeronautics. There is one private company that uses an MPP system to support software development. (Machines that may be installed on equipment vendors' premises for demonstrations and other marketing functions are not included in the survey tables).

In discussing the benefits of their in-house installations versus the use of a shared facility, the comments of industrial users strongly reflect a need for in-house management and control of their processing. They find that the results under such circumstances are of higher quality and are obtained more quickly, and thus better serve organizational goals. Industry also has concerns for security and data bandwidth, as HPC is typically used for competitive-edge applications.

Examples of usage and the accrued benefits are described below.

## 3.3 Some Samples of Dedicated (In-House) Facilities

### 3.3.1 Atmospheric Environment Service (AES)

The AES operates Canada's largest supercomputer centre, located at its facility in Dorval, Quebec. According to an HPC newswire in August 1994, this facility ranked 23rd in computing capability amongst the world's top 176 sites (see Appendix B). No other Canadian facility ranked.

This very modern facility has extensive security and reliability features, which ensure the consistent quality of its weather forecast-support products. The AES application, and justification for the computer centre at the Canadian Meteorological Centre, is virtually singular. Canadian success in numerical weather-prediction research is world renowned, and many other nations are basing their methods on those developed here in Canada. The current very large vector supercomputer supplied by NEC of Japan, operates in dedicated use for two periods each day, 365 days per year. During these periods, which last for about two hours each, the vast majority of other applications are set aside and numerical weather prediction runs are executed. Each run is based on the previous run, with data supplementation occurring in between each one. The output is primarily in the form of graphics, which are disseminated by the AES's sophisticated broadcast networks across the country.

The run period is of consistent duration and is bracketed by the cut-off period for data reception and the hour at which new forecast-support products must be distributed. Additional computer power simply means that a finer mesh model can be executed, providing improved product output.

The balance of the time available on the large AES supercomputer is used for research activity and continual model enhancement. The second supercomputer will again be used for model enhancement, other AES activities such as climate research, and revenue-generation initiatives. The system is also used for research on climate and atmospheric conditions. In addition, the AES has traditionally allowed other users from government and occasionally from industry to use the supercomputer facility; however, external usage has been minimal.

AES has entered into long-term lease contracts with computer vendors with pre-planned updates to their facilities. These multi-year commitments ensure consistent cash flow over long periods while ensuring the availability of leading-edge technology. Minimization of application conversion and the opportunity to plan for new models is a side benefit. The current contract for a NEC SX3/44 commenced in 1991 and a second, more powerful system, an SX3/44R, is currently being installed. Another upgrade is contracted for 1997.

AES and NEC have agreed to continue to operate both the new and old systems at the Dorval facility. This initiative perhaps provides opportunity for expanded Canadian use of this class of HPC system, depending on AES's plans for its ongoing use in support of its existing applications and research.

Funds for this centre are within AES's annual budget, although comments expressed by the centre operators indicate that government budget cuts are having a major impact.

Users and support personnel of the AES facilities have all undergone in-house training and are experienced. New recruits are often recruited through AES support of the co-op programs offered by many educational institutions and through AES involvements in equal-opportunity programs such as those for the handicapped and for Native Canadians. Support services to users include application development and optimization assistance, as well as technical troubleshooting.

Marketing expertise will be an important requirement if AES intends to open its expanded facilities to external users. The availability of the second SX3 was announced at the June 1994 Supercomputing Symposium.

### 3.3.2 Research Computing British Columbia (RCBC)

RCBC is a "logical entity" within a government crown corporation, BC Systems Corporation. RCBC offers commercial services on its massively parallel system. The system, a Maspar MPP was initially donated by Digital Equipment of Canada, but because of its success in commercialization, RCBC has been able to expand the system considerably using its own funding. Principal applications include a large database decision-support and statistical-processing system for medical records and a protein/DNA sequence alignment system.

About 60 users from government, industry and academia use the facility and RCBC also offers the system on a "pay-per-use" basis.

Support services by E-mail and telephone are offered as an extension of BC Systems Corporation services. Training courses are provided by the vendor.

### 3.3.3 University of Alberta

The University began experimenting with HPC using on-campus workstation clusters, based on IBM high-performance workstations. These have been some 12 users on 23 workstations, which were purchased at favourable rates from existing research programs. Interconnection is by Ethernet LAN, whose speed allows for effective use of the cluster on certain parallelizable problems with very large granularity. The load levelling of tasks for these facilities is handled by the Myrias WPAMS software, developed in Canada.

Recently, the University upgraded its facilities with the acquisition of an IBM SP2 MPP system. Like similar institutions, the University is convinced that "high performance computing... provides for the expansion of the frontiers of science, allows problems to be solved that hitherto were impossible to approach, and provides a platform for competitive advantage through partnership with industry and academia".

### 3.3.4 Shell Canada Limited

Shell in Calgary recently replaced its obsolete Cray supercomputer with an IBM MPP system to process a variety of seismic applications. Shell's principal application software was developed by its parent firm in the Netherlands. There appears to be little local application development, though there are some 50 users applying the system to local problems. The main reasons for this in-house system appear to be to take advantage of company-developed software, better quality of results and more rapid processing.

### 3.3.5 University of Calgary

A joint Chemistry/Computer Science facility has been installed at the University to serve its more than 60 users. The 14-processor Power Challenge SMP system provided by SGI has been funded by departmental budgets with support from NSERC and industry. Activities include research into global optimization, contour tracing, image processing, modelling and visualization of biological processes, design of new drugs, optical properties of ring systems, the development of metal-based catalysts and new materials, spectroscopic properties of molecules and biological electron-transfer processes.

## 3.4 Publicly Accessible HPC Facilities

There is currently only one privately operated facility offering HPC services to the public in Canada: the HPC Centre in Calgary. In addition to the Centre, which is described below, we have included brief comments on some other facilities that occasionally offer external services to the Canadian HPC community.

### 3.4.1 HPC Centre in Calgary

The High Performance Computing Centre (HPCC) is located in Calgary. The HPCC offers machine services and user support on a large Japanese-manufactured vector supercomputer, a Fujitsu VPX240/10. HPCC officially opened for business in June 1993. It is a not-for-profit corporation with initial funding of $5,100,000 from the Alberta Government and $5,000,000 from the federal government, through the Department of Western Economic Diversification. Continuing operation will be dependent on success in the marketing of these facilities and fulfilling technology enhancement and renewal plans.

Collaboration between the supporting government groups, industrial interests in seismic processing and in software development, Fujitsu Systems Business of Canada, as well as Alberta universities, led to the establishment of the Centre. While consideration was given to locating the facility at the University of Calgary, which had experience in operating an older vector supercomputer for many years, it was concluded that industry interests would be best served by alternative arrangements. Thus, the vector supercomputer is operated on behalf of the HPCC at the major computing complex in Calgary operated by Systemhouse Ltd., which also provides HPCC users with a "hot line" for problem management.

The HPCC offers a full range of support services to its clients; its professional staff offers application development, application conversion, optimization and operational support. While these services are currently concentrated in Calgary, the HPCC will provide support to users at their local facilities at a level that is appropriate to the contractual commitment.

The HPCC offers its users a number of application software packages. The number is limited primarily because of the cost of acquisition, implementation and ongoing support of such packages. Additional packages will be acquired as usage requirements are identified and contracts are confirmed.

Access to the facility is supported by two high-bandwidth network initiatives, one local to Calgary (OilNet), the other a western university initiative (WNet), which in turn is connected to the cross-Canada CANARIE test network. These networks are described in more detail in Section 3.3. Commercial users outside Calgary, who do not have access to these broadband test networks, can access the facility via Internet, though speeds are considered to be too slow for many applications.

Security of access has been shown as an essential requirement for industry-based users, the primary target client group for the HPCC; however it is interesting that some academic users comment that the extra security procedures are an annoyance. Industry users often expect their data and/or their analyses to be protected, as many are concerned with "beating" their competition to the marketplace or providing a service at a more competitive cost. Extensive security features have therefore been implemented.

As a commercial service, the HPCC maintains an active marketing staff and has established a price structure that has been developed to reflect the revenue necessary to maintain operations. The HPCC is attractive to its commercial and institutional users, yet is competitive with commercial centres that offer similar services in the U.S.

### 3.4.2 Atmospheric Environment Service

The AES, though covered in more detail above, currently provides limited amounts of computer time to external clients on its large vector supercomputer. With the addition of a second system, the AES appears ready to offer expanded facilities and services. Current users have expressed concern over limited support services and recognize that the primary function of

Canadian weather forecasting takes precedence over other user requirements. A charging algorithm is currently under development.

Funding for this centre, the largest in Canada, is provided by the federal government. Access to the centre is primarily provided by Econet, a wide area network (WAN) operated by the federal Department of the Environment and by Internet.

### 3.4.3 MPP Systems

Some public access to MPP systems is currently offered by the University of Toronto and by Research Computing British Columbia (RCBC), which is a division of the British Columbia Systems Corporation (BCSC), as explained above. The University of Toronto has an in-house base of more than 200 potential users, representing 61 principal investigators. One researcher at the University of Calgary claims that he was the largest external user, for his telecommunications network simulations. Based on numbers of users, public access represents an extremely small part of the RCBC activity, probably less than 1%, and is provided without formal support services. Active marketing is not carried out. RCBC is using its system for some research applications but primarily in support of major commercial database applications in the health field.

In January 1995, the Université de Sherbrooke will have inaugurated its Centre de Applications de Calcul Parallèle de l'Université de Sherbrooke (CACPUS). CACPUS will offer services to users on its 16-processor IBM SP2 MPP system using a standard information highway connection. IBM has designated the Centre as a Canadian Centre of Excellence in High-Performance Computing.

### 3.4.4 Multi-Processor Servers

Dalhousie University in Nova Scotia installed a multi-processor server system in 1991. This system was unfortunately supplied by a vendor whose business operations ceased shortly after the installation date. However, the system continues to provide service to some 40 users within Dalhousie and a few from outside the University.

Funding for the system came partially from the government of Nova Scotia and partially from the federal government (Atlantic Canada Opportunities Agency). In establishing the Centre, there was collaboration with a commercial interest that was set up to provide experience and training for industry in Nova Scotia.

### 3.4.5 (New) Ontario High Performance Computing Facility (OHPCF)

With funding by the Ontario Government, the now-defunct Ontario Centre for Large Scale Computation (OCLSC) offered services on a vector supercomputer until 1992. The Centre was located at the University of Toronto.

The Ontario High Performance Computing Facility is a new initiative of the Ontario Government to which a commitment of $29.4 million over five years has been made. The funding for the project was approved in principle by the Cabinet of the Ontario Government in 1992, subject to Treasury Board approval of a detailed business plan.

In January of 1993, the Ministry of Education and Training set up an advisory committee to assist in the development of a business plan. The Committee was chaired by the Ministry with six representatives form the universities, four from industry and four from government.

The Advisory Committee developed a reference plan, which envisioned a facility as an interactive combination of a core facility linked by high-speed networks to local nodes dispersed throughout the province.

The core facility would consist of an appropriate blend of high-performance computing power, high-end visualization capabilities and networking linkages. The local nodes would have network access and visualization equipment allowing for support of and participation from local Ontario industry end-users through access to the full capability of the core facility.

The reference plan proposed that the facility be established as a not-for-profit corporation, independent of government with a board of directors drawn from universities and industry. An interim board of directors has been established with five industry, five academic and two government representatives. The board has developed the following mission statement for the OHPCF:

> To promote and facilitate the effective application and use of networked high performance computing in industry, universities and government, and thereby to contribute to skills development, job creation, economic growth and the international competitiveness of Ontario's industries and research institutions.

The board will incorporate the OHPCF, confirm its name and revise its business plan. It is expected that the facility will be fully operational by the end of 1995. The OHPCF will focus on providing support to its members through partnerships in high-performance computing projects. One of the key mandates of the OHPCF will be to encourage the industrial application of high-performance computing, as well as to increase the number of small and medium-sized businesses engaged in its use.

### 3.4.6 U.S. Centres

With the lack of Canadian centres that are publicly accessible via cost-effective networks, many Canadian users of HPC turn to the very large centres in the U.S. for services. Research collaboration makes these centres accessible at minimal cost via Internet links. These cross-border links, which have traditionally been at a higher bandwidth than those in Canada, also make the use of these U.S. centres more attractive than remote centres in Canada.

Application software availability also makes these centres attractive to Canadian users. Two reasons stand out for this increased availability. First, most of these centres operate large vector supercomputers from Cray Research Inc.—the type of computer on which most of these packages were either developed or have already been implemented. Secondly, these centres have an extensive user base and guaranteed long-term funding to support that user base. Thus they can better afford the high cost of a large variety of software packages.

Support services to Canadian users are available only via a visit to the centre or by telephone. Furthermore, there have been reports that Canadian users have been cut off from access, even in the middle of critical research work, because of increased demands by U.S. users.

### 3.4.7 Centre de Recherche Informatique de Montréal (CRIM)

CRIM is a publicly accessible facility that is unique in its approach. It is one of Canada's major research and development centres in the field of advanced computer science technology. The Centre transfers its findings to users, enabling them to increase their competitive advantage and improve their contribution to economic development. While CRIM has several parallel computers that are accessible to partners, their primary objective is to use this equipment in collaborative efforts to advance the science of parallel computing.

One of seven priority research areas at CRIM is parallel architectures. The group responsible for this activity has forged alliances with industry and universities for precompetitive research in parallel computing architectures. The first large-scale project, started in 1992 and scheduled for completion in 1995, is the development of a portable parallel programming environment. The environment is portable in the sense that the user can develop and fine-tune an application on a workstation and then rapidly port it to a variety of parallel architectures. The components of the environment include a high-performance programming version of the standard "C" language, an advanced compiler technology, a parallel simulator and a performance analyzer. The approximate budget for this project is $9 million, with funding provided by CRIM, Industry Canada and three computer vendors.

The parallel group is also planning new projects to work more closely with industry and government users, starting with CRIM's own partners, in order to use CRIM's expertise and development tools to experiment with new applications for parallel technology. One such project, which is already proceeding, is the development of a parallel spatial analysis and interactive visualization system. This project will provide the necessary computational capabilities of parallel processing for the interpretation and imaging involved with geographical information systems (GIS), remote sensing systems and intelligent spatially based decision-support systems.

### 3.5 Canadian Trends in HPC Facilities

The general trend for HPC in Canada will undoubtedly follow the directions being set by the U.S., Europe and Japan—that is, the number of vector systems in operation will probably not grow, despite moderate increases in computer power. The number of workstations and entry-level HPC systems will increase dramatically as their price/performance ratio continues to improve rapidly. MPP systems will continue to be applied in increasing numbers to specialized applications and dedicated experiments, but also in shared user environments, particularly for commercial applications involving large databases and multimedia.

Canadian HPC facility operators were asked for their opinion as to what they saw as their future technological direction, with varied answers, as summarized below.

The AES large vector installation is being expanded, within the current contractual arrangement. The AES entered into a long-range contract with its supplier in recognition of the extensive effort and potential disruptions to the Canadian Weather Forecast system that could result from a technology change. The AES situation is likely typical of that of many current HPC users. The cost of redevelopment is prohibitive. More than 50 FTE (full-time employee equivalents) were required to make the last major technology change at the AES. The AES vector system program will continue until 1998; nevertheless, opportunities that the more advanced MPP systems may offer are being investigated.

AMOCO indicates it will continue with a medium-scale vector system also because of the investment in existing software. The firm recognizes that it can significantly increase its capacity with improved versions of the same technology for less cost than a major redevelopment of its applications to a new architecture.

Shell, on the other hand, has moved from vector systems to a more parallel environment. Shell sees cost as an advantage, though newer algorithms being employed can benefit from the parallel environment.

While these are just a few representative examples of world-wide trends, the consultants also investigated the interest, opportunity and perceived benefits of collaborative programs where technologies were shared by multiple users. Support for the concept appears universal, despite the identification of some "not insurmountable" concerns. In general, the opportunity to direct applications to facilities that are most suited to the application was considered a most desirable option to have.

One expressed concern was that provincial politics may inhibit participation in a national collaborative program.

As Canada evaluates a national HPC strategy, the following is a partial list of new opportunities that should be examined with respect to facilities and access.

i) Canada currently has in place very few large-scale HPC facilities. However, even the few that are available are not fully utilized. While academic researchers will make a good case that Canada is very impoverished with respect to adequate facilities for computational research, it still makes good sense to examine how to make better use of existing resources before investing in new facilities, especially with today's extremely restricted budgets. With the new broadband networks that are being established throughout the country, along with the CANARIE program to link the regional networks with a national backbone (see Section 3.7 for details), there is an excellent opportunity to develop a national "virtual machine room", known as a "metacentre" in the U.S. With this concept, any HPC user in Canada could theoretically have access to the most appropriate facility, regardless of its location, at the necessary high-speed communications rates. Certainly such a concept would have to be supplemented with an equivalent national training and support network. This concept could also be broadened to an international metacentre by high-speed Internet links, about which more will be said in subsequent sections of this report.

ii) With the dramatic drop in prices for higher-performance servers and parallel systems, it will become more economically viable for certain organizations to acquire and operate their own facilities. However, before this can happen, there is still a need to establish awareness, education and training programs. There will also be a need to couple these internal systems to a Canadian metacentre to provide access to the very large problems that cannot be run internally.

iii) Parallel computing and utilizing the unused capacity of existing workstation clusters will be attractive because of the interesting economic advantages, despite the difficult software challenges. Canada is already quite advanced in the basic research of this concept and therefore there is an interesting opportunity to transfer this knowledge and the required software to user organizations that need a cost-effective way to commence their own HPC strategies.

### 3.6 HPC Education and Outreach

The *Feasibility Study on High Performance Computing Facilities in Canada* (September 1992) by Elizabeth Pearce formally inquired of the 18 Canadian universities interviewed whether their undergraduate programs included courses on supercomputing and computational science. While the results of the 1991 interviews are perhaps a bit dated, the response would not likely be different if the same question were asked today. There are almost no such courses being offered. The education of users of HPC systems comes in post-graduate programs where individuals are, for all intents and purposes, tutored in their use by others who have already familiarized themselves with HPC. One of the impediments then and now is the lack of facilities on which to give students the essential, practical, hands-on experience. One can speculate that in today's course structures there is increased use of high-power workstations. Despite enhanced compatibility with their larger cousins, however, these systems do not provide the opportunity for students to learn how to deal with the increased level of complexity inherent in much higher-performance systems that incorporate features such as vector and parallel processing.

These educational concerns can only be addressed as HPC facilities and associated software become more readily available. These concerns will be dealt with by educators as awareness of HPC increases and as representatives of Canadian industry demand such qualifications of their new employees.

A wide variety of outreach programs have been conducted in conjunction with the few HPC service centres that have existed in Canada. Certainly the universities of Calgary and Toronto both had industry and academic outreach programs in place when they operated their large vector HPC facilities in the late 1980s and early 1990s. Dalhousie University also encouraged local industry in the early days following the installation of its HPC system in 1991. It would have to be said that for academic programs, outreach has been very successful, based on the hundreds of researchers who advanced their domains of science and published creative literature on their successes. The opposite is true of industrial programs, measured by the very few industrial clients that have ever used HPC facilities.

The new HPC Centre in Calgary is taking a new approach to industry outreach in two fundamental respects. First, the Centre was established as a private company, with the facilities managed by an independent professional management company, as compared to previous centres that were managed by universities, as described above. Secondly, the HPC Centre has established an industry program in collaboration with the IRAP group at NRC, whereby NRC and the HPC Centre will work as a team to educate and train people in industry across Canada on the benefits of high-performance computing. (See the description of the U.S. industrial partnership programs in Section 4.1 for examples of highly successful industry outreach programs).

Two other current Canadian programs stand out today: the supercomputing symposia being held annually under the sponsorship of Super*Can, the Canadian Association for High Performance Computing and Communications, and the scholarship program co-sponsored by the Calgary HPC Centre and Fujitsu Systems Business of Canada.

The Canadian annual supercomputing conference has been renamed this year as the High Performance Computing Symposium 95. Though sponsorship and encouragement is under the auspices of the Super*Can Association, it is really driven by the organization that is chosen each year to manage the annual event and by the vendor and user organizations that take an active role. Conferences have been held since 1988 in Calgary (twice), Edmonton, Toronto (twice), Montreal and Fredericton. The 1995 conference is being hosted by CRIM in Montreal.

The HPCC and Fujitsu Scholarship program is now in its second year and essentially offers free computing resources on the vector system at HPCC. Efficient use of these resources is expected from participants, but is not a primary goal. Applicants are graduate students from Canadian universities, but collaborations at the industrial and international levels are encouraged. Their submissions for the annual award program are judged on their scientific merit, feasibility and reasonableness of their goals. An annual review of results identifies which researcher produced the most significant scientific results. The winner is awarded a 12-month allocation of additional system time.

## 3.7 Networks (Information Highway)

### 3.7.1 National Networks

#### 3.7.1.1 Summary

HPC users often require a connection to networks external to their premises for two principal reasons. First, the users sometime require access to external computer facilities that offer much higher performance than what is available within their building. Secondly, researchers and engineers often work collaboratively on large projects, requiring the transfer of data between multiple sites. As the data volumes associated with HPC are normally exceedingly large (it is not uncommon to process billions of bytes of data in a single job) and since there is sometimes a requirement to work interactively with a simulation, these external networks must have very high bandwidth if they are to be effective.

Networks are normally classified in three categories, as follows:

LANs: local area networks, which are located on the premises;

MANs: metropolitan area networks, which connect users within a city;

WANs: wide area networks, which extend provincially, as well as nationally and internationally.

#### 3.7.1.2 CA*Net (Production Network)

CA*net and the regional networks that connect to it constitute the major Canadian component of the world-wide Internet. CA*net is a not for profit organization that provides production network services to hundreds of organizations in all ten provinces, and recently Whitehorse and Yellowknife in the Northern Territories as well. However, CA*net only provides the backbone that connects each of the provincial networks together and the links into the U.S. for connectivity to the Internet. The provincial network operators are responsible for extending the network services to individual sites. Each of these provincial network organizations appoints one representative to the CA*net board of directors.

CA*net began several years ago as a modest effort supported by the National Research Council to provide a network facility that met the special requirements of research and education across Canada, including access to HPC facilities. In recent years, the network has experienced the same phenomenal growth in traffic that has fuelled the expansion of Internet throughout the world. As the Canadian sponsors and users of research and educational networks realized that the CA*net community was one of the engines driving the demand for development of the information highway, the interested parties launched the CANARIE initiative (described below), which incorporated CA*net as the production Internet network component.

The major criterion for being an authorized user of CA*net is that access must be restricted to activities related to research, education or technology transfer. Private companies are welcome, but only when they are engaged in one of these activities. Otherwise, users must resort to more expensive alternatives, i.e. a commercial Internet service, private leased lines, public common carrier network services or private networks. As almost all HPC activity involves research and development, most users are able to use the CA*net backbone.

In CANARIE's *1994 Three-Year Plan—A Statement of Direction*, there is a firm recommendation for government financial support of CA*net as an engine of network growth and application development in Canada, without which it is predicted that the network will shut down because of the excessive costs of upgrades needed to keep pace with growth in demand.

The CANARIE document goes on to explain that the basic dilemma facing CA*net is that tariffs for high bandwidth (i.e. T1 and up) connections in the U.S. are much less than in Canada, typically five to ten times lower in cost. Any attempt to maintain a cross-Canada backbone is constantly under threat from the alternative of simply linking each Canadian provincial network to the U.S. Internet (e.g. NSFNet). Most private Canadian Internet suppliers are already taking this approach and it is suggested that CA*net will be forced to go the same route unless there is a major overhaul of tariffs to bring them in line with real costs of new broadband service costs or unless ongoing financial support is provided to CA*net in the interim. The one user group that needs the highest possible bandwidth consists of HPC users, for reasons that will be explained in more detail below. The risk of not maintaining CA*Net at international levels is that not only will Canada lose its network service business to U.S. bypasses, but also the users will then make more use of U.S.-based HPC facilities in the process, leaving Canada with a crippled HPC infrastructure.

What is the current bandwidth capacity of CA*net? Until last year, CA*net had operated at 56 kilobits per second. Over two years ago, the U.S. upgraded from T1 (1.5 megabits per second) to T3 (45 megabits per second) for the national backbone (NSFNet). This represented an 800-fold increase of U.S. bandwidth over the Canadian 56-kilobit-per-second network. Today, the U.S. is experimenting in many locations with gigabit networks (one billion bits per second), which are about 20,000 times faster than a 56-kilobit-per-second network. However, some progress has been made in that CA*net has just completed upgrading most of its circuits to T1. Unfortunately, however, some of the provincial networks have made little progress in upgrading their facilities beyond 56 kilobits and some have no definite plans to do so. This means that higher T1 speeds only extend to one node in each province (at least for those provinces that have no T1 service) and all the feeder networks are still at very slow speeds. The President of CANARIE explained that there are no plans to upgrade to T3, as the common-carrier tariffs are completely out of reach. However, the CANARIE Phase 2 announcement of $80 million in additional government funding over the next four years lists further upgrades to CA*net as one of the four supported programs.

The projected cost of CA*net for the next three years (before any upgrades beyond the T1 level) is $20.9 million, with $4 million to be recovered by provincial network user fees, the balance being a CANARIE subsidy.

In summary, Canada has recently experienced an important upgrade to CA*net, but the situation is far from being on par with international programs to upgrade the Internet, especially for HPC users.

#### 3.7.1.3 CANARIE (Test Network)

CANARIE is the Canadian Network for the Advancement of Research, Industry and Education. It is a non-profit corporation established in March 1993. It evolved out of the collaborative efforts of 200 volunteer workers from 56 organizations representing Canada's research, university, business and government communities. This organization has ambitious plans to marshal a campaign to invest $1 billion over seven years in the research, development and eventual implementation of a fibre-based infrastructure for the information highway.

The CANARIE program so far encompasses three primary objectives:

- infrastructure—the continual upgrading of CA*net, as described above;
- testing and trials—the creation of a very high-speed testbed to enable Canadian organizations to test advanced networking technology and showcase Canadian products;
- applications—a grant program supporting product and service development to encourage the private sector to bring innovative network-related products and applications to market.

Analogous to CA*net, which is the backbone production network interconnecting regional production networks, the CANARIE high-speed testbed is the national backbone for the interconnection of regional networks dedicated to the development and testing of new broadband communications technology and applications. Since these facilities are test networks, they are not subject to regulated tariffs and in effect involve large subsidies by the common carriers.

Nevertheless, since they are not production networks, there are restrictions as to acceptable applications. From discussions with the management of the various regional test networks and CANARIE, it seems clear that many HPC applications could and should be tested on this facility. Many HPC applications absolutely require the high bandwidths of these test networks. What is required is a facilitator to bring together a group of interested potential participants in a collaborative effort.

From the brochure that has been circulated by CANARIE, there are many typical HPC applications that have been identified as target uses for the new communications infrastructure. Examples are airplane design (aerospace), access to remote sensing data (agriculture, fisheries, natural resources and forestry), access to supercomputers (oceanography, fisheries), high-resolution image transfer (health care), access to supercomputers (to model oil/gas reservoirs), development of new drugs (pharmaceuticals), mapping the human gene (biotechnology), and geographic information systems (urban planning). In most industrialized countries, these are applications that make extensive use of HPC technology and consequently require access to broadband networks.

What is discouraging in Canada is that, outside the mention of these applications in brochures, no champions have stepped forward to develop any HPC-based applications on the CANARIE trial networks. For example, of the 40 shared-cost technology and application-development projects for which funding has been granted to date, none had any HPC component (at least to the best of our knowledge). This is indeed strange when one observes that in the U.S. and Japan, the information highway development programs are embedded in the multi-billion-dollar high-performance computing and communications programs. Much of the information highway technology is developed in these HPC programs and many HPC applications are the first to be tested on the experimental networks.

In summary, HPC users must devise projects to take advantage of the very advanced testbed of CANARIE. CANARIE has almost completed a coast-to-coast linkage of regional ATM test networks, initially at acceptable T3 speeds, but with certain links already running at even higher speeds. These regional activities are indeed world-class, utilizing the most advanced new world standards for multimedia broadband communications, including ATM and SONET. These regional test networks are described in subsection 3.7.2.

#### 3.7.1.4 Other Network Alternatives

In Canada, the only current alternative to the Internet services described above is a service called frame relay offered by the common carriers and certain private network providers. Frame relay, unlike ATM, is a medium-performance service dedicated to data communications. Presently, the highest speed links are restricted to T1 service. Relative to the use of CA*net (which is also now a T1 backbone), the costs to the end user are very high. For example, to provide access from a Toronto user to a facility in Alberta, a cost of approximately $14,000 per month was quoted by one of the suppliers. Therefore, there is no reason for the research and education community to consider this service. However, as there are no restrictions on use of frame relay, it may be an option for certain HPC commercial users whose applications do not qualify for access to CA*net or the ATM trial networks.

Another possible future service would be satellite links. For example, the Communications Research Centre has established a facility in Ottawa to test and demonstrate applications over ATM fibre networks with interconnection to satellite links. This facility is already connected to the CANARIE ATM testbed via a direct connection to OCRInet, Ottawa's ATM test network. CRC, in cooperation with Telesat, is searching for information highway applications that may require a satellite link from the CANARIE ATM test network to some unserved locations. Their brochure specifically mentions interest in engineering and scientific applications such as the high-speed transfer of data and high-resolution imaging. It should be noted that this is a test network, like CANARIE itself, and consequently there are restrictions on its use. The WurcNet group in Western Canada has had conversations with the Telesat firm, which is a partner in the project, and there is a real interest to work collaboratively to identify HPC applications that could be tested over the satellite links.

### 3.7.2 Regional Networks

#### 3.7.2.1 Regional Production Networks

From the above descriptions of the national networks, it should be evident that it is the responsibility of the regional operators to deliver network services to the end users. Most of the regional networks are not-for-profit organizations that provide Internet services for activities involving research, education and technology transfer. They each connect their networks to CA*net for national interconnectivity and, in turn, CA*net connects to the world-wide Internet. Consequently, the total Internet service is only as good as the weakest link, which at this time consists of several provincial networks that have not yet upgraded from 56 kilobits to T1.

We asked regional operators to describe the major strengths and weaknesses of the present networks. The major strength seems to be the provincial networks' contribution to building a strong Canadian component of the international Internet, with all the benefits that this has bestowed upon the research and academic community. All operators were quick to explain that the major weakness is the exceptionally high cost of high-speed circuits (e.g. T1 and T3), especially relative to the U.S. rates. These high tariffs, with the associated difficulty of finding upgrade funds, has placed Canada at a severe disadvantage in terms of implementing new high-bandwidth applications, especially in the area of HPC.

It is beyond the scope of this study to analyze each of the 10 provincial networks and to review their in-depth plans. As a general statement, however, we can report that they all have intentions to upgrade to the CA*net backbone speed of T1, but to date, several provinces are having difficulty finding the funds for the costly upgrades.

As Ontario has discovered a way to raise the necessary upgrade funds, and as they have the largest provincial network and the best-defined upgrade plan, a description of their network, ONet, is provided as an example of one regional production network. The other regions will mostly likely follow a similar plan in time, but what is not clear is the extent to which they will obtain provincial government support in these difficult economic times of high government deficits.

*ONet*

ONet is the largest and most extensive of Canada's provincial research and educational networks. ONet's membership has grown from a few universities in 1988 to an association in 1993 of almost all post-secondary institutions, 20 federal and provincial ministries, and 12 private sector organizations, as well as a number of research centres and hospital research establishments. Very recently, this loose association of members formed the ONet Networking Corporation, with a membership of 108 organizations. The applications run on ONet are very much like those found throughout the Internet and HPC remote execution is very definitely one of these. In fact, the slow-speed 56-kilobit-per-second links have been a severe impediment to HPC access up to now (along with the limited availability of HPC facilities), so it will be interesting to see if the upgraded speeds to T1, even though they are still 28 times slower than the U.S. T3 Internet, will spur more interest in HPC in Ontario.

Until a few months ago, ONet had reached a saturation point, as the number of Internet users had escalated at dramatic rates. While ONet had always been a self-sustaining association, with all costs covered by user fees, it was not practical to increase fees for the major cost to upgrade to T1 service (T3 could not even be considered). However, at least T1 was critically required to handle the growing traffic and to accommodate the increasing demand for new high-bandwidth applications such as HPC access and remote visualization and imaging. The association felt that it was time to request government support for infrastructure costs in order to provide the necessary environment for research and extend equal access to all important regions of the province.

The government has responded positively in two ways. First, the Ministry of Education and Training has made contributions to the 18 universities towards the T1 upgrade. The major inter-city T1 links are already installed. Secondly, through a grant from the province's Ontario

Network Infrastructure Program (ONIP), the speed of the remaining inter-city links and the links to individual members from these city hubs is expected to be increased to T1 over the next two years, subject to individual members' motivation to contribute matching funds. According to the new director of ONet, the establishment of network-connected HPC facilities would provide significant motivation for some of those members.

For the 1993 submission to ONIP, a very detailed report was produced: *ONet, Designing for the Future.* The following is what the report had to say about HPC-type applications:

> The ONet network is being stretched to its technological limits. New computer applications that rely on interactive imaging, large volume data transfer and real-time computing cannot be attempted using the existing network. This is putting Ontario researchers and educators at a substantial competitive disadvantage relative to their counterparts in other countries. Increasingly, researchers in medicine, the physical sciences and technological disciplines are dependent on interactive access to huge databases of research information.
>
> Using an enhanced ONet, students and researchers anywhere in the province could be given access to very costly and expensive computing resources, similar to the earlier OCLSC (the Ontario Centre for Large Scale Computation, which provided access to a CRAY supercomputer until 1992). A very high-bandwidth network would ensure that a facility of this type could be used effectively from remote locations. This would make large investments in specialized resources an economically practical option.

Since this report was written, funds have been allocated for T1 upgrades. While still not adequate for some HPC applications, it is a major improvement.

#### 3.7.2.2 Regional Test Networks

As explained previously, CANARIE will link together regional advanced test networks to form a national testbed for developing new communications technologies and applications. Let us examine what these test networks will consists of and what they will provide for HPC.

In the last year, there have been very exciting projects announced in various regions of the country with respect to very high-speed broadband test networks, on a par with other similar projects throughout the world. All these test networks are utilizing ATM (asynchronous transfer mode) technology, the new international de facto standard for broadband multi-media communications. In fact, when CANARIE fulfils its commitment to link together all these initiatives, Canada will have the largest expanse of ATM test networks in the world. This is very good news for the HPC community, as ATM networks are adequate for all HPC applications in common use or in development today.

Once again, it is beyond the scope of this report to give a description of each initiative, but to give an appreciation of the benefits of these networks to HPC, a few comments are provided on WurcNet, the Western Canadian ATM test network. This example has been selected because WurcNet is concentrating on the use of high-performance networking (HPN) to support activities in HPC and vice versa.

*WurcNet Inc.*

Communication technology is particularly important in Western Canada where the population is sparse. Effective, affordable communication reduces the isolation of university, government and industrial researchers. In more densely populated regions, a critical mass of researchers working in the same area can develop spontaneously. In contrast, in Western Canada, high-performance computing and networking must be in place before effective collaboration can begin.

To that effect, a non-profit corporation called WurcNet Inc. was formed in December 1993. WurcNet Inc. is a collaboration of university researchers, the HPC High Performance Computing Centre (located in Calgary), telecommunications companies and suppliers, and other interested industry and government partners from Alberta, Saskatchewan and Manitoba.

WurcNet Inc. has two goals: to build a high-speed ATM test network for research and education in western Canada and to develop applications that will benefit most from broadband communications. This group is of great interest to the HPC community as it is the only network initiative that has grown out of a need to collaborate in the area of HPC, which is still one of its major focuses. In addition to the network focus described in this section, WurcNet has submitted a major proposal to NSERC to develop collaborative HPC research programs over the ATM network and sharing HPC facilities at the HPC Centre, clustered workstations at the University of Alberta and a parallel HPC facility at Research Computing British Columbia.

The ATM test network initially interconnected all the universities in the three provinces and the HPC Centre at 45 Mb/s. Most of the links have already been upgraded to 155 Mb/s, while the remaining upgrades are scheduled for early 1995. There are also plans to connect other research labs such as ARC and TR Labs, as well as the Banff Centre. Eventually, all of the major cities in these Western provinces would be connected to this ATM testbed network by a backbone which operates at 622 Mb/s. Initially it was intended that B.C. would be part of this network, but B.C. has its own testbed under development and therefore the connection to B.C. will be through the CANARIE backbone.

The 80 researchers participating in the WurcNet consortium have defined 45 HPC projects as part of the NSERC proposal. One of the major projects involves seismic processing and is expected to be a large consumer of HPC resources.

This seismic project will be conducted by the Consortium for Research in Elastic Wave Exploration Seismology (CREWES), sponsored by 25 major international oil, gas and seismic companies. The project team intends to develop techniques to remotely process and interpret 3-D seismic data. This process is extremely computationally intensive and consequently not every organization has access to the necessary computer power on its own premises; thus the requirement for processing over ATM networks. As the user interface will be a graphics workstation, it will be necessary to develop a use-friendly interface with the remote machines. Seismic acquisition, analysis and interpretation in 3-D is producing remarkably accurate and useful pictures of the subsurface to enhance the probability of discovering new harder-to-locate reserves. Developing new algorithms for 3-D analysis is an area of intense research and global industrial competition. Developing procedures for the remote processing of seismic data using workstations connected via high-speed links to a high-performance machine has major ramifications for how effectively the geophysical industry will operate in the future.

### 3.7.3 Metropolitan Area Networks

#### 3.7.3.1 Overview

Metropolitan area networks are facilities designed to extend the reach of data communications access to a city-wide area. The MANs that are expected to be established over the next few years will begin to provide the very high bandwidth required by HPC users within city boundaries. Extending the network beyond city boundaries is currently prohibitively expensive because of the cost of leasing high-speed circuits. This problem could be resolved by the evolution of regional and national broadband test networks into production networks, assuming the link charges are affordable and bear a reasonable relation to real costs.

#### 3.7.3.2 Example of a Production Network: OilNet

An example of the potential of the MAN networks is OilNet in Calgary. The original requirement for this network came from a group of supercomputer users at the High Performance Computing Centre in Calgary. This centre is now operational and is serving users from universities and industry. As this is Canada's only publicly accessible supercomputer, it is a national resource that should be shared by users across the country, thus requiring high-bandwidth networks on a national scale. However the petroleum/seismic and university user community in Calgary provides a good base of HPC users that can be served economically by an ATM MAN.

One of the principal needs of the HPC Centre is to provide high-speed data communications for massive file transfers and visualization. In Canada, such high-speed communications have been exceedingly expensive and difficult to operate. Fujitsu, the supplier of the supercomputer, decided to implement an ATM-based solution to link Calgary users to the supercomputer over fibre-optic lines.

Subscribers to OilNet install a small ATM switch in their offices. These ATM switches connect the customer host computers (e.g. servers and workstations equipped with ATM adapter cards) to the high-speed network. These switches are connected to OilNet, initially at DS-3 (44.5 Mbs) rates, over a fibre-optic backbone that interconnects all the subscribers in Calgary. The DS-3 customer links are upgradeable to OC-3 circuits (155 Mbs); higher speeds are expected to become available in the near future. The fibre links are leased from the City of Calgary by Fujitsu. This fibre backbone uses "*SONET*" (synchronous optical network) add-drop multiplexors to enable a single fibre pair to transport multiple customer circuits simultaneously, thereby reducing the cost to operate the network, resulting in lower rates for end users.

To the end user, these networks provide not only unprecedented bandwidth, but equally important, they guarantee high throughput from the customer workstation or server to a final destination point as a result of the highly efficient use of bandwidth and the low latency of the ATM technology. This will enable users in Calgary to extend their LANs to connect to those of other organizations at the maximum speeds that their host processors can deliver. This means that the network will no longer be the limiting factor and no longer will it be necessary to physically deliver tapes to remote sites for data transfer. It also opens the door to new applications that require remote visualization.

### 3.7.4 Applications Requiring High Bandwidth

There are two principal reasons why high-bandwidth communications are often required today by HPC applications: massive data volumes must often be transmitted to and from the HPC processor and interactive visualization is necessary whenever the simulation process must be monitored and directed during the model execution. If one considers that HPC systems will also be utilized as information servers for access to massive databases, then the transmission of high numbers of simultaneous transaction volumes will become a third requirement for higher-speed communications, especially as more of the content becomes multimedia. The concept of information servers is explained in more detail later in this report.

Massive data transfers are becoming more common as the precision of models is increased to exploit the increased performance capabilities of the processing systems. Increased precision means finer "grid sizes" and more data elements. To put the data volumes into perspective, let us examine a typical 3-D marine seismic survey for off-shore hydrocarbon exploration, covering an area of 100 square kilometers. In this example, survey boats will fire airguns and then record reflections consisting of 1.92 billion channels (traces) with each channel containing 3,000 time samples, at a recording rate of 25 million bytes of data approximately every nine seconds. This data is then processed on a high-performance supercomputer and requires an estimated 6.5 quadrillion ($6.5 \times 10^{15}$) calculations. This would require several years of processing time, 24 hours per day, on any traditional computer or workstation. On the largest supercomputers available today, it would require about one month, 24 hours per day (which is reasonable, considering that about eight months are required for the survey itself). Some seismic processors are beginning to run jobs of this type remotely over a network, as it is not every company that can justify an in-house HPC system with sufficient capacity to process such jobs. When jobs become this large, the only type of network that has proved to be adequate is an ATM circuit with at least T3 speed.

Many other HPC applications encountered in Canada are experiencing similar explosions in data volumes, necessitating access to broadband communications whenever the HPC system is not in the same building as the users. Other examples are environmental and climate modelling, remote sensing and image processing (Radarsat, scheduled for a 1995 launch, will have a major impact on the data volumes available for HPC processing).

The second category of HPC job requiring broadband networks are those where visualization is required remotely to control a simulation in process. Such a method can be very important, as it enables a user to "steer" a problem to an adequate solution in one interactive run, which can be far less time consuming and less expensive than the traditional method of running a series of successive runs, each time trying a new combination of input parameters. This approach is not feasible on a remote system unless the network can transfer image data rapidly enough that the interacting human is not frustrated by excessive response times. Sometimes the requirements are even greater when the visualization involves animation, i.e. refreshing a screen several times per second. This approach is only now being considered in Canada, as the networks have not been in place until very recently to accommodate the image data transfer. To quantify this situation, consider that a typical colour screen contains approximately 24 megabits of pixel data. To transmit dynamic displays would require at minimum a T3 circuit, and even then an animation would have screen refreshing only two or three times per second.

### 3.7.5 New Network Opportunities

The rapid development of high-speed multi-media networks opens the doors to many new HPC opportunities in Canada.

First, these networks will provide almost-universal access to HPC facilities at the required bandwidth for users across Canada. This makes resource sharing and the metacentre concept a real possibility. As discussed in previous sections, the new networks should enable Canadians to share the limited facilities that are already in place and reduce investment costs for new facilities.

The CANARIE and regional network initiatives are encouraging pilot projects and application demonstrations access to the networks at no charge. HPC users can take advantage of this opportunity, and in fact a few HPC projects are now beginning. For example, the University of Calgary has implemented an on-line visualization project connected to a remote supercomputer for the numerical modelling of chemically reacting flows over an ATM network.

The high-speed networks offer the possibility of new applications involving remote data collection of massive data files. For example, the seismic industry is interested in transferring data directly from the acquisition sites to a remote HPC processing facility.

Telesat offers the possibility of remote access to HPC facilities from areas that are not served from fibre lines on the ground. Research groups in Newfoundland have expressed a strong interest in this possibility.

Lastly, Canada should give serious consideration to information highway applications that could benefit from HPC technology. For example, information database servers using MPP systems could be considered for highly cost-effective delivery of massive amounts of government information.

## 3.8 Canadian HPC Applications

### 3.8.1 Industrial and Government Applications

Presented below a few examples of applications from government, industry and universities. The descriptions are drawn from trade journals, vendor publications, and user profiles provided by their respective companies. More than 50 survey forms specifically addressing applications of HPC&C were distributed to these and other organizations, and the following descriptions were augmented and updated where appropriate.

Vendor publications from which description of Canadian applications were drawn included *IBM Visions* and *Cray Channels*. As these publications are basically marketing tools, some minor editing was done to de-emphasize sales rhetoric and scientific jargon. Other applications were drawn from the Public Works and Government Services Canada publication *Research and Development Bulletin (Science and Technology)*. The most up-to-date information was drawn from the electronic publication *HPCwire*, which is distributed weekly by electronic mail and has become the primary source of information for the high-performance computing and communications community world-wide.

In this section, applications from industry are emphasized, while a few examples from the academic sector are provided in the following section. Not all applications currently run on supercomputers. Some applications included here illustrate the migration from 1980s-vintage single-processor mainframe supercomputers to modern single-processor RISC microprocessor workstations. Some applications included here have never been run on supercomputers, but nevertheless use advanced numerical techniques employing application-specific processors (such as digital signal processors) that have the equivalent processing power of some mainframe supercomputers on a board that can be slotted into an everyday microcomputer.

From the Canadian examples that follow this brief overview, it can be observed that Canadians are generally active in many of the areas that their U.S. counterparts label as "grand challenges" but, in Canada, the availability of resources significantly limits the range or complexity of problems that Canadian industry, government, or academic researchers can study. In the words of one respondent to this study, "If Canada cannot afford such facilities, researchers cannot keep up with colleagues in the U.S. and Europe who have access to HPC and, with orders of magnitude less computational power, are forced to work on a very different scale far from the frontiers of their field."

*High-Resolution Short-Range Weather Prediction*

The Canadian Meteorological Centre (CMC) of the Atmospheric Environment Service (AES), located in Dorval, operates Canada's Numerical Weather Prediction (NWP) program which provides the atmospheric and other environmental information.

The computer-based numerical weather-prediction models are designed to simulate the behaviour of the atmosphere. These models, however, are only one of a series of applications that run within the CMC operational suite of programs. This suite consists of four steps: data acquisition, computer analysis of data, computer forecasts, and interpretation and dissemination. The data are received from across Canada and around the world continuously, and as a result of the computer analysis are then used as initial data by a numerical model of the atmosphere to produce forecasts of its future state.

The CMC approach is to perform an analysis for North America as soon as the observations are available and to produce short-range high-resolution regional (North American) forecasts. The medium-range global forecast is conducted in independent steps. The short-term model must be carried out at a rate that results in the production of the forecast states of the atmosphere before they become manifest in the real atmosphere. Thus, this model is a time-critical application.

These models result from many years of research and development aimed at producing highly accurate forecasts over limited regions of the earth's surface.

The twice per day execution is a major HPC activity on the very large-scale vector supercomputer. Essential to the interpretation, however, is visualization graphics, another HPC-based activity that represents a load about equal to that used by the model.

*Coupled Global Climate and Ocean Modelling*

The Canadian Climate Centre (CCC) represents one of the three major users of the operational weather forecasting/climate supercomputer at the Canadian Meteorological Centre (CMC) in Dorval, which is leased by the Atmospheric Environment Service (AES). One particular area of expertise at the CCC involves climate modelling and diagnostic studies; for example, the CCC's Global Climate Model (GCM) is the primary vehicle by which Environment Canada investigates climate-change scenarios associated with global warming.

Environment Canada issued a warning in late 1991 that global warming, brought on by excessive production of carbon dioxide and other human-made pollutants, could cost the

Canadian economy billions of dollars by increasing Prairie droughts, coastal flooding, severe storms, and forest fires. It also warned of a major health impact.

Most climate experts world-wide agree that the earth is gradually warming because of the "greenhouse effect", caused by increased carbon dioxide released into the atmosphere. Even with clear agreement on global warming, there is still a long way to go before we will know the answers to critical questions such as which regions will have less rainfall, whether there will be increased melting of the arctic glaciers, which regions will suffer from flooding, and how hydroelectric reservoirs will be affected. These questions cannot be answered at the moment because computer models are not accurate enough to take into consideration all of the physical influencing factors such as cloud formations and interaction of the oceans. Even if the models are developed with sufficient realism, today's supercomputers are still not fast enough to process such models with the necessary level of accuracy.

It is imperative that we determine what the global ecosystem's response will be to global warming. The federal government had committed significant resources to support climate-change research in Canada under the Green Plan Global Warming Science Program. The program was intended to challenge the Canadian scientific community to reduce the uncertainties associated with the rate, magnitude, and regional characteristics of global warming in Canada and thus provide a sound basis for policy and decision making. The funds originally allocated from the Green Plan to support the modelling of the greenhouse effect have been withdrawn and consequently the funding for a new Climate Centre in Victoria is still in question.

The CCC is currently integrating a Global Ocean Model (GOM) with the CCC's Global Climate Model, significantly increasing the complexity of the models and their demand for HPC. The coupling of a GOM with a GCM is universally acknowledged as a grand challenge. Canada's Climate Centre has been a world leader in this area.

*Three-Dimensional Turbine Flows*

Pratt & Whitney Canada and Concordia University of Montreal have collaborated for many years on a finite-element computational fluid dynamics (CFD) methodology for the analysis and design of gas turbines. Recently, the team developed a code to analyze 3-D viscous flows in complex internal flow passages. The Pratt & Whitney/Concordia collaboration has developed an award-winning application that studies the complex, turbulent, flow through a gas turbine diffuser. Such studies are crucial to improving the aerodynamic design of gas turbines.

The method developed by the team uses a direct approach that was previously considered prohibitively expensive for 3-D flows, in terms of execution time and memory. This approach became feasible, however, on a supercomputer, achieving almost 90% of the peak performance. This highly efficient code provides Pratt & Whitney with the capability of testing and designing gas turbine diffusers by means of numerical models, with potential savings of hundreds of thousands of dollars per design.

*Airframe Design*

For several years, Canadair has been using an in-house CONVEX "mini-supercomputer" to solve problems in computational fluid dynamics and structural analysis related to airframe and component design. The firm's HPC needs have grown to the point where they plan to replace the current technology with a small CRAY supercomputer for CFD and a Silicon Graphics SMP multiprocessor for structural analysis. Also, Canadair occasionally used to use remote computers, for example Boeing's supercomputers, when runs were too large to run in-house. However, a Canadair representative made the comment that the aerospace industry will normally want to do all or most of its processing on in-house facilities, as it is critical for these companies to control their own tight design schedules and to maintain very high levels of security.

Canadair makes effective use of HPC facilities to refine aircraft design, by enabling the engineers to cycle through far more design iterations and by simulating a wider variety of parameters. This represents significant value added in a very competitive industry. Using HPC models, Canadair is migrating from a typical "design-build-test-redesign-rebuild" cycle to a "numerical simulation-build the final product" scenario. Another area where HPC capability will have an increased importance is in computer-aided design and manufacturing (CAD/CAM) applications. Canadair has a goal to go from a computer screen design to a manufactured product in a two-day cycle for smaller components.

An executive of Canadair explained that the most important potential benefit of applying HPC to the aerospace industry would be to allow for "integrated product design (IPD)", i.e. integrating all the design aspects for an entire aircraft. If such a process were automated, it would then be possible to perform design iterations and arrive at an optimum design. The benefits of such a process would be dramatically reduced design costs, reduced risk by avoiding costly redesigns late in the cycle, less costly product primarily because of reduced weight, reduced aircraft fuel consumption, and getting product to market sooner through shorter design cycles, all of which translates into increased competitiveness.

The major problem in the implementation of IPD is not a lack of HPC facilities; it is lack of expertise and software. While major U.S. aerospace companies such as Boeing are already investing heavily in these methods, a relatively small company like Canadair finds it too difficult to justify such an investment. However, if Canadair—Canada's largest aerospace company—cannot justify the investment, it is unlikely that any other Canadian company can. Over time, Canadair may be able to pick up U.S. codes, which will provide the software at a cost that is significantly lower than if the software is developed totally in-house.

The installed and on-order equipment is adequate for Canadair's current method of design, but would be totally inadequate for the IPD and optimization process described above. If Canadair were to begin using such codes, access to a large supercomputer would be required, and because of the cost and the likely sporadic need for large-scale supercomputing, it would only make sense to access a shared facility, provided there is adequate security.

*Hypersonic Flight Vehicles*

The development of hypersonic commercial aircraft (i.e. aircraft that flies at speeds greater than five times the speed of sound) promises to be of great benefit to commercial aviation. Research efforts in various countries are aimed at designing and engineering such aircraft. The evaluation of hypersonic designs requires testing capabilities beyond those provided by wind tunnels. Such design work is an exacting technical challenge being met largely through the use of computational models and supercomputers.

Researchers at the University of Toronto's Institute for Aerospace Studies used the University's former CRAY X-MP system to investigate shock-induced combustion ramjet engines as potential propulsion systems for hypersonic flight vehicles. Shock-induced combustion is not as well understood as diffusive combustion, which is an alternative and more easily implemented type of combustion, but it offers some compensating advantages. During shock-induced combustion, the temperature and pressure of a combustible mixture are raised sufficiently, by means of a shock, to initiate combustion. This type of combustion results in a very short combustion zone and reduced thermal loads within the combustor.

*Fabrication of Aluminum Products at Alcan*

At the heart of Alcan International Ltd.'s business is aluminum smeltering, which involves investment costs in the order of a billion dollars and substantial maintenance costs. Models have been developed to better understand the complex processes of alumina production, such as alumina precipitation from the caustic liquor; assist the cathode design process to ensure longer life and versatility of these expensive components; improve the design of electrolysis cells through better understanding of wave instabilities; design more efficient remelt/casting furnaces; etc. Applications for supercomputers also exist in the fabrication of aluminum products—for example, in simulating processes such as casting, extrusion, rolling, forming aluminum car design, etc.

As one of the world's largest manufacturers of aluminum, Alcan has a vested interest in aiding automotive companies in introducing aluminum-stamped components into their products. Automotive die designers and stamping plants are relatively unfamiliar with the behaviour of aluminum, which reacts differently from steel in terms of stamping and springback. Through the application of metal forming simulations and Alcan's knowledge of its aluminum material, the firms is working with the automotive industry to overcome the initial perceived difficulties in introducing a new material to a mature, experience-based industry.

Alcan sees a great deal of promise in metal-forming applications for high-end massively parallel processing systems (256 processors or higher). Metal-forming problems are ideal candidates for new MPP technology. The tooling or stamping equipment contacts the aluminum sheet at multiple points and is inherently parallel. Each processor of an MPP system can be given a stream of discrete elements on which to work for fast, efficient, problem solving. The competitive advantage for Alcan is that, depending upon the number of processors allocated, a large number of designs could be evaluated in a single working day.

*Computer-Aided Rational Drug Design*

The Queen's University Quantum Pharmacology and Medicinal Chemistry Group in Kingston, Ontario, is using computers to research an important medical problem: epilepsy. Characterized by recurrent seizures during which a person's brain undergoes a burst of uncontrolled electrical activity, epilepsy is second only to stroke as the most common neurological disorder. Only 65% of sufferers experience reasonable control with the available anticonvulsant drugs. Moreover, side-effects range from cosmetic problems to severe blood or liver toxicity.

The past decade has seen rapid growth in "rational drug design"—the development of new drugs based on understanding a given disease at its molecular level. Extending this to computer-assisted rational drug design means using sophisticated mathematical and graphical computer techniques in order to understand the shapes and properties of new drug molecules.

The Queen's group is working on all these challenges using large-scale computer-assisted molecular simulations. First, elaborate prolonged computer simulations were used to create a computational model of the sodium channel protein, the idea being to devise prototype drugs that inhibit the spread of a seizure by blocking the sodium channel. Second, molecular dynamics simulations allow for the evaluation of portions of the calcium channel protein which may bind certain anti-seizure drugs, thereby blocking their function. Third, quantum mechanics calculations enable researchers to model neuro-transmitter receptor sites. Chemical messengers such as the neurotransmitter glutamic acid cause excessive brain excitement, which can trigger seizure.

Advances in computational power are linking advanced mathematical techniques to medical problems on a broad front. Bringing a single drug to market requires approximately 10 years of R&D effort and many tens of millions of dollars. Computer-assisted drug design is just beginning to be used by the world's leading-edge pharmaceutical manufacturers to eliminate years of product development in producing tomorrow's drugs.

*Computer-Assisted Design and Manufacturing*

Exco Engineering of Newmarket, Ontario, started in the 1950s as a machine shop and extrusion die manufacturer. Exco Engineering began using computers in manufacturing in the early 1980s. It took almost four years before the company was able to exploit computer technology in a meaningful way. The years paid off when Exco evolved into building transmission cases for automobile manufacturers. A typical transmission case has in excess of 1,500 discrete surfaces, all of which must be machined in three dimensions, making these housings the most complex parts the automobile industry requires. Exco currently uses multiple high-performance workstations. Given the company's work environment, it would be very useful if all these workstations could be run in parallel as a cluster whenever a large design is required. However, clustering will not be a possibility until Exco's software suppliers provide parallel versions of the codes, as the firm does not develop its own software.

*Computer-Aided Industrial Design and Film Animation*

Among the fastest growing fields of high-performance computing, with special emphasis on visualization, is entertainment. Toronto-based Alias Research is a leader in this youthful industry.

Alias has a world-leading market share in computer-aided industrial design (CAID), where its software is used in "front-end" creative aspects of product design. Alias focuses on aesthetics and ergonomics. To quote Sony, it empowers customers to "use style as a competitive weapon." Alias pulls together the style and design of a car from a design team's first thoughts to final iteration—lines, colours, dashboard, upholstery, and road views from the driver's seat—in short, the functional beauty of the package as a whole. Only then do engineers turn the vision into a product with CAD/CAM.

Software that enables a product's creator to model shapes in every light and colour is but one step away from the immense entertainment and digital media animation market. Alias enjoys strategic agreements with several leading game developers.

Alias recently announced an agreement with the digital production studio Digital Domain. The alliance has a clear focus: to create a screen environment that lets video-game developers and film makers direct the creation of digital imagery as easily as they currently direct live action production.

In its early years, Alias Research was a value-added reseller of computer systems. This line of business was phased out in favour of its core strengths in developing and enhancing its computer graphics software, which runs on a variety of vendors' computer systems. The U.S.-based Alias Computer Graphics is an industrial affiliate of the San Diego Supercomputer Center.

Just before this report was completed, Silicon Graphics, Inc. (world leader in high-performance visualization workstations and servers), Alias Research, Inc. and Wavefront Technologies, Inc. announced that they had entered into definitive merger agreements. As a result of the merger, Silicon Graphics will form a wholly owned, independent software subsidiary that will focus on developing the world's most advanced tools for the creation of digital content.

"With the creation of this new subsidiary, Silicon Graphics extends its reach in defining the standard for visual realism and interactivity in the world's most demanding computing environments: entertainment and industrial design", said Edward R. McCracken, Chairman and CEO of Silicon Graphics. "Leading-edge products from Alias and Wavefront are unique and immensely valuable. Embracing these assets will enable us to design the power tools that artists and innovators need to create the digital studio of the 21st century."

*Non-Destructive Evaluation*

Barrodale Computing Services Ltd. (BCS), based in B.C., has developed a system that uses ultrasound and advanced digital processing to detect damage, for example, in the laminated composite materials of a state-of-the-art fighter-plane wings before the wings experience catastrophic failure.

The company's solution involves non-destructive evaluation (NDE), which involves ultrasound. Pulses are beamed at a point on the surface of a target and the traces of their reflected waveforms are recorded as digital data. As a transducer moves over a surface, images can be created based on internal reflections. Tell-tale signs reveal internal defects in this way.

A new visualization tool called DETECT/NDE processes and visualizes full 3-D data sets, eliminating the need for more than one pass.

Just as important as time-saving 3-D visualization is the high resolution of the features that DETECT/NDE's analysis reveals. Two techniques achieve this: deconvolution compresses ultrasonic pulses into spikes showing amplitudes and positions of corresponding reflections; and inversion screens out multiple reflection coefficients for interfaces in the sample material. This simplifies the interpretation of large volumes of data.

The Canadian Department of National Defense has developed techniques for inspecting critical components on CF-18 fighters using DETECT/NDE. Some units returning from the Gulf War, for example, were subjected to this analysis. British Aerospace and Boeing are other customers. Less dramatic, but no less important, is DETECT/NDE's ability to analyze flaws in girders, bridge support beams and trusses, high-pressure pipes, vessel hulls, and many other situations where materials come under stress.

*Full Flight Simulators*

CAE Electronics Ltd. of Montreal is the world leader in the manufacturing of full-flight simulator technology. In a typical CAE flight simulator, data on up to 2,000 variables flows continuously through a computationally intensive Fortran program created by CAE. The computer processes incoming data and directs real-time responses to a motion system, visual and auditory subsystems, and every instrument in the cockpit. Hydraulics power the CAE simulator's six degrees of freedom motion system, giving the pilot the sensations he/she expects during takeoff, landing and, for example, air turbulence. Meanwhile, a visual system generates a panoramic display of up to 220 degrees, creating computer-drawn images of runways, rainfall, lightning, landing lights, the curvature of the earth's horizon, traffic hazards and a host of other features.

Airplanes themselves are becoming increasingly controlled by computers, with conventional-dials being replaced with computer screens (referred to as "glass" cockpits). Management of these increasingly sophisticated systems requires pilots to undergo continuous upgrading of their skills on simulators.

Since 1991, CAE's standard offering has incorporated high-performance computers. A second system serves as the instructor's workstation, allowing the instructor to monitor, record, respond to and replay each decision taken by personnel in training.

*Optimization*

The art and science of optimization seeks to maximize profit, minimize cost, create the most profitable array of goods with the resources at hand, use scarce resources to the best effect with the least waste, resolve complex scheduling demands (for flight crews, trucks or locomotives, for example), select investment portfolios within given constraints of risk and return, define the most logical switching pathways for communication networks, and determine manufacturing decisions.

TEE Consulting of Oakville, Ontario, is a consulting company that specializes in optimization. The firm has consulted with the Ottawa River Regulation Planning Board on the optimization of water levels in the Ottawa River Basin to balance tradeoffs in hydroelectric power output and environmental impact concerns due to flooding. A host of constraints and demands include navigation, pollution concerns and municipal water requirements.

In business, scientific and commercial high-performance computing are rapidly converging. Scientists studying order, permutations and change in physical systems know that the mathematical techniques they use for problem solving in pure science can also be brought to bear on complex business decisions.

Hilary Markowitz won the Nobel Prize (jointly) for devising a technique of portfolio risk analysis. Markowitz made it possible to analyze the risk of owning an individual stock by correlating with many others, using a co-variance matrix. The Standard and Poor's 500, for instance, demands very large co-variance matrices, followed by a quadratic program solution for a range of values for assumed risk. The resulting analysis plots reward against risk for any selected percentile of return.

There has been recent anecdotal evidence in employment studies that indicate that significant numbers of computationally literate Ph.D.s in the mathematical and physical sciences are being employed by Wall Street firms to apply their mathematical and computational expertise in the area of financial analysis. Many of these firms adopted massively parallel systems early on. In contrast, there has been no apparent move by Bay Street firms to explore the promise of high-performance computing to financial applications such as portfolio analysis.

*Micro-Climate Modelling and CFD*

Rowan Williams Davies & Irwin Inc. (RWDI) is a firm of consulting engineers based in Guelph, Ontario. The firm specializes in predicting and, if necessary, mitigating the effects of wind and snow on all kinds of facilities. In particular, the firm has carried out microclimate studies for stadiums as far afield as Malaysia and Hong Kong.

RWDI's modelling of wind, snow, and air-pollutant transport has historically been associated with wind-tunnel simulations using scale models. However, computer models are now an important and complementary tool in the firm's work.

For example, when RWDI was asked to predict snow-drift loads on the roof of the SkyDome in Toronto, the firm developed a computer modelling tool to complement its wind-tunnel and water-flume tests. The company used its own wind tunnel and water flume to measure average wind velocities close to the roof from several directions; snow depths and drift patterns, particularly at roof steps; the pathways where snow would, at full scale, drop off roof steps and pile up as drifts. This physical evidence was then merged with 21 years of Toronto's weather records to show worst-case snowdrift patterns for the roof. With all this data, and a numerically intensive computer-modelling process, the engineers were able to design the roof's steel structure.

Computational fluid dynamics (CFD) studies are playing a growing role at RWDI. CFD allows for the study of fluid flow at a level of detail not possible in a wind tunnel. CFD can be applied as much to modelling heat transfer through a CANDU nuclear reactor as it can to tracking snow travelling across roofs. As part of a study for the Atomic Energy Control Board of Canada, RWDI is assessing CFD predictions of flow and temperature patterns in the heavy-water moderator fluid circulated through a CANDU reactor vessel. Identifying potential hot spots is important because, in a CANDU reactor, the heavy-water moderator is expected to be an effective heat sink in many postulated loss-of-coolant accidents.

*Computational Hydrodynamics*

The Computational Hydrodynamics Laboratory at the National Research Council's Institute for Marine Dynamics in St. John's, Newfoundland, is developing complex numerical models of ship interactions with heavy seas.

The numerical models treat the wave-ship interaction as a non-linear scattering problem. An efficient numerical model was developed that uses the latest achievements in the development of methods for the computation of radiation potentials and forces, that can be solved on a wide range of computer systems, from workstations to supercomputers.

These numerical tools for the practicing naval architect are currently being ported and optimized for parallel computers under commercial contracts with Alex Informatique of Montreal. The agreement resulted in the installation of a 32-processor parallel computer at the laboratory in the fall of 1993.

*Ice Loading and the Northumberland Strait Crossing Project*

The Northumberland Strait Crossing, now under construction, is a 14-km bridge linking PEI with mainland Canada. The prime contractor is the Strait Crossing Joint Venture of Calgary. CANATEC Consultants Ltd. is assisting the design review team of Buckland and Taylor, which is providing an independent engineering assessment of the overall design.

CANATEC's task is to evaluate the analysis performed by the Cold Ocean Design Associates of the ice loads on the bridge piers. CANATEC built a software model to analyze the range of ice loads in light of approximately 30 different input parameters. The base model was designed for a 486 PC and ran each evaluation in about 2 1/2 hours. The original plan was to run a relatively small number of variations of the base case.

Additional requirements stimulated CANATEC's interest in the supercomputer.

1. The base model had to be extended for much longer time periods.

2. The uncertainties associated with the input parameters required a very large number of runs.

The model designed for the PC has been modified somewhat to achieve speedups of about 15 times over the original performance on the Fujitsu supercomputer at the HPC High Performance Computing Centre in Calgary. To fully take advantage of the vector facility on the supercomputer, a redesign and some reprogramming would be required, promising potential performance gains of several orders of magnitude.

The ultimate benefit of this modelling is increased public safety. The uncertainty runs are very important and, without the supercomputer, this work would have surpassed the capabilities of CANATEC.

*Reduction of Erosion and Increasing Equipment Longevity with CFD*

Syncrude Canada's research scientists are working with a mechanical engineer from the University of Alberta to explore methods to evaluate new designs for a let-down valve at the Fort McMurray site. The researchers are using computational fluid dynamics (CFD) to help design new valves with reduced erosion rates on the internal solid surfaces of the valve.

Currently, because of erosion of the wear-resistant coatings in the valve, the coatings require replacement every few months. The current coatings are made of the most wear-resistant materials available. By modifying the shape and fluid dynamics of the valve, the researchers hope to realize significant reductions in erosion rates. They will use CFD to develop methods for evaluating erosion rates in candidate valve designs.

The researchers are currently using Flow3-D, a CFD code developed in the U.K. This code is currently running at both Syncrude and the University of Alberta on workstations, but in both cases the research requires resources of memory and CPU time that are only available on a supercomputer.

Flow3-D had to be ported to the Fujitsu VPX240 at the HPC High Performance Computing Centre (HPCC) in Calgary by the HPCC technical staff. Initial optimization efforts have resulted in speedups of about 15 times current performance on high-speed workstations. Improved network access between the University of Alberta and Calgary has resulted in relatively easy access to the HPC Centre. Large memory and the vastly improved computational performance of the supercomputer enable these researchers to do work they were incapable of performing with their own local systems.

There are clear financial benefits to reducing erosion rates in Syncrude's let-down valves. A single let-down valve costs $60,000. Extending the life of one valve could mean potential savings of tens of thousands of dollars per year. In total, the company spends millions of dollars in replacing eroded pipes, valves and joints each year. Understanding particle movement within the pipe flow will enable researchers to better design pipe segments to reduce erosion and increase equipment longevity. Ultimately, this could save the company hundreds of thousands of dollars per year.

*Seismic Processing*

Pulsonic Geophysical Limited (PGL) is a Canadian-owned seismic processing company wholly owned by Pulsonic. Pulsonic bases its processing business on two levels of expertise: its geophysical processing teams and its application software, Index.

Pulsonic's strength has been in processing Canadian exploration survey data. This processing is seasonal in nature; the high season starts late autumn and ends in late spring. Companies traditionally have had to purchase computer equipment to handle the peaks in order to remain competitive. This means significant capital investments which depreciate rapidly.

Pulsonic's strategy to combat these influences was twofold. First, they decided to aggressively pursue international processing business in order to ensure a steadier supply of work. Secondly, they negotiated the purchase of high-performance computer time to augment their baseline computer facilities during peak periods, giving them quick turnaround of large jobs.

Index, the seismic software package designed and developed by Pulsonic, was ported and optimized for the Fujitsu VPX240 at the HPC High Performance Computing Centre (HPCC). The project, which involved the conversion and optimization of approximately 250,000 lines of Fortran and C code, involved all the technical staff members at the Centre as well as the software development teams at Pulsonic. Performance increases ranged up to 23 times because of enhancements made through optimization and vectorization of various modules. Pulsonic Geophysical is now running in production mode on the VPX240 at HPCC. The project has grown from two users initially to access by four full geophysical processing groups totalling approximately 11 users.

The purchase of supercomputer time helped Pulsonic in several ways:

1. The firm's money is no longer tied up in capital investments to meet peak demands. Pulsonic is paying only for computer resources consumed in revenue-generating projects.

2. Access to the power and speed of the supercomputer enables Pulsonic to bid on bigger projects and international business.

3. Access to computer service bureau offerings allows Pulsonic to use the latest technologies affordably.

4. Access to the HPC Centre's staff expertise has helped improve software performance on Pulsonic's in-house systems, as well as on the supercomputer.

*Processing Radarsat Data to Track Sea Ice*

Environment Canada's Ice Centre is gearing up to make full use of RADARSAT, Canada's first earth observation satellite, scheduled for launching in 1995. RADARSAT will carry Synthetic Aperture Radar (SAR) to produce reliable, all-weather images and provide an up-to-date picture of sea-ice conditions, vital to fishermen, shipping and Coast Guard operations.

The large volumes of data generated by RADARSAT will require fast, computerized analysis, which can later be augmented by expert interpretation, as required. Noetix Research Inc. of Ottawa, Ontario, is developing a set of algorithms to support computerized tracking and classification of sea ice.

The Ice Centre will be the single largest customer for ice reconnaissance data from RADARSAT. Computer software algorithms will be necessary to greatly speed up interpretation of ice motion and type. Human interpreters would be overwhelmed by the volume of satellite data.

With the introduction of algorithms to interpret ice data and the resulting large volumes of information, better knowledge of actual ice dynamics will drive improvements in kinetic modelling. The ice tracking algorithm will help validate ice models and is an important step towards the goal of producing better models within three to five years.

Classifying sea ice by using algorithms poses a more sophisticated challenge. Separate algorithms must be developed to identify spatial and textural features. The first step will be to distinguish between open water and ice. The next step is to use algorithms to determine whether the ice in a given region is homogeneous or heterogeneous. It is not yet possible to label ice automatically according to age or other specific features. The algorithm is therefore known as an unlabelled ice classifier. A labelled classifier will not be available for several years and will require significant increases in computer processing.

The data generated by RADARSAT will require significant storage, as well as high-speed networks to distribute to commercial interests. Much of the value of the data generated has real-time constraints that requires images to be transmitted across the country in minutes or, at most, within a few hours to be commercially marketable.

*Acoustic Imaging Using DRUMS*

Guigne' International Ltd. (GIL) is a high-tech research firm based in Paradise, Newfoundland, that designs and produces "intelligent" sonar. GIL researchers and a Department of Fisheries and Oceans team are investigating the effects of fishing trawlers on bottom-dwelling sea creatures. Under a series of contracts sponsored by the department's Northern Cod Science Program, the team is using GIL's patented Dynamically Responding Underwater Matrix Sonar (DRUMS ™) to record and analyze acoustic images of burrowing sea animals and the tunnels that they dig.

Through advanced digital signal processing, including fractal processing, the researchers can make correlations between the attenuation, reflectivity and scattering of the sonar's broad-frequency-spectrum signals, and the animals and structures within the collected samples. The acoustic imaging of the ocean's subsurface provides researchers with an exciting new tool for research studies.

*Ministry of Health Decision Support Using Tablemaker*

The B.C. Ministry of Health is now downloading the medical insurance plan claims database to the MasPar parallel computing system. The data are loaded into the TableMaker decision support database. Research officers at the Ministry offices can access the data directly from their PCs using the client/server TableMaker query tool. The database receives approximately 2 million claims every two weeks resulting in table sizes in excess of 250 million rows, constituting a 40-gigabyte database. The MasPar is ideally suited to large data applications and typically performs queries 50 times faster than an IBM ES9000 mainframe computer.

*Protein and DNA Database Query Tool for Genetic Researchers*

BC Systems and the University of Victoria have collaborated on developing the SeqseeMP protein and DNA sequence alignment tool for the MasPar massively parallel computer. The Uvic protein sequencing laboratory is able to compare protein sequences against the standard public databases using the electronic mail interface with SeqseeMP. Although other Internet sites around the globe offer similar services, the algorithms used in SeqseeMP are unique and may provide information not found elsewhere. BC Systems and the University of Victoria are continuing to develop this technology. The protein sequencing code has recently been extended to process DNA sequences.

### 3.8.2 Academic Applications

As examples of current research being performed on HPC facilities in Canada, the following are descriptions of nine projects that were conducted this year on the HPC facility in Calgary. We have chosen these examples, as the Centre in Calgary represents the only supercomputer-class facility in Canada that has been widely available to all universities over the past year, and these were the only Canadian applications presented at this year's Supercomputing Symposium. (For complete documentation on hundreds of university applications in Canada over the past seven years, see the Proceedings of the annual Canadian Supercomputer Symposia for the years 1988 through 1994).

At the Supercomputing Symposium 93, held in Calgary, Alberta, in June 1993, Fujitsu and the HPC High Performance Computing Centre jointly announced a university scholarship program, open to all university researchers across Canada. The intent of the program was to promote the use of high-performance computing in Canada by offering free access to the HPC Fujitsu VPX240/10 supercomputer to any university research project that was deemed to have reasonable scientific merit in the opinion of an evaluation review committee, composed of experts in supercomputers and science from the HPC Centre and the University of Calgary (which administers the program).

For this first scholarship program, there were 35 grants awarded to a wide cross section of applicants across Canada, representing a very diverse range of applications. From this group, approximately 15 submitted results and documentation from which a winner was selected by the committee: Yong Luo, Department of Electrical Engineering at the University of British Columbia. At the 1994 Symposium, HPCC and Fujitsu announced that this successful program was to be extended for the following year and that applications were being accepted.

The following is a brief description of the nine research projects that were presented at the Supercomputer Symposium 94:

1. *Computational Aspects of Monte Carlo Simulations with Long-range Interactions.* A. B. MacIsaac and J. P. Whitehead, Department of Physics, Memorial University of Newfoundland.

This research is in the domain of elementary particle physics, where quantum mechanics theories apply. Such studies tend to be extremely computationally intensive. The particular study involves magnetic systems with spins coupled by a long range dipolar interaction. The interest in these systems stems from efforts to interpret the large number of experimental studies on the magnetic properties of high-temperature superconductors. While there is a substantial body of literature concerned with the computational aspects of magnetic systems with short interactions, it is only recently that much attention has been given to the effects of long-range interactions and how to effectively treat these interactions within a Monte Carlo simulation. While limited studies on small lattices have been possible on workstations, supercomputers are now required, along with better algorithms, in order to adequately study the complex phenomena under investigation.

2. *Direct Simulation of Oil-Boom Failure.* E. J. Clavelle and R. D. Rowe, Department of Mechanical Engineering, University of Calgary.

Oil booms are often used to contain or divert oil spills floating on the surface of water, be they on the ocean or a river. Successful containment necessitates knowledge of the dynamic behaviour of the slick behind the boom so that the slick thickness, required boom draft and maximum volume that may be contained can be determined. Oil-slick behaviour is dependent upon the relative oil-water velocity, the oil volume per unit length of barrier and the physical properties of the oil and water.

Oil booms are subject to three modes of failure: drainage failure, droplet entrainment and critical accumulation, the last being the least understood. The analysis of oil boom critical accumulation is complex and computationally intensive on account of the numerical solution of fluid flows at the moving interface between oil and water. An exact solution using Navier-Stokes equations would be computationally prohibitive on any computer system, so the requirement is to develop numerical approximations that adequately simulate the real phenomenon.

The VPX240 supercomputer is now being used to improve the accuracy of the study so that eventually oil booms can be better designed and thus be far more effective in specific oil-spill situations.

3. *Image Coding using Vector Quantization with Fujitsu VPX240/10.* A. Jain and S. Panchanathan, Department of Electrical Engineering, University of Ottawa.

With the emerging information highway, there are increasing demands to discover more efficient methods of compressing image data in order to conserve both bandwidth for transmission and storage space. The objective is to reduce bit rates while maintaining acceptable quality.

An image-coding technique called "vector quantization" (VQ) is studied through computer simulation. In VQ, each image is partitioned into a set of non-overlapping blocks, and then each block is organized as a vector. Each vector is represented by a closest matching "code word" from a table to form a new set called a "code book". This system of coding, which results in compression, is a time-consuming iterative process, which is continued until an acceptable level of distortion is realized. Compression is achieved by transmitting only the index (label) of the closest matching code word for every vector in the image. The image is reconstructed by a simple table lookup operation using the label as an address to the table that contains the code words.

This VQ process is very time consuming if processed on a workstation; however, considerable speedup is possible on a vector supercomputer (like the VPX240) and even more speedup would be achieved on a parallel machine. On a 486 workstation, a single 512 x 512 image took five days to process, whereas an image of this size required only one minute on the VPX240. Obviously this research is able to advance at a much improved rate with the availability of the supercomputer.

4. *Large Scale Numerical Modelling of Postglacial Rebound.* Z. Ni and P. Wu, Department of Geology and Geophysics, University of Calgary.

An ongoing revolution in the geological sciences has come about because of the widespread acceptance of the theory of plate tectonics, which postulates that continental masses have moved horizontally with respect to one another throughout geological time. Implicit in this theory is that mantle material must be able to flow like a viscous fluid over geological time. The crucial questions that have to be asked and answered in order to believe or deny these theories are as follows: What is the viscosity of the mantle material? Is the viscosity too high for the mantle material? Or, is it low enough to flow geological time? What is the viscosity variation within the mantle? The key to answering all these questions lies in the study of postglacial rebound.

Postglacial rebound refers to the process of the Earth's adjustment following the last ice age, 20,000 years ago. After deglaciation, with ice loads removed, the land slowly returns to its initial state before glaciation, causing earthquakes and faulting as movement occurs. Also, the melting ice causes a global rise in sea levels. How fast the Earth rebounds depends on its flow properties, whereas the timing and magnitude of the rebound depend also on the history of the ice load. Therefore, postglacial rebound is often used to infer the flow properties of the Earth and the ice history during the last deglaciation.

Through the use of supercomputing, large-scale computationally intensive finite-element models are created to model this postglacial rebound and extract information about the flow properties of the Earth. The supercomputer models differ from previous, less-sophisticated models in that they consider the Earth as a sphere whose properties can vary in both radial and tangential directions. Very simple models were previously developed on the University of Calgary local servers, but the research was never completed because of lack of memory, lack of disk space or excessive run times. Very encouraging results have been obtained, however, on the VPX240 using a finite-element program, ABAQUS. The main run required 8608 seconds of computer time, even on this very fast supercomputer.

These studies are significant for the understanding of earthquake activities and the state of stress in the Earth's lithosphere, which will be beneficial for mining, petroleum exploration and civil engineering. The research will also have impact on studies such as global climate change, glaciology and the migration of North American Indians, to name just a few.

5. *Marine Controlled-Source Electromagnetics for Hydrocarbon Exploration.* T. T. Boyce and I. J. Ferguson, the Department of Geological Sciences, University of Manitoba.

Marine Controlled-Source Electromagnetic (CSEM) surveys map the electrical conductivity of the Earth's subsurface, the seafloor in this study. This conductivity mapping may be done by measuring the travel time of an electromagnetic (EM) impulse diffusing between transmitting and receiving antennae located on the seafloor. The electrical conductivity is interpreted in terms of geological structure. One important application for CSEM is hydrocarbon exploration in marine sedimentary basins. Rock types important to hydrocarbon exploration can be differentiated on the basis of conductivity values. While EM methods do not directly identify oil deposits, the identification of favourable rock types and the presence of geological structure, capable of trapping hydrocarbons, are critical to successful exploration.

Considering the costs and increasing difficulty of offshore hydrocarbon exploration, EM methods must be considered as a supplemental exploration method. Seismics, the most common technique, often provide poor resolution of structures such as steeply dipping faults or regions beneath high-velocity layers such as volcanic rock. As seafloor CSEM is costly, numerical design studies are needed, both to examine the feasibility of the method and provide estimates of the signal levels to be expected at the EM receivers.

Such studies are being conducted using a finite-element model at the University of Manitoba. In order to model realistic seafloor structures, and in particular to examine the sensitivity of the seafloor geological response to these structures, it is necessary to compute the response for a relative complex conductivity model and for a range of recording times. Early studies were done on a SUN workstation and then on a CRAY XMP supercomputer. With the Fujitsu VPX240, the research is being expanded with more complex models operating on larger mesh sizes.

6. *Modelling Heterogeneities in the Earth's Crust: Important Effects of Small-Scale Statistical Variation.* D. Cao, Department of Geology, Queen's University.

Analyzing and interpreting seismic surveys is today the most important methodology for locating hydrocarbon reserves under land or under the seabed. As exploration companies seek the remaining more obscure reserves, it is important to refine the interpretation of the seismic data. It is also critical to avoid the drilling of very expensive "dry holes". This study is one of many world-wide to identify techniques to improve the results of seismological studies.

Recent seismological studies have demonstrated the importance of considering statistical models of small-scale variation within larger geological structures. For example, the images of seismic reflectors (as seen in seismic sections), associated with large-scale discontinuities in seismic velocity beneath the Earth's crust, can be distorted by statistical distributions of small velocity heterogeneities below the crust. The imaging of structural features below the crust is seriously affected by statistical distributions of small heterogeneities. Numerical simulations of the propagation of elastic seismic waves within the crust can demonstrate and quantify the effects of random distributions of heterogeneities on the structure and continuity of the reflectors. The finite difference method has been used to simulate the seismic responses of crustal models. The validity of results is critically dependent upon the speed and memory of the computer used in the modelling process. Comparison between results from the VPX240 supercomputer simulation and that of a workstation have demonstrated that a realistic simulation requires the speed and memory of the supercomputer.

7. *Numerical Simulation of a Transient Close-Contact Melting Problem.* C. Wu, D. Phillips and A. Jeje, Department of Chemical and Petroleum Engineering and Academic Computing Services, University of Calgary.

Close contact melting of sulphur, in which the liquid produced is continuously squeezed out of a narrow gap, was numerically modelled in this study. The liquid viscosity and heat capacity exhibit strong temperature dependence. A hot cylinder forced into close approximation with a flat wall of sulphur at its melting point supplied the latent heat and caused progressive deformation of the wall. Hydrodynamic and energy equations were solved simultaneously using a time-marching procedure based on a combined finite difference and element scheme in a fixed grid domain.

The program was initially developed on a local workstation, but it was taking two to four weeks, for example, to advance through approximately five minutes of simulation. For accelerated calculations, the program was modified and transferred to the Fujitsu VPX240. Although a speedup of 10 has been achieved, the vectorization percentage is only 20%, leaving ample room for further speedups through code optimizations.

8. *Quantum Monte Carlo Simulations for a Model of High-Temperature Superconductors: Effect of Next-Nearest-Neighbour Hopping.* A. Veilleux, A. Daré, L. Chen, Y. M. Vilk and A. Tremblay, Phisics Department, Université de Sherbrooke.

Superconductors are characterized by their unusual electromagnetic properties. For example, they can transport current without loss and exhibit perfect diamagnetism. The microscopic understanding of superconductors came in 1957 with the BCS theory. In this theory, electrons form pairs whose wave-functions combine in a coherent manner. This is the fundamental explanation of superconductivity.

The microscopic understanding of high-temperature superconductors still poses a challenge. The difficulties arise from the fact that one must work in the regime where interactions between electrons are strong. The simplest "Hamiltonian" which contains the essential physics of the problem takes into account electrons hopping on a square lattice and screened Coulomb repulsion. This new study at the Université de Sherbrooke uses the so-called determinantal quantum Monte Carlo approach to study magnetic and superconducting correlations in a model that includes more realistic kinetic energy (band structure) terms.

The resultant models typically take days or even weeks on workstations, so the codes have been ported to the VPX240 to accelerate the research. Monte Carlo methods are very parallelizable, as one can run independent random number sequences simultaneously on each processor. Consequently, the research team is looking forward to next simulating this code on a massively parallel system.

9. *Simulating Some 2-D Seismic Reflection Models on the Fujitsu VPX240 by Polynomial Time-Marching*. Y. Luo and M. J. Yedlin, University of British Columbia.

Numerically simulating wave propagation in heterogeneous media has become one of the major tools for studying seismograms. Many different numerical algorithms have been applied in this area. Particularly, finite-difference time-domain algorithms have been intensively implemented on supercomputers for seismic forward modelling, which starts with an assumed Earth model to generate the wave field by solving elastic wave equations. Because of their high resolution and accuracy, spectral methods have also focused the attention of seismologists for some years. Conventionally, the spectral methods are applied to the space discretization and time-marching is still done by the finite-difference explicitly or implicitly. These time-marching schemes have a finite-order accuracy in the time domain and a critical stability limitation in step size.

In this study, a newly developed time-marching technique, based on complex polynomial interpolation, is applied to the numerical seismic wave field simulation with realistic boundary conditions. Based on this more accurate time-marching scheme, a balance of accuracy in both time and space dimensions can be achieved. Due to its looser stability limitation in time step size, fewer time steps are needed for simulating the wave propagation.

The matrix operations of the algorithms were such that the model was very adaptable to the VPX240 supercomputer and achieved extremely high performance relative to the theoretical peak of the machine. The complete application achieved 96.8% vectorization and the kernel itself ran at 2.44 gigaflops on a 2.5-gigaflop machine, a performance record for the Calgary HPC machine.

### 3.8.3 New Application Opportunities

The sections above review some of the typical applications being processed on HPC systems in Canada today. When searching for new application areas, perhaps the first step is to determine whether other countries are using HPC in ways that are applicable to Canada. There are five areas that should be examined:

- In the U.S., Europe and Japan, the pharmaceutical industry is using HPC molecular models to shorten the total design time required to discover new drugs. In Canada, even though there is an active pharmaceutical research community, the research appears to be mostly concentrated on laboratory experiments and clinical trials. Could Canadian companies not supplement these traditional methods with HPC techniques to accelerate their success rate in new drug designs?

- The chemical industry also uses molecular modelling extensively in other countries to design new compounds and materials. One interesting pilot project has been recently completed with a small chemical company in Calgary. The firm has been working with the IRAP consultants of NRC and with the Calgary

HPC Centre to develop a new molecular modelling research methodology for better understanding of rust-inhibiting compounds for the petroleum industry.

- Alcan has used HPC models for several years to study and improve the processes and products in the aluminum industry. In the U.S., there have been similar developments in the steel industry. It should be possible for other materials companies in Canada (steel, plastics, etc.) to take advantage of similar HPC techniques.

- Canadair mentioned integrated product design and product optimization as very high-pay-back applications for the aerospace industry (and presumably any heavy manufacturing industry). The barrier to implementation of such processes is the lack of experts, the lack of software and the high cost of application development. Is there some form of collaboration between research labs, universities and industry that could introduce these techniques to Canadian corporations?

- Can we not learn something from the financial institutions on Wall Street that have introduced parallel computing to applications such as portfolio analysis and testing of investment strategies?

- One Canadian airline stated that there are several optimization applications that could result in large potential cost savings—for example, crew scheduling, flight scheduling and yield management (optimizing the loads on all flights through discount plans, flight scheduling, etc.). Such applications should run on high-performance computers, not mainframes that are dedicated to administration or flight reservations. Investing in the development and implementation of these applications may require pilot projects with the help of HPC experts.

A second opportunity for Canada, as mentioned several times throughout this document, is the use of parallel processing servers for information highway applications. Interest has already been expressed for applications such as the following:

- a national climate database operated by Environment Canada;
- Geographical Information System (GIS) databases;
- Earth observation satellite (EOS) databases;
- Natural resources databases (e.g. petroleum information);
- DNA sequencing databases;
- Protein sequencing databases;
- Health-care databases containing both patient records and scan images;
- Digital library information.

There are organizations in Canada that are prepared to work in partnership to build these applications and deliver the services. Two of these applications are already being implemented in B.C.: genetic sequencing and health care.

Another opportunity is to assist Canadian industry, perhaps through pilot projects and demonstrations, in developing new commercial applications that utilize the power of multi-purpose parallel processors. Subsection 4.1.3.2 includes several examples already implemented in the U.S. including:

- processing massive volumes of insurance claims;
- micro-marketing, whereby massive databases of sales statistics are searched for buying patterns in retail sales companies, airlines and telephone companies;
- searching for patterns in credit-card transactions that identify instances of fraud;
- enterprise information systems that allow all corporate data to be kept on-line, with very fast search times for decision support systems.

A final opportunity involves the utilization of HPC technology for multimedia applications on the information highway. Multimedia, because it includes an increasing amount of compressed digital video information, requires an inordinate amount of storage space—many terabytes for some applications—and a corresponding very high system bandwidth for dissemination. These applications require very large parallel multimedia servers. For these servers, world-wide HPC research efforts involve adapting the same MPP systems developed for scientific applications, but with innovative new software, algorithms and intelligent interactive tools for creating and managing distributed multimedia databases and for browsing, navigating, searching, retrieving, combining, displaying, visualizing, and analyzing very large amounts of information that are inherently in different formats. Research will also address standards that allow for interoperability. Multimedia applications that may require HPC technology include:

*Education*

- Life-long learning, at work and at home
- "Edutainment"—games set in an educational environment
- Virtual field trips

*Digital Libraries*

- Knowledge centres without walls

*Remote Simulations and Animation*

- Visualizations and real-time animation for remote HPC simulations and for educational purposes

*Museums*

- Virtual visits to libraries and galleries world-wide

*Health Care*

- Remote consultations
- Medical imaging

*Entertainment*

- Video on demand

Because of the growing importance of multimedia applications, CITI has commissioned a separate extension to this HPC study to examine the possible relationship between HPC and multimedia. A report on this study will be available in May 1995.

## 3.9 Canadian Software Development for HPC Systems

The approach employed in order to identify and contact Canadian companies and individuals engaged in HPC software development was to ask all the respondents to the HPC questionnaires to indicate the names and contacts for firms with which they were familiar. In addition, the consultants already had some contacts. A review of Industry Canada literature identified others, though Industry Canada does not appear to specifically categorize the computer industry in terms of HPC.

The search produced 25 names, 21 of which were contacted to obtain more details on their products. The input from some of these contributed to the descriptions below.

The response does generate concern that there are only a few individuals or groups in Canada producing software for a potentially lucrative market outside of Canada. This concern expands with the realization that there are entirely new markets opening now that parallel architectures are finally beginning to reach commercial viability.

### 3.9.1 World-wide HPC Software Market

The following market projections are based on the 1993 report of the Superperformance Computing Service (Palo Alto Management Group, Inc.).

The world-wide market value for all classes of HPC systems approached $19 billion (U.S.) in 1993 and should rise to $22.5 billion (U.S.) by 1997. The market for the largest systems (supercomputers) remains constant at about $2.5 billion (U.S.) throughout this forecast period, with the growth appearing in small to medium-size systems. Desktop systems are not included in this analysis; the figures include only "technical" systems and exclude those used for commercial purposes.

It is noted that MPP systems are on the verge of becoming the force majeur for technical computing and are on the threshold of making rapid inroads into the commercial market. MPP systems represented about $6 billion (U.S.) of the 1993 market, with an expected rise to $3.8 billion (U.S.) by 1997. The numbers of MPP systems is expected to rise from 735 in 1993 to 4,090 in 1997. Commercial uses are expected to represent 35% of the MPP systems by 1997. The Gartner Group and IBM are forecasting much higher numbers for commercial MPP systems.

Canada does little manufacturing of these systems and therefore cannot anticipate a place within these projections that could have a positive effect on our balance of trade. Where it could have a major impact is in a Canadian software industry that recognizes the potential of the HPC software market and aims to capture a respectable portion. The Europeans and the many U.S. firms that are currently active in software will undoubtedly also recognize the potential, but Canadians can and have competed effectively in the general software market.

Using a standard rule of thumb that for every dollar spent on hardware there are two to three dollars spent on software, Canadian software developers would be missing out on a multi-billion-dollar annual opportunity. With the anticipation that MPP systems will start delivering on their exciting promises, especially in the commercial sector, the opportunities will become enhanced even further.

As recognized by the U.S. and Japanese hardware developers, there is also a tremendous benefit from the "trickle-down" effect of the knowledge and experience gained during the research and development stages. The sophisticated software components developed for HPC are continuously being migrated down to the smaller-scale (desktop) computing systems. Moreover, this is well recognized as the fastest growing market for scientific computing, expected to reach $40 billion (U.S.) by 1997.

### 3.9.2 Examples of Canadian HPC Software Development

The following examples of Canadian organizations that are developing software specifically for HPC systems serve to highlight Canada's strengths and weaknesses in this marketplace.

It is interesting to note the comments of respondents to the study questionnaires. These can only be summarized as general trends, as there was too small a response sample to provide detailed statistics. In any case, it was noted that of the 14 HPC applications software packages currently in use, five were developed in-house in response to a local need, while the balance was acquired from outside Canada. For these five internally developed programs, there seemed to be no intention to commercialize the software. This is understandable, as the cost and effort of taking a good locally usable package to a full market position would probably involve a tripling of the resources. Market establishment, distribution and ongoing support will also substantially add to the basic package cost.

However, in the area of systems software and tools, a few companies were identified that have firm intentions to pursue global marketing of their product, and two companies have already established themselves as world market leaders in their niches.

It has also been observed that most HPC software developments begin in the academic community, especially for complex programs related to fields such as rational drug design and material science. Canada has been quite successful at doing the basic research for these applications, but not all that successful in diffusing this knowledge and expertise to the private sector.

In passing, it is worth noting that there are two other important Canadian software success stories—Alias Research of Toronto (recently merged with Silicon Graphics Inc.) and Softimage of Montreal (recently acquired by Microsoft)—both world leaders in the development of film animation software and visualization tools for industrial design. They are not included in this section, as their products run on workstations; however, as the line blurs between entry-level HPC systems and very high-powered graphic workstations, it may be correct to include these applications as part of HPC.

#### 3.9.2.1 IBM Canada's Development of a Database System

One of the less-known success stories in Canadian software development is the world-wide role that IBM Canada's Software Development Laboratory has enjoyed. For several years the laboratory has been responsible for the development and support of software packages, including compilers, database-management and software-development support tools (CASE) for the IBM RS6000 workstation product. Considering that the technology of the IBM RS6000 forms the base for their MPP and clustering HPC systems, it was not unreasonable for the Canadian group to be selected to lead the development of IBM's parallel edition of its database software. The product has recently been released for IBM HPC systems, including workstation clusters and IBM's SP2 MPP system. Target applications are stretching the role of HPC into on-line transaction processing and decision-support systems.

The benefit to Canada so far can be measured by approximately 100 jobs over the year-long development period, and similar numbers for the long-term support and product enhancement.

#### 3.9.2.2 Waterloo Maple Software

Waterloo Maple Software is a Canadian success story, where software developed by university researchers was moved to the private sector for global commercialization. The function of this software is to simplify the programming of scientific HPC applications using symbolic mathematical terms. The eight years of development represented a part-time activity for the researchers from the University of Waterloo, but the project went commercial in 1988. Since then, growth has been exponential. Maple V, as the product is known, has taken a position in HPC systems, but the majority of the product's revenue comes from the lucrative desktop marketplace. The product is no longer just a research tool, but is now used in education as well, particularly in high school math programs.

Further success is coming from partnerships where other developers are embedding Maple within other math "toolboxes". For example, in 1993 400,000 installations of the Mathcad package included Maple.

#### 3.9.2.3 Optimax Software Packages

This Halifax-based group identified the requirement for sophisticated computer-based tools to assist developers of scientific applications and other major software systems. It is technology-independent, though emphasis is naturally placed on computer architectures that are available today or are expected to emerge in the near future. In addition to providing a general environment to improve the efficiency of software development, Optimax has produced very specific tools to assist developers in optimization of codes that are targeted for parallel and/or vector machines. Realizing that fully automatic optimization is not realistic, Optimax has taken the approach of developing interactive tools, whereby computer analysis procedures guide the user as much as possible through the optimization process, thus arriving at the end result, a highly optimized code, as rapidly and as efficiently as possible. The HPC community is naturally a target for this software, considering the complexities of these systems and the rapid change in the technologies that are being implemented, but it will also benefit the smaller-system users.

The benefit of these software tools is in the computer-aided management of the complexities of the software-development process for individual developers and for large groups that may be widely dispersed. The result of the utilization of the Optimax products should be increased efficiency at the implementation stage of the software-development life cycle.

Development has been financially supported by a wide number of government and private-sector backers. Optimax software has been under development since 1990 and was formally released into the marketplace in December 1994. Optimax has now entered the critical marketing phase of the product cycle; new sales and marketing experts have been hired and negotiations are in progress with organizations that could market and distribute the products through efficient channels.

#### 3.9.2.4 Myrias Computer Technologies

Myrias' present position in HPC software grew out of an attempt in the 1980s to establish a Canadian MPP system. While the hardware interests of the firm did not survive, like many similar ventures in the U.S. and Europe, their sophisticated software continues to offer new opportunities. Development now includes the use of the software on large-scale MPP systems and also on clusters of workstations linked by relatively slow networks.

The Myrias product, PAMS (Parallel Applications Management System), enables typical programmers, not just parallel computer specialists, to relatively easily harness the collective resources of multiple workstations or CPUs in an MPP system. The result is faster execution of large applications. In the workstation world, PAMS makes use of wasted cycles by executing large parallel applications in the background, while monitoring interactive foreground usage to ensure an absolute minimum impact. In short, PAMS enables customers to make better use of what they already own, to achieve faster turnaround on time-sensitive applications, and to execute large applications without having to upgrade hardware.

The University of Alberta has been a test site for the workstation version of PAMS and seems to be very satisfied with its ease of use and performance. Over the past year, Myrias has ported PAMS to Sun and IBM systems and hopes to use these two large vendors to help their international sales.

### 3.9.2.5 Platform Computing Corporation

The Platform Computing Corporation of Toronto, Ontario, which has developed a product called LSF (Load Sharing Facility), is another Canadian success story. A spin-off from the University of Toronto, Platform Computing has received impressive international acceptance of its LSF software package—a load-sharing and distributed-batch-processing software package that provides integrated support for sequential and parallel jobs on clustered workstations in a network.

The following are quotations from a recent HPC Newswire service:

> "We observe a tremendous need for tools that enable the efficient shift from mainframe to distributed, heterogeneous computing", said Judith Hurwitz of Hurwitz Consulting Group. "We're impressed with LSF's robust capabilities."
>
> "LSF is the technology glue that provides us with a cost-effective, low-maintenance, computing environment", said Neal Linson of Western Digital, Irvine, Calif. "With LSF, we're able to perform all computing on workstation clusters, and our design engineers maintain peak productivity because they can now access superior computing resources."
>
> "LSF 2.0 represents the maturity of distributed load sharing technology, and is the result of cooperative efforts among Platform, our system vendor partners such as, H-P/Convex, Digital and Silicon Graphics, and our rapidly expanding partner and user base world-wide", said Dr. Songnian Zhou, president of Platform Computing. "LSF answers the critical market needs for integrated distributed computing solutions which facilitate downsizing from mainframes and supercomputers. We are helping organizations achieve far more effective and economical network computing."

In Canada, LSF is used by organizations including Pratt and Whitney, Northern Telecom and de Havillard.

3.9.2.6 Advanced Scientific Computing Ltd.

ASC of Waterloo Ontario was founded in 1985. Its 39 employees in Canada and 11 outside of Canada provide marketing, support and enhancement for their TASCflow software package. TASCflow simulates gas and liquid flows in both 2-D and 3-D through turbines, pumps, compressors, rocket motors, pipes and ducts. It is also used by engineers and technical specialists in manufacturing processes (pulp and paper, steel, aluminum, oil, food, chemicals) and in consumer products (automobiles, boats, sporting equipment, etc.). According to ASC's estimates, the international market for CFD has reached $70 million per year and may be growing at 30% annually. ASC currently has 70 installations. The firm comments that the marketing of HPC software is highly technical and requires demonstrations on specific problems of interest to specific customers, i.e. an unusually labour-intensive and expensive sales activity.

3.9.2.7 Alex Informatique Inc.

Alex is a Montreal-based developer and manufacturer of parallel hardware and software systems. Alex has been supporting and encouraging Canadian software development for use in parallel computing environments, as well as doing its own internal development.

Alex management suspects that the firm will gradually move away from hardware to concentrate on software, again a recognition that the performance of the workstation will support clustering and will make use of the company's sophisticated parallel techniques.

3.9.2.8 Pixel Motion Images Inc.

Pixel Motion is a recent arrival on the Canadian software scene. This Halifax-based firm grew out of an HPC training and software-development group called Hypercomp. Pixel Motion is concentrating development on a package for the production of entertainment, and for comprehensible scientific visualization for non-technical audiences, for example, for environmental assessment and impact.

3.9.2.9 ACTC Technologies Inc.

ACTC is a Calgary-based company that offers applications-development and software-support services for HPC systems. ACTC does not develop or market specific software packages.

ACTC was one of the founding partners of the HPC Centre in Calgary, as it recognized that Canada continued to lag behind other nations in the use of HPC systems. As a partner of the HPC Centre, ACTC has been involved in conversions and optimizations of clients' application software for the supercomputer.

3.9.2.10 Can-Sol Computer Corporation

Can-Sol is a startup company based in Kanata, Ontario. The firm is pursuing the development of a parallel processing system and high-performance computing applications. While this company has plans to manufacture special hardware, it also will be developing systems and application software.

The company is developing a general-purpose parallel processor based on an architecture that is a hybrid of SMP and distributed-memory organizations. The first application being pursued is high-quality image generation. Can-Sol's product for this application is referred to as the Multiprocessor Image Creation System. The company believes that technologies to be developed over the next few years will allow this application to create real-time images, providing one of the first photo-realistic virtual-reality engines.

Discussions with other HPC software companies are underway to determine appropriate relationships for porting other applications such as seismic modelling, CFD analysis, climate and weather modelling and communications system design.

### 3.9.3 Summary of Canada's HPC Software Industry

Despite the enthusiasm of the representatives of the organizations listed above for the products they are developing or the services they provide, they too suffer from a lack of familiarity within the Canadian community of HPC users, or potential users. In most cases, their marketing strategy to compete on a world-wide basis was not obvious from documentation that was available to the consultants. Concern arises when one realizes that of the many Canadian software companies contacted in the study's information-gathering stage, only four were actually listed in response to the specific question asking respondents to indicate their knowledge of any current or past Canadian software companies. In the cases where individuals were using specific packages from the U.S. or Europe and were asked if a Canadian company existed with a comparable product, not one was identified.

Given that there is a vast untapped market for software packages for HPC-based applications and development tools, as detailed above, there appears to be a need for the identification of the most marketable products to the HPC user community world-wide. Software development will also benefit from the same "trickle-down" effect as has been experienced by HPC hardware developers—that is, the market for these software products will eventually expand to the much larger desktop-system market.

Encouraging a Canadian-based HPC software industry, targeting the full spectrum of applications, may present Canada with a much-required new job-producing opportunity.

### 3.9.4 New Software Opportunities

Subsection 3.8.3 above describes the many new application opportunities available in Canada, especially with respect to providing software environments that can harness the power of parallel machines. The domestic and world market for this software will be very large as parallel processing becomes mainstream for information highway and corporate applications. Canada is already quite advanced in the development of the basic parallel software, and a few commercial firms are now beginning to ship products for environments and development tools, e.g. Myrias, Platform Computing and Optimax. The next opportunity will be to develop end-user applications.

### 3.10 Canadian Associations with Interest in HPC

While there is only one association—Super*Can—that concentrates exclusively on the field of HPC, a few other application-specific groups have very close links with and dependency on HPC.

The name Super*Can originally stood for Supercomputing Canada. The group was initially a special-interest group composed of individuals who had involvement in Canadian supercomputing. It has been described as a loose association of individuals and organizations who come together once a year at the Canadian Supercomputing Symposium. There have been seven such symposia so far, the latest of which was held in Toronto under the auspices of the University of Toronto. Membership in Super*Can has basically resulted from attendance at one of these annual events and therefore currently numbers about 120.

Super*Can has recently determined that a closer alignment with the broader category of high-performance computing was more in line with its goals. Thus, at an executive meeting last summer, the name was expanded to Super*Can, The Canadian Association for High Performance Computing and Communications. The 12 members of the executive are elected annually by the membership at an annual general meeting held each year in conjunction with the symposium. Other executive meetings are held sporadically, primarily because of the lack of funds to support meeting expenses, travel and accommodation. In the past, the generous participation of interested Canadian organizations has supported staff participation in such meetings.

The current executive has committed itself to reviewing the bylaws of Super*Can, as those that were developed in 1989 are likely due for modernization. The executive has also undertaken to coordinate the publication of an annual Canadian HPC review, which will see broad distribution and provide the Canadian HPC community with a forum for sharing activities and interests.

The 1995 Symposium will be held under the auspices of CRIM, in Montreal. A feature topic will undoubtedly be the results of this CITI study of Canadian high-performance computing.

Since Super*Can was formed, the organization's role in relation to other Canadian organizations with an interest in information technology has been a continuing agenda item. The benefit of such an alliance is a broader audience for its activities and increased opportunity for educating other individuals in Canada who are knowledgeable in information technology in the area of HPC. Another perceived opportunity would be to share in the influence and recognition acquired by the more successful Canadian information-technology groups within Canada. The need for expanded awareness of Canadians at both the potential user level and the levels that influence funding for technology has consistently stood out as a requirement for the furtherance of Canadian HPC.

The question of need and benefit of application-specific organization is continually on the minds of their executive members. For example, there have been two meetings of a Canadian Computational Chemistry Conference: the first in Sherbrooke in 1991 and the most recent in Kingston in May 1994. The next conference will be in Edmonton in 1997. The participants are researchers who are developing avant garde algorithms and software for large-scale supercomputers, e.g. for complex molecular systems. The goal of the conferences is to bring together leading computational chemists to share ideas on large-scale computations.

Similarly, researchers who employ computational fluid dynamics techniques have formed a very successful Canadian society and also hold an annual meeting. Again, while they are perhaps more interested in the theoretical approach to CFD, their principal tool is the high-performance computer. The implementation of HPC is an important item on their meeting agenda. Sponsorship of the CFD Society has come from NRC, NSERC, Canadian universities, and various industry groups.

The question to be examined by all such groups is whether the research community can support multiple meetings of multiple societies and organizations, all of which devote considerable time and effort to topics of high-performance computing. It seems opportune for Canadian researchers who use HPC as a tool to now consider the most cost-effective approach to an annual update of their HPC information base, perhaps by joining together to pursue such information.

# 4 HPC STATUS IN U.S.A., EUROPE AND JAPAN

## 4.1 U.S.A.

### 4.1.1 Overview

The U.S. has always been the undisputed world leader in the development and use of HPC technology. The HPC industry started to take form in the late 1970s with the introduction of (vector) supercomputers in the U.S. By 1982, when the Japanese first introduced their own supercomputers, the U.S. market accounted for over 80% of all installed systems. Since then, the increasing rate of acceptance of supercomputers by Japan and Europe has reduced the U.S. share to approximately 50%.

The U.S. nevertheless continues to be the world trend-setter for new improved technology. It is the U.S. that first introduced the vector supercomputer. The U.S. developed and continues to dominate high-powered technical workstations, and it is the U.S. that dominates today's development and utilization of the new massively parallel machines. The U.S. develops and supplies most of the world's HPC application software.

The U.S. has far more programs than any other country to do the basic research using new HPC technology and to transfer this technology to industry. The first adopters of HPC were the U.S. national labs, primarily for military purposes. Over the last decade, we have seen a rapid shift from these defense projects to the industrial use of HPC to strengthen the U.S. competitive position internationally. The U.S. has realized that there is a wide range of HPC applications that will significantly enhance U.S. industrial competitiveness, productivity, energy conservation, quality of life, basic science and national security. The U.S. also understands the "trickle-down" effect i.e.—the fact that investment in HPC, the leading-edge of information technology, will benefit other areas of science and technology. A good example is the HPC program support for the development of the underlying technologies needed to implement the information highway.

HPC is widespread throughout the U.S. A $1-billion-per-year High Performance Computing and Communications (HPCC) program (explained below) has been established for basic and applied research that is in live with government priorities. At the state level, there are numerous state-supported HPC centres throughout the country, most of which have industrial partnership programs for diffusing the technology to industry. In addition to the HPCC program centres, many universities have their own dedicated systems and HPC research programs. Lastly, many private companies, especially in the aerospace, automotive, petroleum, chemical, and pharmaceutical industries, have realized sufficient benefits from HPC that they have installed their own dedicated supercomputers.

### 4.1.2 HPCC Program

#### 4.1.2.1 General Description

The U.S. federal High Performance Computing and Communications (HPCC) program was created to accelerate the development of future generations of high-performance computers and networks and the use of these resources in the U.S. federal government and throughout the U.S. economy.

In the early 1980s, U.S. scientists, engineers, and leaders in government and industry recognized that advanced computer and communications technologies could provide vast benefits throughout not just the research community but the entire U.S. economy. Senior government, industry, and academic scientists and managers initiated and are implementing the HPCC program to extend U.S. leadership in these technologies and to apply them to areas of profound impact on and interest to the American people. The federal High Performance Computing program was formally established following the passage of the High Performance Computing Act of 1991, introduced by then Senator (now Vice-President) Gore.

The scalable high-performance computing systems (i.e. a system whose overall performance grows linearly with the addition of processors), advanced high-speed computer communications networks, and advanced software being developed in the HPCC program are necessary for science and engineering and will contribute critical components of the National Information Infrastructure (NII). This infrastructure is essential to U.S. national competitiveness. It will enable the U.S. to build digital libraries, enhance education and lifelong learning, manage energy resources better, monitor and protect the environment, and improve health care, manufacturing, national security, and public access to information.

The program is planned, funded and executed through the close cooperation of federal agencies and laboratories, private industry, and academia. These efforts are directed toward ensuring that to the greatest extent possible the program meets the needs of all communities involved and that its results are brought into the research and educational communities and into the commercial marketplace as effectively as possible.

The program is well on its way toward meeting its original goals:

- As part of the HPCC program, more than a dozen high-performance computing research centres are in operation nationwide. More than 100 scalable high-performance systems are in operation at these centres. These include large-scale parallel systems, vector parallel computing systems, hybrid systems, workstations and workstation clusters, and networked heterogeneous systems. The largest of these systems can now perform more than 50 billion operations per second (50 gigaflops) on large problems. The HPCC program's 1996 goal of developing high-performance systems capable of sustaining trillions of operations per second (teraflops) performance is well on the way to being met.

- The HPCC-funded communications networks, part of the Internet, continue to experience phenomenal growth in size, speed, and number of users. These networks enable researchers to access high-performance computers and advanced scientific instruments easily and effectively, and have begun to allow for independence of geographical location.

- One illustration of the global reach of HPCC technology is that the Internet now extends across the country and throughout much of the world. The Internet links more than two million computers, more than 15,000 networks in the U.S. and more than 6,000 outside the U.S., 1,000 colleges and universities, 100 community colleges, 1,000 high schools, and 300 academic libraries in the U.S.

- More than half a dozen "gigabit testbeds" are being used to conduct research in high-capacity networks. These network testbeds connect 24 sites, including many of the high-performance computing research centres. The testbeds develop technologies to handle the NII's increased demand for computer communications, along with greater accessibility, interoperability, and security. The program goal of demonstrating gigabit (billions of bits) per second transmission speeds by 1996 is well on the way to being met.

- Teams of researchers are using scalable systems to discover new knowledge and demonstrate new capabilities that were not possible with earlier technologies. These researchers are addressing the "grand challenges", fundamental problems in science and engineering with broad economic and scientific impact whose solutions can be advanced by applying high-performance computing techniques and resources. Agencies are increasingly working with U.S. industry to use their grand-challenge application software and develop new software that will improve commercial and industrial competitiveness. We are beginning to understand our world better and to improve our chances of dealing effectively with our environment and with diagnoses and treatments of diseases such as cancer and AIDS.

- The thousands of researchers who develop fundamental HPCC and NII technologies and applications form the vanguard that hardware and software vendors rely upon to promote the use of high-performance computers and communications throughout the U.S. economy. Teachers and students can readily access HPCC resources, and the program conducts training events for multitudes of trainees. Smaller systems are being installed at colleges and universities. The goal of these efforts is to train a nation of knowledgeable users and thereby fully incorporate HPCC and NII technologies and applications into the U.S. economy.

#### 4.1.2.2 Five HPCC Programs

The HPCC program is organized into five integrated components with the following key aspects:

*High-Performance Computing Systems (HPCSs)*

The HPCS component extends U.S. technological leadership in high-performance computing through the development of scalable computing systems, with associated software, capable of sustaining a performance of at least one trillion operations per second (teraflops).

*National Research and Education Network (NREN)*

The NREN component extends U.S. technological leadership in computer communications through a program of research and development that advances the leading-edge of networking technology and services. NREN will widen the research and educational communities network connectivity to high-performance computing and research centres, as well as to electronic information resources and libraries. It includes nationwide prototypes for terrestrial, satellite, and wireless communications systems, including fibre-optics, with common protocol support and applications interfaces.

*Advanced Software Technology and Algorithms (ASTA)*

The ASTA component demonstrates prototype solutions to "grand challenge" problems through the development of advanced algorithms and software and the use of HPCC resources.

Grand challenge problems are computationally intensive problems such as forecasting weather, predicting climate, improving environmental monitoring, building more energy-efficient cars and airplanes, designing better drugs, and conducting basic scientific research.

*Information Infrastructure Technology and Applications (IITA)*

The IITA component demonstrates prototype solutions to "national challenge" problems using HPCC technologies. National challenges are informationally intensive applications such as education and lifelong learning, digital libraries, health care, advanced manufacturing, electronic commerce, and environmental monitoring.

*Basic Research and Human Resources (BRHR)*

The BRHR component supports basic research, training, and education in computer science, computer engineering, and computational science, and enhances the infrastructure through the addition of HPCC resources. Initiation of pilot projects for kindergarten through grade 12 and lifelong learning will support the expansion of the NII.

#### 4.1.2.3 Agencies Involved

The HPCC program reports to the Director of the Office of Science and Technology Policy of the Executive Office of the President.

HPCC program agencies work closely with other government agencies, industry, and academia in developing, supporting, and using HPCC and other NII-associated technologies. The 10 agencies that participate in the HPCC program are:

| | |
|---|---|
| ARPA | Advanced Research Projects Agency, Department of Defense; |
| DOE | Department of Energy; |
| ED | Department of Education; |

| | |
|---|---|
| EPA | Environmental Protection Agency; |
| NASA | National Aeronautics and Space Administration; |
| NIH | National Institutes of Health, Department of Health and Human Services; |
| NIST | National Institute of Standards and Technology, Department of Commerce; |
| NOAA | National Oceanic and Atmospheric Administration, (Department of Commerce); |
| NSA | National Security Agency, Department of Defense; |
| NSF | National Science Foundation. |

Through coordinated planning, research and development, these agencies are developing an integrated infrastructure of HPCC and NII technologies. No individual agency has either the mission or the expertise to develop all components of the infrastructure, but each plays a necessary and unique role in the overall program.

#### 4.1.2.4 HPCC Budgets

The total fiscal year (FY) 1994 HPCC program budget for the 10 participating agencies is $938 million (U.S.). For FY 1995 the proposed HPCC program budget for nine agencies is $1.155 billion, representing a 23% increase over the appropriated FY 1994 level. While these budget levels may appear staggering to Canadians, it should be noted that this is only one component of the U.S. expenditures in HPC. In addition, many state governments are also supporting large regional centres; the computer manufacturers and software developers invest large sums annually in R&D programs for new hardware and software and industry invests in purchasing its own systems and in collaborative programs with the government-supported centres. In short, the federal government funding is highly leveraged.

FY 1994 Budget (in millions of dollars)

| Agency | HPCS | NREN | ASTA | IITA | BRHR | TOTAL |
|---|---|---|---|---|---|---|
| ARPA | 103.9 | 48.7 | 32.3 | 95.6 | 18.4 | 298.9 |
| NSF | 19.8 | 47.9 | 119.3 | 19.0 | 61.0 | 267.0 |
| DOE | 10.8 | 16.6 | 73.8 | 0.3 | 20.7 | 122.2 |
| NASA | 11.5 | 13.2 | 73.1 | 12.0 | 3.2 | 113.0 |
| NIH | 3.4 | 3.0 | 26.1 | 13.6 | 11.7 | 57.8 |
| NIST | 0.3 | 1.9 | 0.6 | 15.3 | | 18.1 |
| NSA | 21.5 | 7.1 | 11.2 | 0.2 | 0.2 | 40.2 |
| NOAA | | 1.0 | 9.8 | | | 10.8 |
| EPA | | 0.7 | 5.9 | | 1.3 | 7.9 |
| ED | | 2.0 | | | | 2.0 |
| TOTAL | 171.2 | 142.1 | 352.1 | 156.0 | 116.5 | 937.9 |

FY 1995 Budget Request (in millions of dollars)

| Agency | HPCS | NREN | ASTA | IITA | BRHR | TOTAL |
|---|---|---|---|---|---|---|
| ARPA | 110.7 | 61.1 | 29.6 | 140.8 | 15.2 | 357.4 |
| NSF | 21.7 | 52.9 | 141.2 | 50.6 | 62.2 | 328.6 |
| DOE | 10.9 | 16.8 | 75.5 | 1.2 | 21.0 | 125.4 |
| NASA | 9.7 | 12.7 | 81.2 | 17.5 | 3.8 | 124.9 |
| NIH | 4.9 | 8.4 | 23.8 | 29.1 | 15.6 | 81.8 |
| NIST | 6.8 | 3.7 | 4.6 | 41.4 | | 56.5 |
| NSA | 16.1 | 11.8 | 11.8 | 0.2 | 0.2 | 40.1 |
| NOAA | | 8.7 | 16.1 | 0.5 | | 25.3 |
| EPA | | 0.7 | 11.7 | 0.3 | 2.0 | 14.7 |
| TOTAL | 180.8 | 176.8 | 395.5 | 281.6 | 120.0 | 1,154.7 |

#### 4.1.2.5 HPCC Program Strategies

The following is the list of current HPCC program strategies:

- Develop, through industrial collaboration, high-performance computing systems using scalable parallel designs and technologies capable of sustaining a performance of at least one trillion operations per second (teraflops) performance on large scientific and engineering problems such as grand challenges.

- Support all HPCC components by helping to expand and upgrade the Internet.

- Develop the networking technology required for the establishment of nationwide gigabit-speed networks through collaboration with industry.

- Demonstrate the value of wide area gigabit networking to support and enhance grand challenge applications collaborations.

- Demonstrate prototype solutions of grand challenge problems that achieve and exploit teraflops performance.

- Provide and encourage innovation in the use of high-performance computing systems and network access technologies for solving grand challenge and other applications by establishing collaborations to provide and improve emerging software and algorithms.

- Create an infrastructure, including high-performance computing research centres, networks, and collaborations that encourage the diffusion and use of high-performance computing and communications technologies in U.S. research and industrial applications.

- Work with industry to develop information infrastructure technology to support the National Information Infrastructure (NII).

- Leverage the HPCC investment by working with industry to implement national-challenge applications.

- Enhance computational science as a widely recognized discipline for basic research by establishing nationally recognized and accepted educational programs in computational science at the pre-college, undergraduate and postgraduate levels.

- Increase the number of graduate and postdoctoral fellowships in computer science, computer engineering, computational science and engineering, and informatics, and initiate undergraduate computational sciences scholarships and fellowships.

#### 4.1.2.6 HPCC Research Areas: Grand Challenges

The solution of grand-challenge problems is a the major objective of most of the agencies in the HPCC program. Grand challenges are fundamental problems in science and engineering, with broad economic and scientific impact, whose solution can be advanced by applying high-performance computing techniques and resources. These problems continue to tax any available computational and networking capabilities because of their demands for increased spatial and temporal resolution and increased model complexity. The fundamental physical sciences, engineering and mathematical underpinnings are similar for many of these problems. To this end, a number of multiagency collaborations are underway.

As new application software becomes more efficient, stable and robust, applications researchers are porting their software to the new parallel systems more quickly and are achieving faster run times. They are also obtaining more realistic results by taking advantage of the faster speeds, larger memory, and the opportunity to add complexity, which was not possible before the new architectures became available. These results are being obtained through the following:

- Higher resolution (for example, modelling a beat of the human heart at time steps of a tenth of a second instead of each second, and at every hundredth centimeter rather than every centimeter).

- Faster execution times (for example, models that took days of execution time now take hours, enabling researchers to explore a wider range of parameters and time scales—100-year climate models can now be executed in the same time it used to take for 10-year models).

- More realistic physics—for example, including in weather models physics that better model the effects of clouds.

- More realistic models—for example, one "multidisciplinary" model combining separate atmosphere and ocean models, or one combining "single discipline" air and water pollution models.

As grand-challenge applications prove the mettle of scalable parallel architectures, commercial software vendors are becoming more active in moving their software to these new machines. The success of commercial applications software is crucial to the success of the high-performance computing industry.

The various agencies, often in collaboration with industry, are addressing problems in the following areas.

*Aerospace*

Improved and more realistic models and computer simulations of aerospace vehicles and propulsion systems are being developed. These will allow for analysis and optimization of designs over a broad range of vehicle classes, speeds, and physical phenomena, using affordable, flexible, and fast computing resources. These applications are computationally unrealistic with traditional computing technology.

*Energy*

Grand-challenge projects are exploring the following:

- Mathematical combustion modelling.

- Developing algorithms on massively parallel machines for high-energy physics and particle physics applications (quantum chromodynamics).

- Oil-reservoir modelling—constructing efficient parallel algorithms to simulate fluid flow through permeable media.

- The numerical Tokamak project—developing and integrating particle and fluid plasma models on massively parallel machines as part of the multidisciplinary study of Tokamak fusion reactors (note that Canada has one of these reactors at Hydro-Québec's Institut de recherche en électricité du Québec [IREQ]).

*Environmental Modelling and Prediction*

The environmental grand challenges include weather forecasting, predicting global climate change and assessing the impacts of pollutants. High-performance computers allow for better modelling of the Earth and its atmosphere, resulting in improved weather forecasts and warnings, and improved global-change models.

Models of the atmosphere and the oceans are being rewritten in modular form and transported to several large parallel systems. A much greater level of detail is now possible—local phenomena such as eddies in the Gulf of Mexico are being modelled, thus allowing for better warning of weather emergencies and improved design of equipment such as oil rigs.

Separate air- and water-quality models are being combined into a single model which is being transported to a variety of massively parallel systems. These models are needed to assess the impact of pollutant contributions from multimedia sources and to ensure adequate accuracy and responsiveness to complex environmental issues.

In FY 1995, complexities such as aerosols, visibility, and particulates will be added; entire environmental models will be integrated into parallel computing environments with a focus on emissions-modelling systems and integration with geographical information systems.

*Molecular Biology and Biomedical Imaging*

FY 1994 accomplishments include the following:

- Development and field testing of the Diagnostic X-ray Prototype Network, a nationwide radiology research system.

- Implementation of an Internet-based system that enables integrated searching of DNA and protein sequences and the medical literature linked to those sequences.

- Usage of the Internet-based BLAST (Basic Local Alignment Sequence Tool) for advanced biological similarity searching reached 3 million queries per year.

- Improved understanding of heart function through the 3-D simulation of a single heartbeat (this required 150 hours on the fastest Cray and received a Smithsonian Award).

- First simulation of an entire biological membrane, including all lipid and protein components, allowing for better understanding of the mechanism of inflammation of tissues in diseases such as asthma and arthritis (collaborative work with Eli Lilly).

- New algorithms for constructive 3-D images from 2-D clinical images and micrographs.

- Assistance in the design of new drugs to inhibit HIV replication.

- Further elucidation of the structure of the herpes virus using high-performance computing to obtain 3-D images from electron micrographs.

- Identification of the 3-D structure of several proteins by using a genetic algorithm approach to automate the spectral assignment task in NMR spectroscopy.

FY 1995 plans include the following:

- Expand research in algorithms for real-time acquisition and processing of multi-model medical images for use in telemedicine and virtual environments for surgery.

- Develop new algorithms and software for receptor-based drug design; study basic mechanisms of receptor structure and function; and develop software to model specific populations at risk for disease.

- R&D in medical image processing for basic research, clinical research, and provision of health care.

- Support development of telemammography, addressing high-resolution, wide field-of-view displays and high-performance, low-cost networks for image transmission.

- Understanding human joint mechanisms through advanced computational models.

*Product Design and Process Optimization*

Beginning in FY 1995, NIST will develop new high-performance computing tools for applying computational chemistry and physics to the design, simulation, and optimization of efficient and environmentally sound products and manufacturing processes. The initial focus will be on the chemical-process and microelectronics industries, and on biotechnology manufacturing. New computational methods for first-principles modelling and analysis of molecular and electronic structure, properties, interactions and dynamics will be developed. Macromolecular structure, molecular recognition, nanoscale materials formation, reacting flows, the automated analysis of complex chemical systems, and electronic and magnetic properties of thin films will receive particular emphasis. NIST researchers and their industrial partners are developing methods for monitoring and controlling diamond turning machines to improve the precision and production of highly efficient optics such as mirrors for laser welders.

*Space Science*

NASA's Earth and Space Science projects support research in:

- Large-scale structure and galaxy formation;
- Cosmology and astrophysics;
- Solar activity.

NSF has funded grand-challenge applications groups to address fundamental problems in:

- Black-hole binaries;
- The formation of galaxies and large-scale structure;
- Radio synthesis imaging.

#### 4.1.2.7 HPCC Research Areas: National Challenges

In his state of the Union address on January 24, 1994, President Clinton called for connecting every classroom, library, clinic, and hospital in America into a "national information superhighway" by the year 2000. He said, "Instant access to information will increase productivity. It will help educate our children. It will provide better medical care. It will create jobs." The HPCC program is helping to fulfil this vision by developing much of the underlying technology for the NII, enabling the development of national-challenge applications, and demonstrating a variety of pilot applications. National challenges are fundamental applications that have broad and direct impact on the nation's competitiveness and the well-being of its citizens, and that can benefit from the application of HPCC technologies and resources. They include crisis and emergency management, digital libraries, education and lifelong learning, electronic commerce, energy demand and supply management, environmental monitoring and waste minimization, health care, design of manufacturing processes and products, and public access to government information. Some of the early national-challenge applications, such as electronic commerce, will mature to become services in the NII, enabling a wider range of future national-challenge applications.

National challenges to be addressed by the HPCC program in FY 1994 and FY 1995 include the following:

*Digital Libraries*

A digital library is the foundation of a knowledge centre without walls, open 24 hours a day and accessible over a network. The HPCC program supports basic and strategic digital library research and the development and demonstration of associated technologies. These technologies are used in all of the other national-challenge applications.

Beginning in FY 1994, the program will support the following types of R&D:

- Technologies for automatically capturing data of all forms (text, images, speech, sound, etc.), generating descriptive information about such data (including translation into other languages), and categorizing and organizing electronic information in a variety of formats.

- Advanced algorithms and intelligent interactive Internet-based tools for creating and managing distributed multimedia databases and for browsing, navigating, searching, filtering, retrieving, combining, integrating, displaying, visualizing, and analyzing very large amounts of information that are inherently in different formats. Research will also address standards that enable interoperability.

- Internet-based access to and distribution of remote sensing imagery and other satellite products from its geostationary and polar orbiting operational environmental remote sensing satellites, available to state and local governments, the agriculture and transportation industries, and to libraries and educational institutions.

*Crisis and Emergency Management*

Large-scale, time-critical, resource-limited problems such as managing natural and man-made disasters are another vital national challenge. Effective management involves the use of command, control, communications, and intelligence information systems to support decision makers in anticipating threats, formulating plans, and executing these plans through coordinated response.

*Education and Lifelong Learning*

HPCC support for this national challenge involves making HPCC technologies a resource for the nation's education, training, and learning systems for people of all ages and abilities nationwide. The NII approaches this challenge from several directions:

- Distance learning will bring specialized resources in a timely manner to geographically widespread students.

- Teacher training and coordination enhances the resources available to teachers at all educational levels.

- Students throughout the country will have access to information and resources previously only available at research and library centres.

- Lifelong learning provides educational opportunities to populations regardless of age or location.

- Digital libraries will make information available throughout the network—for both professionals and students at all levels.

*Electronic Commerce*

This national challenge integrates communications, data management, and security services to enable different organizations to automatically exchange business information. Communications services transfer the information from the originator to the recipient. Data-management services define the interchange format of the information. Security services authenticate the source, verify the integrity of the information received, prevent disclosure by unauthorized users, and verify that the information was received by the intended recipient. Electronic commerce applies and integrates these services to support business and commercial applications such as electronic bidding, ordering and payments, and exchange of digital product specifications and design data.

*Energy Management*

Oil consumption, capital investment in power plants, and foreign trade deficits all benefit from improved management of energy supply and demand. Beginning in FY 1994, the U.S. Department of Energy and the power utilities will document and assess the tools and technologies needed to implement the national challenge of energy supply and demand management. They will also document the expected economic benefits and identify policy or regulatory changes needed so that the utilities can participate in the deployment of the NII.

*Environmental Monitoring and Waste Minimization*

Improved methods and information will dramatically increase the competitiveness of U.S. companies in the world's $100-billion-per-year environmental monitoring and waste-management industries. Beginning in FY 1995, digital libraries of the large volume and wide range of environmental and waste information will be assembled and tools will be developed to make these libraries useful.

*Health Care*

Advanced HPCC communication and information technologies promise to improve the quality, effectiveness and efficiency of today's health-care system. This national challenge will complement the biomedical grand challenges. It will include testbed networks and collaborative applications to link remote and urban patients and providers to the information they need; database technologies to collect and share patient health records in secure, privacy-assured environments; advanced biomedical devices and sensors; and system architectures to build and maintain the complex health-information infrastructure. NIH is funding the following activities:

- Testbed networks to link hospitals, clinics, doctors' offices, medical schools, medical libraries, and universities to enable health-care providers and researchers to share medical data and imagery.

- Software and visualization technology to visualize the human anatomy and analyze images from X-rays, CAT scans, MRI scans, PET scans, and other diagnostic tools.

- Virtual reality technology to simulate operations and other medical procedures.

- Collaborative technology to enable several health-care providers in remote locations to provide real-time treatment to patients.

- Database technology to provide health-care providers with access to relevant medical information and literature.

- Database technology to store, access, and transmit patients' medical records while protecting the accuracy and privacy of those records.

*Manufacturing Processes and Products*

Advancing manufacturing through the use of HPCC technologies in the design, processing, and production of manufactured products is another national challenge. A key element is the development of the infrastructure technology and standards necessary to make processes and product information accessible to both enterprises and customers. Ongoing multi-year projects include the following:

- ARPA's MADE (Manufacturing Automation and Design Engineering) in America Program, will develop engineering tools and information integration capabilities to support future engineering and manufacturing processes.

- NASA efforts in both industry and academia in the multidisciplinary design of aeronautical airframes and aircraft engines will develop an integrated product/process development capability. This is intended to shorten the product-development cycle, maximize design capability, lower the life-cycle cost, and obtain new insight into and understanding of the advanced manufacturing process, all critical to more competitive airframe and propulsion industries.

- NIST's System Integration for Manufacturing Applications Program emphasizes technologies that support flexible and rapid access to information for manufacturing applications. It includes a standards-based data-exchange effort for computer-integrated manufacturing that focuses on improving data exchange among design, planning, and production activities.

- At NIST, an advanced manufacturing system and network testbed supports R&D in high-performance manufacturing systems and testing in a manufacturing environment. The testbed will be extended to include manufacturing applications in the mechanical, electronics, construction, and chemical industries; electronic-commerce applications for mechanical and electronic products; and an integrated

Standard Reference Data system. Initial focus will be on virtual and rapid prototyping.

*Public Access to Government Information*

This national challenge will vastly improve public access to information generated by federal, state, and local governments through the application of HPCC technology. Ongoing efforts include connecting agency depository libraries and other sources of government information to the Internet to allow for public access; and demonstrating, testing, and evaluating technologies to increase such access and effective use of the information.

### 4.1.2.8 NII and its Relationship to the HPCC Program

The previous section explained the HPCC national challenges that develop the underlying technology and prototype applications for the complementary National Information Infrastructure (NII) program.

HPCC and the NII are tightly interwoven. While the federal agencies that participate in the HPCC program are working to develop the NII, it is the private sector that will deploy it. Many policy and regulatory obstacles will be cleared through the workings of the interagency Information Infrastructure Task Force.

The capabilities being developed under the HPCC program provide much of the technology base critically needed for the National Information Infrastructure. The following is a list of several examples:

- Scalable computing technologies provide the foundation for the computing systems needed by the NII. For example:

- The microprocessors at the heart of today's most powerful scalable parallel computers can also be found in relatively inexpensive desktop workstations and current personal computers and will be found in the information appliances in the homes of tomorrow.

- The same technology that provides high-speed connectivity between computer processors and memory can be used to build switches for the NII's high-speed networks.

- The same client/server technology used to supply data from remote computers to grand challenge computations can be used to disseminate information to American households.

- NII applications such as the entertainment industry's "video on demand" will soon use the scalable parallel computing and mass-storage systems that have long been part of high-performance computing research centres.

- The technology for the high-bandwidth networks needed by grand challenge researchers to move data between computers and to distribute information to users of the NII is being developed at the gigabit testbeds (see next section).

- The software technologies, including operating systems and software development environments, being developed under HPCC have widespread applications to NII systems, not just those identified as "numerically intensive".

- Many national-challenge applications depend upon computationally intensive grand-challenge applications. For example:

  - Improving health care requires biomedical research (such as improved molecular design of drugs).

- Improved environmental monitoring requires computationally intensive environmental models.

- Effective response to natural and man-made disasters depends on weather forecasting and environmental models.

- Developing advanced methods for product and process design and manufacturing requires numerically intensive prototyping of those products and processes.

- The information in the National Information Infrastructure will include the data used by the grand challenge research as well as the research results.

The NII places additional demands on HPCC research and development. Scalability issues must be addressed—the technologies used to serve thousands of researchers and educators must be scaled up to serve millions of NII users. With the larger number of NII users and their greater diversity of interests and skills, more ubiquitous alternative "on-ramps" to the emerging high-bandwidth networks of the NII must be examined.

New services are being developed to support multimedia applications, digital library technologies, appropriate privacy and security protection, and higher performance levels. Software development environments and building blocks for interfacing humans and computers must be created. These new tools will enable application developers to construct complex, large-scale, network-based, user-friendly and information-intensive applications. These technologies provide the foundation upon which national-challenge applications will be developed.

#### 4.1.2.9 High-Performance Networks

New technologies are needed for the new breed of applications that require high-performance computers and that are in demand by users across the U.S. These technologies must move more information faster in a shared environment of perhaps tens of thousands of networks with millions of users. Huge files of data, images, and videos must be broken into small pieces and moved to their destinations without error, on time, and in order. These technologies must manage a user's interaction with applications. For example, a researcher may need to continuously display on a local workstation output from a simulation model running on a remote high-performance system in order to use that information to modify simulation parameters.

For over two years, the HPCC program has provided a national network backbone operating at T3 speeds (45 Megabits/sec). Many research programs are now experimenting with gigabit networks that operate at billions of bits per second.

As these gigabit-speed networks are set up, the current barriers to more widespread use of high-performance computers will be surmounted. At the same time, high-speed workstations and small and mid-size scalable parallel systems will gain wider use.

A teraflop (a trillion operations per second) computing technology base needs gigabit-speed networking technologies.

HPCC-supported gigabit testbeds, funded jointly by NSF and ARPA, test high-speed networking technologies and their application in the real world. The HPCC program is developing a suite of complementary networking technologies to take fullest advantage of this increased computational power. R&D focuses on increasing the speed of the underlying network technologies as well as developing innovative ways of delivering bits to end users and systems. These include satellite, broadcast, optical, and affordable local area designs.

The HPCC program's gigabit testbeds are putting these technologies to the test in resource-demanding applications in the real world. These testbeds provide working models of the emerging commercial communications infrastructure, and accelerate the development and deployment of gigabit speed networks.

In FY 1994-1995, HPCC-funded research is addressing the following:

- ATM/SONET (Asynchronous Transfer Mode/Synchronous Optical Network) technology—"fast packet-switched" cell relay technology (in which small packets of fixed size can be rapidly routed over the network, integrating multimedia information) that may scale to gigabit speeds.

- Interfacing ATM to ultra-high-speed computer interfaces and switches, to make heterogeneous distributed high-performance computing systems available at high network speeds.

- All-optical networking.

- High speed LANs (Local Area Networks).

- Packetized video and voice and collaborative workspaces (such as virtual reality applications that use remote instruments).

- Telecommuting.

- Intelligent user interfaces to access the network.

- Network management (for example, reserving network resources, routing information over the networks, and addressing information not only to fixed geographical locations but also to people wherever they may be).

- Network performance measurement technology (to identify bottlenecks, for example).

- Networking standards (for interoperability) and protocols (including networks that handle multiple protocols such as the Internet-standard TCP/IP, the evolving International Standards Organization OSI, and popular proprietary protocols).

### 4.1.3 <u>High-Performance Computing in Industry</u>

#### 4.1.3.1 Private Industry Use and Applications

The primary objective of most HPC programs in the U.S. is to ultimately improve the competitiveness of U.S. industry. Even for the programs that support basic research, there is an increasing expectation of results that will eventually benefit industrial competitiveness. (Having made this general statement, we do not wish to diminish the importance of certain HPC programs that are directed towards other societal benefits, such as improving the health-care system, combating disease with more effective remedies, developing more effective programs for cleaning up the environment and improving the efficiency of government operations).

There are several hundred HPC systems installed in various industrial companies for their exclusive use in the design and development of products, processes and services. It is difficult to obtain precise count, however, as each survey uses a different definition of HPC.

The following is a brief summary of a few major industrial application areas of HPC in the U.S. This is far from an exhaustive list, as HPC is pervading so many industrial application areas that it is not possible to cover the full spectrum in this introductory report.

*Aerospace*

Civilian and military aircraft and space vehicles increasingly take advantage of HPC capabilities during all aspects of the design and development process in order to produce better designs and to reduce the time required in introducing new products. The most significant application is the aerodynamic design and analysis of airframes, where exceedingly complex computational fluid-dynamics models, which can only be simulated on very powerful HPC systems, are increasingly capable of replacing expensive time-consuming wind-tunnel experiments. A second area where HPC systems are routinely used by aerospace engineers is in the structural design of the airframe and various components, including engines.

As a result of improved numerical methods of designing aerospace vehicles by means of HPC technology, designers have been able to accelerate the attainment of important objectives, including weight reduction, higher performance, improved fuel-efficiency, larger higher-thrust engines, reduced development costs, higher quality and shorter design times.

*Automotive*

Automotive designers are routinely using supercomputer models in ways that parallel those in the aerospace industry. Again, computational fluid-dynamic models are used to reduce drag to design more efficient cars, structural analysis is used to improve the safety of automobiles that are lower in weight and cost and, analogous to the replacement of wind-tunnel testing in the aerospace industry, automotive engineers are able to reduce the time and cost involved in prototype wall-crash testing by simulating crashes on the computer.

Automotive suppliers are likewise using HPC systems to design various sub-components to reduce cost, improve quality and contribute to overall fuel efficiency by reducing weight through use of lighter materials.

*Engines*

For the foreseeable future, the largest percentage of energy needs will be met by the combustion of fossil fuels. Two of the largest uses of fossil fuels are electric power generation and automobiles. Computational models offer a means of improving the design and efficiency of internal combustion engines. One of the grand challenges requiring an inordinate amount of HPC resource is the modelling that needs to be carried out in order to determine the effects of emissions in the atmosphere. This modelling requires complex models that consider the interaction between fuel, the vehicle engine and the atmosphere, including several hundred chemical interactions that are known to occur in internal combustion engines.

*Microelectronics*

With semi-conductor circuits now exceeding one million transistors per chip, it would be inconceivable to design these circuits without the use of HPC systems, especially in light of how rapidly the technology is evolving and the challenge of remaining competitive. Today, HPC models are mandatory for the intricate interconnection design and logical testing within a chip, design and testing external interfaces, design of packaging such as multi-chip modules and simulation of entire electronic systems.

At a more basic research level, supercomputers are used to evaluate new semiconductor materials, which will eventually become more cost-effective replacements to silicon, at least in certain specialized applications.

*Petroleum*

As explained in Section 2 with reference to Canada, petroleum and seismic processors were among the early adopters of HPC, since computer modelling has obvious cost benefits. The major application is seismic analysis, which provides interpretation from seismic exploration surveys of the potential reserves underground. As HPC systems become more powerful, the petroleum/seismic industry is being quick to improve the accuracy of its models. At the same time, there is a corresponding demand for more computer power in order to provide more precise indications of these reserves. For example, these models have been substantially improved in recent years by moving from 2-D to 3-D analyses. The real payoff is the reduction of the number of "dry holes" that are drilled through more accurate determination of where the hydrocarbons are located below the surface. A single marine dry well can represent a $50-million write-off, which itself is enough to pay for the most powerful supercomputer.

A second application is reservoir simulation, which enables a petroleum company to better manage the operational recovery phase of its operations.

*Pharmaceutical*

Today, many major pharmaceutical companies rely upon supercomputers to save millions of dollars on the design of new drugs by using molecular modelling techniques to supplement traditional laboratory experimentation. Considering that each new drug developed typically costs $150 million and can take up to ten years, and that hundreds of millions of dollars can be received from sales of a new drug in its first year on the market, it does not take very much speedup in the design process to justify the cost of supercomputing. Supercomputers and numerical experimentation are used very effectively at the front-end of drug design to help better focus on molecular interactions of interest, thereby reducing the number of possible considerations to the most relevant ones, which nevertheless still require verification in the laboratory.

Advancement in molecular computational methods coupled with the power and flexibility of general-purpose supercomputers is accelerating the development of more effective pharmacological therapy for cancer and AIDS. The new methods are improving the chances of developing an actual cure for these diseases, which currently resist therapies of minimal toxicity.

*Metals*

Manufacturers of base metals such as steel and aluminum have been using HPC for a variety of applications, including improving the metal production process itself, designing new improved alloys and composite materials, and designing new products.

As just one example, Pittsburgh-based USS has turned to supercomputing at the nearby Pittsburgh Supercomputing Center as part of the industry's drive toward the technological leading-edge. In the 1980s, foreign steel and new materials cut deeply into the markets served by North American steel, and the industry in the 1990s is responding with a strong effort to modernize its production facilities.

An important application of the supercomputer at USS is investigating ways to improve the process-control systems used in steel manufacturing. There has been a virtual explosion of intelligent devices, sensors and actuators directly connected to the processes and networked to higher-level process-control computers. These devices can be mathematically modelled not only for off-line simulation studies, but also and even more importantly, in the core of process-control software. Developing more accurate, versatile computer models is one of the keys to producing high-quality steel products at a competitive cost.

#### 4.1.3.2 Applications of Parallel Information Servers

While HPC has traditionally been the private domain of scientists and engineers, the technology under development is beginning to see its way into non-technical information applications. The information highway initiatives linked to the HPCC program are a good example, as explained above. The parallel computers developed initially for large scientific problems can also be used today for the very large-scale information servers that will be required to deliver information to the multitude of individuals that are connecting to the highway. These ultra-large servers require the massive data-storage capacities and the rapid simultaneous access capabilities offered only by massively parallel systems. Furthermore, the gigabit networks, also developed and tested within the HPCC program, will be required to connect the many users to the networks.

The business community itself is starting to use massively parallel information servers for commercial applications. While the first users are pioneering the development of new applications, many industry observers believe that widespread software development will happen sooner rather than later and that within three or four years MPP systems will be used much more widely for commercial than scientific applications. As an indication of the fast-growing interest in parallel information servers, the largest data-management software company in the U.S. has informed us that it has a waiting list of 400 organizations that have requested benchmark demonstrations of their parallel server product over the last several weeks. A few U.S. examples will demonstrate the advantages of parallel information servers.

- Prudential Securities is using several parallel computers to test hundreds of separate hypotheses for investment strategies for mortgage-backed securities. This application did not run fast enough on conventional computers, since market rates could change several times in the 5 to 11 minutes it took to provide answers. With the parallel systems, the same analysis can be done in 30 seconds, i.e. close enough to "real time".

- At Allstate Insurance, an increase in the number of claims was pushing the company's mainframe to the limits of its capabilities. A parallel computer was installed, at much less cost than a mainframe upgrade and with much better response times, translating to improved customer service.

- At Hallmark Cards and K-Mart, the tedious task of searching through huge databases is being carried out on parallel systems. The firms claim that large retailers can truly benefit from this approach, as they must collect large amounts of data from hundreds of stores and other distribution channels (normally collected through point-of-sale cash registers). With MPP technology, it is far more cost-effective to sift through this massive data to look at combinations of products, stores, seasonal sales and other criteria in order to develop business strategies.

- Booze, Allen & Hamilton, a management-consulting firm, purchased its own MPP system to develop massively parallel software and to convert existing applications to an MPP system for its corporate clients. For example, they are helping organizations in the health-care, financial services, insurance and telecommunications sectors (all data-intensive) to reduce the substantial losses from sophisticated fraud schemes. While these organizations often have the raw data about both legitimate and fraudulent transactions, they lack the technology to separate and identify transactions and develop patterns of interest. Booze Allen has developed MPP tools that facilitate detailed analysis on extremely large databases.

#### 4.1.3.3 Benefits to Industry

In the sections above, descriptions of various industrial applications are accompanied by explanations of the benefits of each of these applications in qualitative terms. We shall now attempt to measure these benefits in quantitative terms.

In 1991, the U.S. Department of Energy commissioned the Gartner Group to study the economic impact of implementing the initial HPCC program, which at the time represented a commitment to increase investment by $1.9 billion over a five-year period, over the base spending of approximately $500 million per annum. The study analyzed the impact of this incremental investment.

The quantification process began by analyzing the productivity improvements that had been realized by early adopters of HPC and then, through a series of interviews with industry executives, estimated the expected increases in productivity with and without the new HPCC program. The analysis concentrated on sectors where the benefits of HPC are relatively well understood, i.e. aerospace, chemicals, electronics, petroleum and pharmaceuticals. In each sector, a low-end conservative productivity number was derived and a high-end optimistic figure. The annual productivity increase projections ranged from a low of a fraction of 1% for pharmaceuticals (where HPC usage was just beginning, so executives were very conservative in their impact estimates) to a high of 1% (conservative) to 5% (optimistic) for petroleum, where HPC usage is mature and the executives are very bullish about even much more positive impact.

Although the individual percentage increases in productivity are small numbers, their cumulative effect on the national economy is very significant. Within the companies themselves, R&D is highly leveraged so the impact within the industry can be quite dramatic. Also, because of the interlocking structure of various industrial sectors, an improvement in one sector can spill over to many other related sectors. These productivity improvement parameters were fed into an input-output econometric model (the University of Maryland's Long-term Inter-industry Forecasting Tool) to study the impact on the national economy.

The economic predictions of the model, forecast over the next decade in constant dollars, based on this $1.9 billion incremental HPC investment, were most impressive:

| **Indicator** | **Conservative Case** (in billions of $) | **Optimistic Case** (in billions of $) |
|---|---|---|
| Increased gross national product | 172 | 502 |
| Increased gross private domestic investment | 57 | 199 |
| Increased gross exports | 8 | 31 |
| Decreased federal deficit | 75 | 190 |

While this analysis took into consideration only the five major sectors using HPC at the time, an even greater impact is to be expected, though perhaps not until the year 2000, in new application areas such as finance, transportation, construction and entertainment.

Presumably the U.S. government must have believed that the economic impact of an HPCC program would be substantial, as the full program was approved and passed as law after the Gartner study was submitted.

#### 4.1.3.4 Collaborative Partnership Programs with Institutions and Industry

There is a very healthy collaborative relationship in the U.S. between industry, government and academia in the HPC sector. The federal government participation stems primarily from HPCC program national research centres and the national labs such as Lawrence Livermore, Los Alamos, Argonne and Sandia. The state government participates through state-supported HPC centres. The universities participate primarily through NSF programs, including their involvement in the HPCC program.

Two examples will illustrate the magnitude and far-reaching effects of some of these U.S. programs. First we will provide an insight to the Industrial Partnership Program at the Illinois HPCC Centre and then explain the important technology transfer program (CRADA) at the national labs.

*NCSA Industrial Partnership Program*

As one of the most successful collaborative HPCC programs is the Industrial Partnership Program at the National Centre for Supercomputer Applications (NCSA, the NSF Centre in Illinois), we provide a few interesting comments made by the director of this program in a recent interview. We shall focus on this example, as it seems to capture the essence of a successful collaborative program in HPC.

- The industrial program at NCSA has been enormously successful, resulting in over $50 million of commercial revenue for the Centre.

- Twelve major corporations have signed up as corporate partners, including:

| | |
|---|---|
| AT & T, Inc. | Eli Lilly & Co. |
| Caterpillar, Inc. | FMC Corp. |
| Dow Chemical Co. | Motorola Inc. |
| Eastman Kodak Co. | Phillips Petroleum Co. |
| Amoco | Goodyear |

- The partners have considered the program to be a major financial and research success. The payoffs have been impressive.

- The results and benefits of the program have been featured on six U.S. Television programs. For example, Caterpillar has explained how HPC has totally revised its approach to research and product development, resulting in a forecasted saving of $200 million in next year's budget. Dow claims that its investment at NCSA has paid back 100-fold.

- NCSA targeted certain key industries at the outset (e.g. heavy equipment manufacturing, pharmaceuticals) and they have stayed with this plan.

- To sign up a new major partner has taken considerable time and effort; the minimum time so far has been 14 months.

- Corporate partners typically commit $4 to $5 million when they initially join the program.

- NCSA management believes that a clear mission statement of an HPC centre is required before approaching industrial partners. The cornerstone of NCSA's mission is to improve American industrial competitiveness. This implies the need for corporate partnerships in order to develop new approaches to applied R&D that put the member companies in leadership roles of their respective industries.

This is quite different from the missions of most other centres, where the objective is to offer the lowest-cost computer time. In the mission, a centre must determine if it is to be a leader or a follower. NCSA has decided to be a leader, on the forefront of HPC technology and its applications. They emphasize that it is very expensive to maintain this leadership role. Maybe certain centres can only afford to be providers of computer time.

- NCSA stresses importance of working with company executives to show them the impact of the HPC partnership on their company results. Companies are interested, as always, in return on investment, improvement of their bottom line, and being ahead of their competitors. They are not interested in the latest kind of computing platform, nor do they particularly care about the underlying technology that solves their problems.

- The Development of contracts, client by client, is a very long, arduous process. NCSA is asking clients to develop a partnership to work on mission-critical activities, including the development of new R&D methodologies that will determine how competitive their organizations will be in the future. The projects they work on at the centre have major strategic value. These companies have never in the past gone so far outside their own four walls to conduct such research. Consequently, months are devoted to negotiating contracts that give absolute protection and security to this proprietary research. Licensing, copyrights and patent rights are major issues. The centre must be very sensitive and flexible in this regard. Many other centres have failed in their industry programs when the management does not understand the importance of or even existence of proprietary research.

- Many of the largest corporations are nervous about making a major long-term commitment to an HPC vendor these days. This hesitation is not so much related to the uncertainty of the technology, but more to the uncertainty about which vendors will be the important HPC providers in the future. In this environment, centres like NCSA have an excellent opportunity to convince clients that they should leave these difficult technology decisions to the large centres.

- There is an appreciation that in Canada, companies are much smaller and much less R&D is carried out. However, there are still ample opportunities to demonstrate to smaller companies how they can become much more world-competitive by embracing HPC strategies. If they have smaller R&D budgets, they must spend their limited funds more wisely, so it is up to centres like NCSA to explain and demonstrate the high-pay-backs from supercomputing applications.

- Training is an ongoing critical component of the NCSA product. Corporate partners send staff, including executives, for regular training sessions, sometimes for weeks at a time. This is a continuous process as technology advances. Training is offered at all levels. Even CEOs attend two-day sessions at least once a year to learn more about the strategic importance of HPC. The centre absolutely needs their support for ongoing commitments. Equally important is training for customers' internal management, as it is important for the them to learn how to manage the new HPC environment.

- Most application development projects are managed and staffed by the client because of their highly proprietary nature. Many of the staff work at the centre to have good access to equipment and support. There are also many collaborative projects (involving customer personnel and centre experts) in progress at all times. At the time of the interview, there were 42 such projects in progress.

- Another important component of the product they sell is support staff, as there is still a severe shortage of trained computational scientists in industry. There are 40 resident scientists on staff to give expert guidance to clients in all the major disciplines on which they have decided to focus.

- Another element of the NCSA package is visualization. They even use this as a marketing tool for CEOs, as they are able to show them how visualization takes the risk out of their decisions that are made from simulation results.

*National Labs CRADA Program*

The U.S. national laboratories are world-class research facilities, supported by the Department of Energy. The Lawrence Livermore laboratory alone has a research budget in excess of $1 billion, employs 8000 people and has a $4-billion-dollar plant that houses a supercomputer equipment that is the envy of the world. Whereas the original mission of theses labs was to conduct basic and applied research for national defense purposes, with the end of the cold war, these labs are devoting more attention to civilian industrial applications. They all have developed technology transfer programs that take many forms, encouraging industry to adopt their technology or collaborate on research through a number of mechanisms including: licensing, partnerships, sponsored research, employee exchange, consulting and training. All the labs are intensively involved in the development of hardware and software for HPC.

The most important partnership program to motivate the labs to work with industry is the Cooperative Research and Development Agreement (CRADA), a cost-sharing agreement under which the labs and the company contribute personnel, equipment and facilities to a cooperative effort. CRADA overcomes the proprietary research concerns of industry by withholding sensitive data for up to five years.

The largest announced CRADA program, related to HPC R&D, is a $52-million technology transfer program led by Cray Research in collaboration with Los Alamos, Lawrence Livermore and 15 leading U.S. industrial firms. This is a cost-sharing program to develop scientific and commercial software for MPP systems. The program includes 15 projects in fields such as environmental modelling, petroleum exploration, materials design, advanced manufacturing and new MPP systems software. Participating companies include Alcoa, Amoco, AT&T, Biosym, Boeing, Exxon, Hughes, Schlumberger and Xerox. The goal of this program is to boost industrial competitiveness and advance science by exploiting the inherent power of the new MPP systems.

### 4.1.4 Opportunities for Canadian/U.S. Collaboration

During this study, the President of Super*Can (the Canadian Association for High Performance Computing and Communications), Elizabeth Pearce, visited executives of the National Science Foundation in Washington, D.C. These executives are responsible for NSF participation in the U.S. HPCC program. The key spokesperson was Dr. Robert Borchers, NSF director of Advanced Scientific Computing.

Dr. Borchers was very interested in further exploring the possibilities and ramifications of Canadian/U.S. collaboration in HPC. He stated that there is already a well-developed collaboration among university computer science departments in the two countries. He could visualize a more extensive relationship whereby Canada becomes the first international extension of the U.S. Metacentre Program. The metacentre is essentially the networking of HPC facilities and human resources such that a "virtual machine room" is made available to users throughout the country, with access to technical support on the network regardless of the physical location of the technical support staff. He believes there is a growing interest in extending this concept internationally and sees Canada as the first country to consider.

In terms of sharing computer resources, some Canadian university researchers have already made their own arrangements to access U.S. HPC facilities. However, these are very informal arrangements, subject to termination at any time. Also, there are few or no support services offered in these ad hoc arrangements and there are no provisions for high-speed network access. A formal agreement with NSF may open the possibility to Canadian researchers, at least in the universities, to have access, with some support services, to the high-end equipment that is not available in Canada. The U.S. researchers, on the other hand, apparently have access to all the computer facilities they require; however, there appears to be an interest in gaining access to Japanese supercomputers located in Canada, as U.S. policy prohibits acquisition of Japanese supercomputers by government and universities.

Dr. Borchers indicated that by statute, the U.S. cannot fund research outside its borders, but this does not preclude collaboration. The NSF has representatives stationed in London and Tokyo, and the question was raised as to the possibility of a representative in Canada.

Independently of the NSF discussions described above, the President of WurcNet (the Western University Research Consortium) has initiated discussions with directors of the NSF centres in Urbana-Champagne, Illinois (NCSA), and in San Diego, California (SDSC). These discussions have centred around the possibility of linking the Canadian ATM test networks (the CANARIE linkage of regional ATM networks, including WNet of WurcNet) to equivalent ATM networks in the NFS gigabit test-network initiatives. This would provide very high-speed communications between all major HPC facilities in North America, thus providing the computing and network infrastructure for the collaborative efforts discussed with Robert Borchers. To quote from Larry Smarr, Director of NCSA in Illinois, "My concept was to begin a metacentre planning process with first Canada and then Mexico to construct a NAFTA type ATM testbed, built on an enlarged version of vBNS and NAPs [two U.S. ATM test networks]. It will take a few years in my guess, but it is a good way of enlarging metacentre activities in new ways and setting the information infrastructure standards for the largest trading zone in the world."

There is much follow-up required to these discussions, and the question to be answered first is Who will represent Canada for these high-performance computing and communications international negotiations?

#### 4.1.5 Comparison of HPC in Canada and the U.S.

There are several factors which have led the U.S. to its dominant position in both the development and usage of HPC systems. Initially, in the 1970s, it was understood that supercomputers were necessary tools to maintaining U.S. dominance in strategic weapons, especially nuclear, and in space programs (also controlled by the military to a large extent). There were no equivalent stimuli in Canada to encourage the adoption of supercomputers.

As a result of the military support of R&D in supercomputing, the universities began to understand how HPC could benefit basic research. In fact, a strong case was made that while all supercomputers were manufactured in the U.S., the only machines installed for university access were in Europe. By the late 1980s, the NSF Supercomputer Centres were approved and university access to HPC became widespread as NSFNet unfolded. In Canada, we started down the same path with the provincial government support of supercomputer centres in Ontario in Alberta, both in the late 1980s. However, while the U.S. centres flourished with continuous federal government support for technology renewal, the two Canadian centres ceased operations in 1991 because of an absence of ongoing provincial funding and no federal government support whatsoever. Both centres were technically obsolete when they shut down for lack of funding, while the U.S. had meanwhile opened dozens of state and national centres with annual programs of technology renewal. The eventual establishment of the national HPCC program in the U.S. has placed the country in the forefront of all current HPC technology developments.

In the U.S., it was only after the establishment of healthy university HPC centres that industrial partnership programs began to emerge for the diffusion of the technology to the private sector. After several years of education, training and industry outreach programs, some U.S. companies developed important R&D partnerships with state and national HPC centres, while others eventually developed sufficient justification to install their own supercomputers. Canada has not graduated to this stage of technology transfer to industry and at present there is no infrastructure support to enable the charting of such a course.

When one observes the high level of industry acceptance of HPC in the U.S. as compared to Canada, the following questions should be examined: Why were there so few industry users of the previous centres in Ontario and Alberta, and why is it so difficult to find industrial users for the new Alberta HPC Centre? While we hear many reasons for this situation, three principal explanations are most often provided:

- There was never sufficient attention or funding to address industry's need for awareness programs, education, training, development assistance, etc.

- Industry needs a long-term guarantee of availability of service before it will invest any efforts to develop HPC applications; with the sporadic funding of previous centres, there has been no confidence by private industry in the longevity of the centres.

- Canada spends only 1.5% of GDP on R&D compared to 2.75% in the U.S.; furthermore, the type of R&D that is best suited to HPC solutions is mostly performed at corporate R&D facilities, which in very many cases means the U.S. parent of Canadian companies.

In summary, the following lessons can be learned by studying the U.S. experience:

- A long-term commitment to funding and technology renewal is essential for building a national HPC program.

- A multi-agency coordinated program promotes collaboration by all sectors, synergy between partners, industry partnerships and technology transfer.

- Knowledge and experience is first developed in the universities and government labs, and then transferred to industry through industry partnership programs.

- As individual organizations, especially small and medium-sized businesses, cannot justify their own dedicated HPC facilities, it is important to provide infrastructure support with government assistance for the establishment of shared centres; even very large organizations begin building knowledge and experience at such centres.

- Interconnected high-speed networks open doors to many new innovative applications at many new organizations.

- Awareness programs, training, education and industry outreach programs are critical to success.

- The technologies developed in an HPC program can be effectively applied to new information highway applications; HPC and information infrastructure initiatives should be tightly coupled.

## 4.2 Europe

### 4.2.1 Background

Europe was one of the early adopters of HPC technology, but has never been a significant developer of HPC systems. In other words, European countries have learned to effectively use HPC to enhance science, improve industrial competitiveness and address some of society's complex problems in areas such as the environment and combating disease. In fact, when vector supercomputers were first introduced into the market in the late 1970s, it was the Europeans that first acquired several large machines for installation in centres that could be shared by industry and the academic community. Most U.S. acquisitions were for government labs that were off limits for industry and academia at the time. Consequently, U.S. researchers went on to Europe to do their supercomputing work on machines that were manufactured in the U.S. The U.S. HPCC program described in Section 4.1 was designed to redress this anomaly. Today, Europe has a community of leading-edge users of HPC equipment, which accounts for an estimated 25% to 30% of the use of these systems world-wide.

The Europeans have nevertheless left most of the development and manufacturing of the underlying HPC technology to the U.S. and Japan. With the advent of parallel systems in recent years, there has been some change in this strategy as we observe the emergence of a very vibrant software industry for the exploitation of parallel systems and we have even seen the emergence of a small number of parallel hardware companies that have introduced world competitive products. The Europeans seem to accept that they have totally missed out on any participation in the development of traditional vector supercomputers and even the downsizing to workstations, but they predict a major shift to parallel technology over the next decade and are committed to playing an important global role in this revolution.

Today, Europe's strength in HPC is that it has developed a strong user base and relatively high awareness of benefits; moreover, there is a good nucleus of a network of R&D activity, especially with respect to the development of tools and environments for the use of parallel systems, and there is an excellent level of education in science and technology by world standards. Europe suffers, however, from a fragmentation of efforts because of the many separate nations in the European Community (EC), a history of not participating in the supply side of HPC, the absence of the entrepreneurial culture that fosters the creation of start-up companies and the absence of an adequate affordable pan-European telecommunications network to support cooperative HPC activities. Realizing the potential benefits of a European cooperative effort, the EC launched the High Performance Computing and Networking (HPCN) program in 1993, following several years of studies and recommendations by teams of experts from the participating countries.

### 4.2.2 High-Performance Computing and Networking (HPCN) Program

#### 4.2.2.1 Overview

The EC has recently embarked upon a major program, the HPCN program which is aimed at the widespread implementation of high-performance computing and networking technology, and concentrates on the development and use of parallel systems. While there was significant pressure by some EC member countries to have an equal balance between the development of both hardware and software, the final consensus was to focus all attention on the software and applications aspects of high-performance computing and networking.

This emerging program is based on a very detailed study and resulting recommendations by the EC Advisory Committee on High Performance Computing and Networking. This committee, with representation from industry, government and academia from all the member countries, presented its final report in October 1992. The program is proceeding along the guidelines presented in the committee's very comprehensive report.

The goals and objectives of HPCN are very similar to the U.S. HPCC program, with the exception of the U.S. goal to develop future generations of high-performance hardware platforms. The emergence of parallel computing systems, coupled with broadband high-speed networks, offers the potential for a massive increase in performance, which is essential for meeting the increasing demands of industry, universities and government to address the competitive pressures on a global scale. This program is expected to have a major far-reaching impact on productivity, industrial competitiveness, environmental management and many other aspects of society in general. The program will be driven by user needs and will be industry-oriented.

The EC began soliciting proposals in 1993 for the first phase of the HPCN program, which has received funding of $100 million European currency units (ECUs). As this is one of the EC's fastest growing programs in science and technology, it is expected that funding levels will increase to $1 billion ECUs per annum by 1995, putting this program on a scale equivalent to the U.S. HPCC program, but with a somewhat different focus. The EC would supply approximately 20% of the total funding. The program will run for 10 years, with constant reviews and monitoring along the way. Again the concentration will be on software and applications, compared to U.S. and Japanese programs which also invest large sums toward the development of hardware platforms. The EC believes that it is not realistic to attempt to catch up with the U.S. and Japan in the development of semi-conductors and hardware systems, and that the major criterion for success will be the development of software systems that can truly take advantage of the latent potential of parallel processing, including MPP systems, clustered workstations and heterogeneous distributed processing.

According to Jean-François Omnes, Deputy Head of HPCN, this program will focus attention in an area where Europe is already a world leader, i.e. implementation of parallel systems. Industry, universities and government have been installing hundreds of parallel systems over the past decade, and there are thousands of people who are well trained in the use of this technology. There is a widespread diversified effort to embed parallel systems deep within the European industrial complex, and many more systems are being installed for experimental purposes in the universities.

The EC program would like to work closely with other world-wide efforts to develop parallel computing standards so that users do not have to reinvest time and energy every time a better architecture hits the market. The program members will also work closely with application software companies to port existing applications to parallel systems for higher performance at lower cost to industry.

#### 4.2.2.2 Goals and Strategies of the HPCN

The two goals of the HPCN program are as follows:

- To ensure European competitiveness via the exploitation of HPCN.

- To improve the opportunities for the European IT supply industry by creating domestic market conditions comparable to those of its competitors; the focus will certainly be on software development.

The achievement of these goals requires collaboration among users in all sectors, suppliers, network operators, educational and training institutions, researchers, investors in high technology, and EC-level government organizations.

The following is a list of some of the essential elements of the unfolding strategy:

- produce significant industrial, commercial, scientific and social applications on HPCN systems (note that the sole emphasis is no longer scientific applications);
- ensure the general availability of advanced and productive HPCN systems, including the establishment of world-class regional centres;
- develop a closely linked community of innovative HPCN users;
- develop education, training and awareness;
- provide adequate and affordable pan-European network services.

From these goals and strategies, the program will be implemented via collaborative projects driven by user needs, paying special attention to industrial requirements and selecting from competitive proposals. The program is to run for ten years, which is considered sufficient time to establish long-term infrastructures that can become economically self-sustaining.

#### 4.2.2.3 Lines of Action for the HPCN

The HPCN committee has defined four synergistic lines of action to maximize the effectiveness of the program, as described below.

*i) Applications Support and Development*

The aim is to promote real-world applications to form the basis of future exploitation and to provide both a proof of concept and a demonstration of benefits. There are three elements to this line of action, as follows:

*Existing Applications*

First there will be programs to convert existing applications to the new parallel platforms in such a manner that they will be scalable to larger systems in the future and portable to new architectures as they emerge. The intent is to approve programs and funding for six major applications (each consuming an average of 50 person-years of effort over four years and $5 million ECUs of HPCN facilities) and 12 small applications (each consuming an average of 10 person-years over two years and $1 million ECUs of HPCN facilities). It is expected that this program will build confidence in and acceptance of parallel HPCN, generate experience in use of the technology and stimulate innovative ideas for future applications.

*Novel Applications*

Secondly, there will be programs to identify and develop novel applications of HPCN technology. These programs will entail the following:

- basic research in simulation and modelling;
- challenging high-risk but high-payoff applications;
- development of underlying algorithms needed to support the new applications;
- development of new optimization, visualization and data-management techniques required by the new applications.

It is expected that these programs will stimulate interest in the use of HPCN in areas where it has not traditionally been applied.

*Development of Tools*

Thirdly, this line of action will promote the implementation of user-friendly and productive environments for the exploitation and development of parallel and distributed systems. This will include the development of tool sets for the many software challenges that are barriers to the use of parallel technologies by application developers today. These will include parallel languages and compilers, migration tools, optimization tools, and performance-monitoring tools, all integrated in a user-friendly environment. These environments are essential in order to increase confidence in application developers for parallel platforms, to increase their productivity and to promote collaboration in the development process.

*ii) Research and Development and Demonstration of Key Technologies*

The aim of this line of action is the development of new HPCN technologies and the demonstration of their potentialities for solving real problems. It consists of two elements:

*R&D in Key Technologies*

This element is closely linked to the tool-development element of the first line of action, described above. This is the R&D component that will develop and integrate the basic technological and scientific knowledge necessary for the HPCN, for example:

- novel application algorithms for optimization methods, simulated annealing, genetic algorithms, dynamically adapting systems, etc.;
- software and hardware interfaces for gigabit/sec networks into HPCN systems;
- visualization and multi-media human interfaces for HPCN systems.

One key objective of this element is to consolidate and expand the efforts of the many existing R&D activities throughout Europe.

*Technology Demonstrators*

This element aims at demonstrating the possibilities of new technologies and applications, including those under development in other parts of the HPCN program.

Advanced hardware and software testbeds will be developed to test and demonstrate new technologies such as:

- cost-effective performance scaling beyond trillions of operations per second;
- gigabit networking and HPCN interfaces;
- massive data management and visualization, utilizing parallel systems.

This will be the European testing ground for the products resulting from the HPCN program.

*iii) HPCN Supporting Infrastructure*

The aim of this line of action is the creation of a pan-European HPCN infrastructure, consisting of facility and support nodes and a broadband interconnecting network. The line of action consists of two elements:

*Networking*

This element is devoted to R&D and deployment of wide area network connections for remote interworking between HPCN systems and user access. The importance of this initiative is that it will provide easy affordable high-speed access from each user site to major HPCN facilities and expertise, often concentrated in a few sites. Furthermore, it will encourage and facilitate the cooperative R&D process, which is essential to the ability of Europe to compete with the U.S. and Japan in the HPCN arena.

The HPCN network is envisioned with the following hierarchy:

- an experimental gigabit testbed between a few of the major facilities;
- a high-speed backbone between all European hubs (starting at 34 Mbits per second and evolving as rapidly as possible to gigabit speeds);
- national and regional networks connecting to the backbone at the hubs;
- local area networks of users connecting to regional networks.

As in the U.S., the HPCN program will be the driving force behind the trans-European information highway. It is estimated that the HPCN program will contribute about 70% of the general infrastructure costs. The national developments will be covered by existing network programs. It is expected that 700 to 800 HPCN facilities and user groups will be connected by the pan-European backbone in the first five years of the program.

*HPCN Support Nodes*

The networks will connect HPCN "support nodes", which will include support resources and expertise, large-scale advanced facilities and software support and distribution. The EC role will be to ensure the integration of these nodes with the entire HPCN program and to provide start up funding to get the centres established with adequate facilities and resources. Nodes will be selected from competing proposals on the basis of their compliance with the overall objectives of the program. It is expected that the initial nodes will be extensions to existing centres for economic reasons, but also because pools of expertise will be more readily available at an established centre. There are plans to establish three levels as follows:

A-Node: A node operating an HPCN system costing $20 million ECUs; 20 support staff; eight type-A nodes are planned in the first two years.

B-Node: A node operating an HPCN system costing $5 million ECUs; five support staff; four type-B nodes are to be established each year for the first five years.

C-Node: A network access node to external HPCN systems; five support staff; four type-C nodes are to be established each year for the first five years.

Note that the support staff referred to above is the group of experts that will provide support services to the researchers and developers who will be the real users of the facilities.

It will be the role of these nodes to ensure that users are able to access advanced HPCN equipment and have the support needed to exploit it effectively. They will provide the ongoing training and education to the user community and they will arrange for the dissemination of information and software.

*iv) Education and Training*

The two objectives of this element are to:

- ensure an adequate supply of skilled staff to meet the requirements of the emerging HPCN infrastructure;
- de-mystify HPCN so that users and managers can understand and exploit it effectively.

The delivery of education and training will not be restricted to universities, but must also include other institutions, support nodes (the centres), private organizations and industrial or commercial sites. The role of the HPCN will be to coordinate the efforts on multiple levels and to ensure adequate access to facilities. It is estimated that one million students per annum will require training and access to facilities at a cost of $110 million ECUs per year.

#### 4.2.2.4 Application Focus of the HPCN

Three broad classes of applications have been defined for the HPCN program.

*i) Simulation and Design in Engineering and Science*

The objectives of this aspect of the program are almost identical to those of the U.S. HPCC program, therefore not much more need be added at this point. In summary, the program will be directed at replacing traditional laboratory experimentation by computational science in areas such as physics, chemistry, biology, medicine and engineering. The following are some typical examples:

- industrial design and testing, leading to faster less costly product development;
- rational design of new materials, e.g. ceramic coatings to improve efficiency of engines and a decrease in pollution;
- fluid-flow and combustion simulations, e.g. engine design, pollution reduction, noise control, hydrocarbon reservoir modelling, ground-water studies;
- design of new chemical and pharmaceutical products;
- design of new medical devices.

*ii) Information Management*

The parallel systems developed and tested in the HPCN program will be applied increasingly to information-management problems involving massive databases that will be searched concurrently by multiple users over the HPCN international broadband networks. The following are examples of applications that will demand massive computing power and intelligent classification schemes:

- searching patents or inventions databases;
- searching legal cases for precedents and jurisprudence;

- searching medical records for similar cases and history of treatment;
- exploiting information in large-scale financial, commercial, engineering and genetic databases.

*iii) Embedded Systems Applications*

This element of the program will strive to find effective ways to embed HPCN technology within other products. In these cases the technology will be used to improve the performance of some higher-order technology and the HPCN component will be totally transparent to the user. A few examples of applications are:

- improved scanning for postal automation;
- image processing for medical, industrial and seismic images;
- virtual reality for training and entertainment;
- management of telecommunication networks;
- compression and decompression of massive databases;
- real-time systems using more intelligent robots.

4.2.2.5 Funding for the HPCN

As indicated above, it is expected that the new financial commitments to the HPCN program will rise to over $1 billion ECUs per year by 1995. With the buildup from lower amounts in the early years, it is anticipated that the new money that will be invested in the first five years will total $3.5 billion ECUs, distributed over the lines of action as follows:

| | |
|---|---|
| Application and support | 30% |
| Research, development and demonstrations | 25% |
| Supporting infrastructure | 35% |
| Education and training | 10% |

#### 4.2.2.6 Benefits of the HPCN

The case is made for strong support for the HPCN program by the advisory committee, as follows:

> HPCN will be crucial for European competitiveness in the context of a rapidly evolving global situation and the realization of the single market. However, the new opportunities will be beneficial mainly to those sectors and companies able to exploit sophisticated technologies such as HPCN to offer higher added value in their products and services. It is important that large scale industrial and scientific areas where Europe is currently strong should take advantage of HPCN in order to maintain and enhance their world-wide competitiveness. In addition, HPCN would be instrumental in gaining a strong position for a vast range of small and medium enterprises which will benefit from an increased use of this technology.

It is interesting that the committee attempted to quantify the economic benefits of the HPCN program by drawing parallels to the U.S. Gartner study which we have described in subsection 4.1.3.3 above. In the Gartner study, there was a prediction made of productivity improvements in key industries that will benefit from HPC. In Europe, the key industries to benefit the most are aerospace, automotive, electronics, chemicals, pharmaceuticals, computers and construction. In these industries, OECD statistics indicate that the total EC R&D annual expenditure is of the order of $40 billion ECUs in 1992. The Gartner study has demonstrated that a small percentage increase in R&D productivity in these industries (1% to 5% in the U.S. examples) will lead to total GNP increases of about 0.5%, which represents about $17 billion ECUs per year for Europe.

Furthermore, the European IT industry alone generates revenues of $175 billion ECUs, representing 5% of the GNP, but it is expected to represent 10% of GNP by the year 2000. As at least 50% of the HPCN funds will be spent with European suppliers, there is an obvious direct and indirect benefit to this strategically important IT industry.

### 4.2.3 Use and Applications of HPC in Private Industry

There is a recognition in Europe that a focusing of innovative HPC solutions on industrial applications will strengthen the competitive position of the adopter industries with the added spinoffs that will improve the entire economic fabric of the EC in the emerging information society.

The industries that have benefited the most from HPC to date are basically the same ones that have been described in the preceding section on the U.S. There are major industries of strategic importance in the sectors of aerospace, avionics, automotive, ship building, chemical, petrochemical, pharmaceutical, consumer electronics and IT, all of which were early adopters of HPC. We will not repeat the specific applications and their benefits, as this has been described adequately in the U.S. section. There is, however, very limited usage of HPC by small and medium-sized businesses, primarily because of a general lack of awareness of capabilities, cost-effectiveness and overall benefits that can be achieved in design/development time and cost. The HPCN program will strive to address this critical problem of awareness.

RCI Europe (an HPC research consortium) sent surveys to 400 known HPC sites in September 1992. The 51 respondents indicated that they collectively had an installed base of 54 vector supercomputers and 12 massively parallel machines at that time. Close to one half of the respondents were private corporations, primarily in the sectors of aerospace, automotive, electronics, petroleum, telecommunications and utilities. The industrial HPC applications that these private companies expect to be most prevalent are fluid dynamics, chemical and pharmaceutical rational design, structural analysis, signal/image processing, database applications and electromagnetics.

An application area that does not appear much in previous studies because it is so new is the application of parallel technology to commercial applications, especially for large information servers on the networks. One company, Meiko, has sold 12 systems to private companies for massive database applications. The firm sees this trend to be so significant that it has repackaged its MPP system as an "enterprise server" and is shifting most of its focus from scientific applications to commercial information-processing applications.

### 4.2.4 Network Situation in Europe

While there are several innovative test-network initiatives now emerging throughout Europe, the development of high-speed networks to support HPC and other broadband activities has been painfully slow, especially relative to the situation in the U.S. (according to an extensive OECD study in 1993).

Users throughout Europe complain about the high cost and limited availability of data-communications services. The Public Telecommunications Operators (PTOs), which holds a monopoly position, have kept prices at 5 to 10 times U.S. levels for equivalent services. They have expanded the Internet in Europe in terms of nodes and connectivity, but the limited bandwidth and high tariffs have been severely restrictive to the HPC community. It was only in February 1993 that the EC announced plans for a 2 megabits per second pan-European backbone, at a time when the U.S. backbone was already 45 megabits per second. The PTOs were slow to accept that there is a significant commercial market for high bit-rate networks, while the HPC community has been reluctant to pay high rates for the current low bandwidths; hence the situation has not progressed. Presumably the HPCN program, deregulation and new demonstrations of the benefits of information highway applications should open the doors for rapid progress.

In some European countries, pilot projects and testbeds are developing and implementing new broadband communications technology including ATM and SONET (called SDH, or synchronous data hierarchy, in Europe) over fibre. For example, the most promising project is SuperJANET in the UK (the UK Internet organization), which plans to connect 50 educational institutions, medical centres and research facilities in an ATM/SDH network, using 140 megabit per second links to each site with a backbone of up to 622 megabits per second. This multimedia broadband network will support applications such as high-performance computing, distance learning, electronic publishing, library document distribution, medical imaging and multimedia information services. In HPC, the planned applications are data visualization and interaction, molecular modelling, oil-reservoir studies, computational fluid dynamics, global atmospheric modelling and heterogeneous distributed supercomputing. One important reason the UK is taking a lead in broadband networking is the fact that deregulation is very advanced in relation to other European countries.

Certain other countries in Europe have announced programs for building ATM testbeds for wide area networks, mostly at speeds of 34 megabits per second (lower than North American standards for the lowest level). Research networks across national boundaries within Europe are coordinated by RARE (Réseaux associés pour la recherche européenne). Of course, the most prominent effort towards a comprehensive European program is embedded in the HPCN program, described above, whereby the EC proposes an investment program of $5 billion ECUs over the next 10 years.

### 4.2.5 Comparison of HPC in Canada and Europe

There are more parallels between Europe and Canada than there are between the U.S. and Canada. The major reason for this similarity is that Europe comprises many countries, each on average being closer in size and economic development to Canada than to the U.S. These countries, until the recent HPCN program of the EC, have operated rather autonomously with respect to the development and implementation of HPC technology. There has been no major military imperative similar to that which gave the U.S. its head start. There has not been the same drive to dominate specific strategic industries as has been witnessed in the U.S., with perhaps some minor exceptions in Germany, France and Great Britain.

The other similarity with Canada is that Europe has never succeeded in developing a globally competitive computer-manufacturing industry, even though a few modest attempts have been made. In high-performance computing, Europe has clearly made a conscious decision to withdraw from any serious attempts to manufacture HPC hardware, while concentrating all efforts on software R&D.

The major difference between Europe and Canada is that the EC now has a full-fledged HPCN program that will pump $1 billion ECUs per year in development of high-performance computing and communications applications. By bringing the EC countries together in a collaborative effort, Europe can compete on the same level as the U.S. and Japan, with each individual country receiving the benefits.

Where does this leave Canada in the international HPCC or HPCN community? Canada has no national program to keep up with even the European nations. Canada can compete, provided that we are selective and strategic in our undertakings.

In summary, the following lessons can be learned by studying the European experience (note that some of these points are repeated from the U.S. list):

- As in the U.S., a long-term commitment to funding and technology renewal is essential for building a national HPC program (in Europe's case, also an EC program).

- A multi-nation coordinated program promotes collaboration by all sectors, synergy between partners, industry partnerships and technology transfer.

- Knowledge and experience is first developed in the universities and government labs, and then transferred to industry through industry partnership programs.

- As individual organizations, especially small and medium-sized businesses, cannot justify their own dedicated HPC facilities, it is important to provide infrastructure support with government assistance for establishment of shared centres; the EC

HPCN program will allocate 35% of its budget to shared facilities and network infrastructure.

- Europe has realized that its network infrastructure lags significantly behind that of the U.S. and that this is a severe barrier to promoting the use of HPC; a major European-wide initiative has been established to rectify this problem.

- Awareness programs, training, education and industry outreach programs are critical to success.

- The technologies developed in an HPC program can be effectively applied to new information highway applications; HPC and information infrastructure initiatives should be tightly coupled.

- It is not feasible to consider competing with the U.S. and Japan in hardware development and supply; the EC countries will concentrate on software development.

- Europe believes that there are enormous opportunities in the application of parallel systems, which can only be exploited by innovative new software environments; consequently they are focusing their attention and their R&D budgets almost exclusively in this direction.

## 4.3 Japan

There is far less published literature available (at least in English) on the HPC situation in Japan than that in the U.S. and Europe. Therefore, the information provided in this section on Japan was compiled primarily from the following sources:

- previous studies by the consultants while working in Japan;

- Internet files provided by David K. Kahaner of the U.S. Office of Naval Research, Asia;
- surveys by Nikkei Computer;
- personal experience of the participating consultants.

### 4.3.1 Overview

Home to approximately one quarter of the world's high-performance computing systems, Japan gets high marks for both effort and achievement in all aspects of high-performance computing. The country funds research in information technology more heavily than all other fields except the life sciences and the environment, and has the only serious supercomputer manufacturing industry outside the United States.

Comparisons of the American and Japanese approaches to and accomplishments in supercomputing are inevitable, as are the periodic claims, by manufacturers of either country, of having the world's most powerful machine. Japanese supercomputers are built by the same vertically integrated giant companies that produce all kinds of technological products, from cellular telephones to semiconductors. Therefore, their approach is hardly surprising: extremely fast individual supercomputer processing units built around very high-speed emitter-coupled-logic (ECL) semiconductors of their own fabrication, which are second to none.

As in the United States, the information-technology research establishment in Japan has industrial, governmental, private, and academic components. With total R&D budgets of approximately U.S. $10 billion per annum, industry investment accounts for nearly 95%, with the balance coming from private research institutes, universities, and government research institutes.

There are two significant measures of success that are often mentioned in relation to Japan's adoption of HPC compared to that of other countries, especially the U.S. First, there are many surveys that indicate that Japan has the world's highest number of installations of machines in proportion to the population or the GNP. However, some caution is required when making this claim, as Japan tends to manufacture a larger percentage of lower-cost lower-performance supercomputers, but the surveys do not compare the relative installed aggregate performance levels. Also, the surveys used to provide a somewhat meaningful comparison of installed traditional vector supercomputers, but with the rapid rise of installations of parallel supercomputers, clustered workstations and multiprocessor servers in the U.S., and the corresponding slow adoption in Japan, the balance may have shifted in favour of the U.S. Today, with such a diversity of HPC systems, the surveys have not reflected the real installed HPC performance capacity in comparing one country to another.

Secondly, Japan appears to have a much higher percentage of HPC systems in use by industry than other countries. In Japan, industrial installations account for about 65% of the total HPC systems. By comparison, the U.S. is currently at close to 50% usage, but adoption by U.S. industry is relatively recent, as in the 1980s, the lion's share of HPC machines were installed in U.S. national labs. Japan has used HPC systems extremely effectively for R&D support in building globally competitive strategic industries.

### 4.3.2 Government HPC Programs

#### 4.3.2.1 Earlier Government Programs

At the end of the 1970s, it became apparent that new computer architectures and new devices would be necessary for future needs in information processing. MITI went the usual way in bringing together experts from universities, government research laboratories and industry to formulate a project proposal. The outcome was quite unusual, as MITI decided to run two large projects in parallel, the High Speed Computing System for Scientific and Technological Uses Project, dubbed the Superspeed Project, (1981-1989) and the Fifth Generation Computer System Project (1982-1991). Where the FGCS Project aimed at a risky, new computing paradigm (not directly related to HPC), the Superspeed Project can be seen more as an extension of the present HPC systems. It aimed at the development of a high-speed computing system for scientific and technical applications. The target system was supposed to operate at a rate that was 100 to 1,000 times faster than the speed of conventional computers at the time. Two major R&D projects were conducted: one on high-speed novel devices and one on computer architecture, algorithms and languages for parallel computing.

The project was concluded in 1990 by demonstrating the prototype system to the evaluation team. The prototype of the high-speed parallel system using four processors ran at the target levels. This was not a prototype of a machine that could be directly commercialized. Gallium arsenide devices were used, though not as extensively as envisioned; other novel devices, such as Josephson junctions, were not used at all. Less tangibly, the project focused the private sector on supercomputers at a critical time, earlier and more intensely than would have been done individually. Of course, cooperation also meant that work was done faster and more economically. Individually, the Japanese companies were also investing heavily; some estimates were as high as three to four times the government figure, $300 million to $500 million by each of the three major participants.

The second architectural subproject focused on developing a machine with 128 processors, a precursor to a massively parallel machine with 1,024 processors. The research group successfully completed the 128-processor machine in 1989.

The third subproject was a satellite-image data processing system.

The Superspeed Project was considered by the Japanese companies as helpful, but the results were not directly incorporated into individual products. Also, some differences can be observed between the attitude of the three giant computer conglomerates (Fujitsu, Hitachi and NEC), which could do without the governmental subsidies, and the minor computer players (Mitsubishi, Oki, Toshiba), which would have had more difficulty embarking in these new directions without the extra funds.

#### 4.3.2.2 Real-World Computing Program

A cornerstone of future research in parallel computing is the ambitious new Real-World Computing (RWC) program, which has aroused tremendous interest in the West. This is to be a 10-year Ministry of International Trade and Industry (MITI) program, which started in 1993, with a budget of about 60 billion yen (almost US $600 million at today's exchange rate), focusing on R&D for what is called flexible or intuitive information processing—the way human beings absorb information and make decisions.

The program organizers envision the need for a substantial computational base comprising both massively parallel computing systems and neural systems. These might be physically distinct, although it is hoped that they will be integrated into useful platforms. Research into optical computing is planned as part of the technology base for both systems, either for interconnection or if possible for actual processing. Thus, research will be supported on all aspects of these systems, but the computational base is seen mostly as a platform upon which more theoretical and functional work will be done to support the higher-level research thrusts of the program.

Work on the massively parallel system will involve all aspects. During the first of the project's two phases, various competing systems will be developed with as many as 10,000 processors. During the second phase, the best ideas would be retained and eventually a million-processor system built. The EM-5 already being implemented at the Electrotechnical Laboratory, affiliated with the MITI agency for Industrial Science and Technology, is planned to have 64,000 processors, hopefully achieving teraflop speed.

Research in neural systems will focus on software and hardware, in two phases, as above. The final goal is to develop a one-million-neuron network composed of 1,000 sub-networks of 1,000 neurons each, with a total processing speed of 10 teracups (10 trillion connection updates per second).

A novel feature of the real-world computing project is that it is to be international, interdisciplinary and open. A research association, called RWC Partnership, has been set up by a half-dozen Japanese computer manufacturers, but foreigners may participate. To aid this effort, Japanese intellectual property rights laws have been altered, and project documents are to be issued in English as much as possible.

With such projects as RWC, Japan will strengthen and diversify its position in parallel processing. At the same time, semi-commercial massively parallel systems will become more widely available to Japanese researchers, accelerating progress and paving the way for the introduction of commercial systems. It can safely be assumed that Japan, with its traditional inclination to setting performance records, will stay in the race for teraflop machines. Quite possibly, it will dazzle the world once again with its prowess in a technology critical to economic progress in the coming years and beyond.

### 4.3.3 High-Performance Networks in Japan

HPC networking in Japan has improved of late, but there are more high-performance networks in the United States than in Japan. U.S. network interconnectivity is also much better, although several more or less independent Japanese networks are supported by different government ministries.

Counterparts to the very high-performance networking projects in progress or planned in the United States have been slow to develop in Japan. However, the country has excellent—and in some cases, unique—technology, including a large infrastructure in integrated-services digital networks (ISDNs), and any obstacles to networking there seem to be more social, organizational or cultural than technological.

As for academic supercomputer networks, they are less ubiquitous than in the United States. The prestigious Japanese universities have excellent services, but many lesser universities have none at all.

In April of 1994, an ATM-based gigabit testbed experiment was announced between the Communications Research Laboratory of the Ministry of Posts and Communications and NTT, the major telephone carrier. This project, with an investment of approximately U.S. $500 million, has the objective of establishing the infrastructure to transfer multimedia information to step up the economy over the balance of the decade and to stimulate the entire Japanese industry. Once the infrastructure is in place, the focus of the program will be to encourage private laboratories to develop applications that can take advantage of very high-speed communications. This program has many similarities to equivalent initiatives in the U.S. HPCC program and the European HPCN program.

### 4.3.4 Japanese Information Technology Industry

#### 4.3.4.1 Big Three Vendors

Japan's "big three" in supercomputers are NEC Corp., Fujitsu Ltd., and Hitachi Ltd., all headquartered in Tokyo. Japan to date has concentrated most of its efforts on designing and manufacturing traditional vector supercomputers. Consequently, the users, which tend to favour Japanese products, have focused on the implementation of these vector machines. About half of the 125 vector supercomputers in the country are Fujitsu models; Hitachi, NEC, and U.S.-based Cray Research Inc. share the balance.

Japanese vector supercomputers clearly have very fast central processing units (CPUs), much faster than their U.S. counterparts. It is only very recently, however, that the Japanese have introduced multi-processor versions of these machines, i.e. parallel-vector machines. Users claim that the U.S. is far more advanced in the use of multi-processors.

Of course, the trend today is toward parallelism. While the U.S. is concentrating almost all its resources on achieving orders-of-magnitude performance improvements by incorporating more parallelism, the Japanese have been slow to move away from their practice of building faster single processors. In practice, high-performance is much harder to achieve on multiple processors because of the lack of software. Therefore, the Japanese seem to be gambling on introducing moderate parallelism while continuing to build exceedingly fast processing units for each of the individual processors. They continue to build fast custom processors, sometimes using new semi-conductor materials such as gallium arsenide. Meanwhile the U.S. is moving rapidly to massive parallelism, employing much larger numbers of inexpensive off-the-shelf microprocessor chips, usually the same ones that are mass-produced for workstations.

#### 4.3.4.2 Towards Parallel Computing (MPP) in Japan

Research into parallel computing started in Japan about 15 years ago at universities and national laboratories. As in the United States, early interest in the technology was tied up with physics, and this branch of parallel computing spawns new ideas and architectures to the present day. Special-purpose machines have been built for the study of quantum chromodynamics (the world's most computationally intensive application), gravitation, and thermonuclear plasmas, and other subjects.

Many other Japanese parallel computers serve such special purposes for neural network simulation, digital signal processing, checking the logic of planned semiconductor circuits, fingerprint matching, deoxyribonucleic acid (DNA) sequence search processing, and image or graphics processing. At least 50 such special-purpose systems have been developed, and at least 50 varied research projects on parallel computing are under way, mainly at universities.

These parallel computer projects are mostly stepping-stones for further work and will not lead to commercial products. In fact, Japanese manufacturers have only very recently begun to offer the first commercial massively parallel computers, although there had been earlier introductions of moderately parallel machines using small numbers of very high-performance, relatively expensive, individual processors, i.e. very different from the U.S. MPP systems.

There are several obstacles to faster progress in the development of commercial massively parallel architectures in Japan. One is the general feeling that conventional (modestly parallel) systems have advantages and greater potential. Others include a limited market for conventional supercomputers, recent evidence of downsizing computer systems, and predictions that stand alone workstations will flourish most in the computer market.

Another, more profound, obstacle is the lack of small venture start-ups in the country's computer industry. The large vendors are loath to develop the architectures out of fear of losing the markets for their existing products. Perhaps the developmental role will be played by other manufacturers who have until now been preoccupied with home electronics; some of these are eager to move into the high-performance computing business and are slowly succeeding.

Nonetheless, the failure of Japanese parallel computers to reach the West must not be construed as evidence that the projects are not active or well established. For the most part, a lack of resources has dictated that most of these systems satisfy educational needs, rather then blossom into commercial products. Given this limited hardware availability, plus the well-known Japanese interest in building hardware systems, it is only to be expected that the kind of research into parallel software applications so common in the West should be much less common in Japan.

Despite these and many other development efforts, there is only a handful of operating parallel computers in the hands of industrial software developers, even fewer at universities, and very few available to users who want to solve real problems. This situation seems poised for change, as manufacturers' systems are now making their way into the hands of Japanese researchers.

Research in software for parallel computing is far behind that in the West, but is growing and will accelerate further as more researchers gain access to systems. There is also significant work in distributed computing, which can be performed on networks or clusters of workstations. However, large companies are still engaged with vector supercomputers. Lacking parallel applications, they are cautious about moving commercially into even more risky massively parallel systems.

#### 4.3.4.3 HPC Software in Japan

Software for supercomputers includes compilers, libraries, operating systems, networking support and software tools. The Japanese have been slow in comparison with the U.S. to adopt UNIX, the de-facto international standard for HPC operating systems. In recent years, all the manufacturers have introduced UNIX as an option, but many users are still clinging to the legacy proprietary operating systems. The industry is only now coming to grips with the need to assess software costs, and moving to Unix is seen as one way to reduce costs for end-user and vendor alike.

In the past, applications developed in the West have been installed very slowly, which has impeded foreign sales of Japanese supercomputers. Using Unix will ameliorate this situation, to the extent that it guarantees greater software portability and shorter development time. Efficient program execution, of course, is another matter entirely. As yet, no shortcut to maximum performance seems to exist besides incorporating knowledge of the hardware into the algorithms and software.

In the early days, Japan's development of supercomputer software was limited to producing Japanese-language interfaces for Western software products. Though this activity is still important, a number of first-rate packages designed and produced in Japan are appearing. Time will tell whether they will have any success in packaging and marketing this software to international markets.

There is also a trend to enlist Western scientists with appropriate experience. NEC and Fujitsu have established facilities, as well as collaborative research with groups in the United States, Australia, and Europe, for this purpose. Thus, Japanese vendors are becoming more effective at accessing expertise beyond their shores.

For those users who need to write programs, standard compilers run on all Japanese supercomputers, and the vendors scrupulously meet all announced standards while offering various enhancements and extensions.

For the Japanese market, there is relatively little demand for standardized software; vendors and users still develop libraries and user interfaces for their own platforms and applications. Japanese computer users tend to write their own application software, and people who have studied it from the inside claim it is often quite good. In summary, Japanese software for supercomputers, especially for multiprocessors, is only now beginning to emerge.

### 4.3.5 Uses and Applications of HPC in Japan

#### 4.3.5.1 Academic Use of HPC

Most university scientists in Japan can obtain supercomputer time, but rarely on top-end machines, which are mainly to be found at industrial laboratories or in the prestigious national universities. Even so, this is an improvement on the state of affairs a few years ago, though it is still below what is available to U.S. academics. There is nothing comparable to the four HPCC national supercomputer centres in the United States partially sponsored by the National Science Foundation, even though supercomputer centres exist at leading Japanese universities and at several government and some corporate laboratories.

#### 4.3.5.2 Industrial Use of HPC

As explained above, what is very impressive about HPC in Japan, is the very high rate of acceptance by industry. The latest statistics available for the industrial use of supercomputers in Japan is a Nikkei Computer survey in 1992. With the recent recession in Japan, there has probably not been a major shift in the allocation of investments in HPC technology since that time.

Japan has been the world's largest industrial adopter of supercomputers in proportion to the size of the country. The table below summarizes the industries that have their own dedicated supercomputers installed:

| **Industry Category** | **Number of Supercomputers** | **Percentage of Total Installations** |
|---|---|---|
| Electric / machinery / telecommunications | 33 | 13.2 |
| In-house computer vendors | 30 | 12.0 |
| Automotive | 29 | 12.0 |
| Finance / services | 29 | 11.6 |
| Chemical / pharmaceutical | 12 | 4.8 |
| Construction | 10 | 4.0 |
| Steel | 9 | 3.6 |
| Other manufacturing | 5 | 2.0 |
| Unknown | 17 | 6.8 |
| Non-industrial (government, academia) | 75 | 30.0 |
| Total | 249 | 100.0 |

A few interesting observations are of interest:

- In the three-year period preceding this survey, the total number of installations doubled, with the industry portion gaining about a 6% additional share of the total.

- The automotive sector increased its share in this same period from 8.5% to 12% as the large manufactures installed multiple supercomputers, e.g. Toyota already had four supercomputers in 1992. In automotive companies, supercomputers are completely integrated in the design/development cycle and have become indispensable for the major manufacturers to introduce new designs quickly enough to react to shifting market demands. The major applications are crash analysis, which alleviates the need to build multiple prototypes to be physically crashed against a wall, and combustion analysis, for more fuel-efficient engines. Toyota even installed a production parallel system for analyzing light-meter readings and selecting optimum colour schemes for the lighting conditions in different countries.

- Many of the supercomputers in the Electric/machinery category are suppliers to the automotive industry (an application area that may have relevance in Canada).

- There is growing acceptance in the pharmaceutical industry, as companies realize that it is no longer possible to be competitive if new medicine must be developed exclusively in the laboratories. To quote from Kissui Pharmaceutical, "It is typical that it takes 10,000 laboratory trials to discover one new drug". We developed new products only by experimentation with living bodies in the past, but such methods have reached their limits. Considering cost and time, it is necessary to focus on targets by numerical experimentation [with supercomputer molecular models] and to then carry out the chemical experimentation in the labs. Furthermore, the computational approach has become more feasible with the introduction of lower-priced supercomputers and the availability of living body databases and application software.

- In the construction industry, it is remarkable that supercomputers are used not only as a design tool for civil engineering projects, but also as a strategic sales tool. Construction companies improve their chances of winning bids by demonstrating much higher-precision earth quake-proof designs through the results of computer modelling.

- It is interesting that it is not only the very large manufactures, e.g. automotive, that use supercomputer modelling to design new products. Many smaller industries, including manufacturers of consumer products, are using supercomputers to improve their competitive position. HPC modelling shortens design time, enables earlier introductions of new products and allows for more precise higher-quality designs. Supercomputers are being used by the following kinds of manufacturing companies in Japan: consumer electronics, rubber and tires, sports equipment, sewing machines and cameras.

- Probably the one application area that has been penetrated more in Japan than in any other country is finance and services (29 supercomputers). While it has been difficult to obtain information on the actual applications, Nikko Syoken, a stock-brokerage firm is one example. Nikko has stored five years of market-fluctuation data for each trading stock (a single stock can represent 70,000 records of information). The firm has developed a neural network learning algorithm on a vector supercomputer to predict future stock movements. Since this is a computationally intensive process that involves accessing masses of data and requires rapid response, it necessitates the use of a supercomputer.

### 4.3.6 Collaborative Situation in Japan

Some years ago, the key phrase *san-ganku-kan* began to appear in every document on Japanese research policy. It is a short form for research cooperation between industry (*san*), universities (*ganku*) and governmental research institutes (*kan*). Discussions showed that in Japan this cooperation was not well established (nor is it yet); the greatest problem exists within the government itself. In principle, the Ministry of Education is concerned with basic research, the Science and Technology Agency with "big science", e.g. nuclear energy, air and space development, and MITI with applied research, but there is naturally an overlap between these areas. In order to minimize organizational problems, there is no cross-funding between MITI and the Ministry of Education. Within the Real-World Computing project, which will be closer to basic research than MITI projects in the past, a softening of these strict regulations is expected.

At the same time, Japan is changing its laws and regulations to make participation by foreign researchers in national projects easier.

### 4.3.7 Comparison of HPC in Canada and Japan

Of all three economic units (U.S., Europe and Japan), a comparison of Canada with Japan probably yields the fewest parallels or similarities, and Japan is probably the most difficult model to emulate. These difficulties are the result of major cultural differences, along with very different business practices and work ethics.

The first major difference is that Japan decided several years back that information technology was to be a major strategic target industry for development. Being already masters of producing the world's fastest chips and having been quite successful at producing large powerful reliable mainframe systems, it was quite natural for the Japanese to establish goals to build the world's most powerful supercomputers. Today they are in this race and their only competition is the U.S. Countries that manufacture supercomputers tend to establish programs for their application in domestic markets much faster than countries that only import the technology. Consequently, Japan and the U.S. have been the most successful at developing HPC applications, while Canada has had minimal isolated successes.

Secondly, since the Second World War, Japan has been building high-tech industries with enormous investments in R&D (an R&D-to-GDP ratio of 2.87%, compared to 1.5% for Canada). Computers, and specifically HPC, have long been recognized as a critical tool to support new product design and development. Japan applied HPC to its industrial R&D far more aggressively and far sooner then the U.S. (currently 65% of supercomputers are used by industry in Japan, compared to 50% in the U.S.). Lastly, Japan has found a way for HPC techniques to be applied to R&D in small to medium-size industries. Consequently, the Japanese situation is most unlike the situation in Canada, even with respect to applications of the technology.

While Canada has no aspirations to design or develop HPC hardware, perhaps we have much to learn about the application of the technology to industry from the Japanese.

In summary, the following lessons can be learned by studying the Japanese experience:

- Being a world-competitive supplier of computer hardware systems requires billions of dollars of R&D investment by large multi-national conglomerates.

- Exporting software requires rigorous adherence to international open standards; a second requirement, which has proven a major difficulty to Japan, is that the software functionality, ease of use and documentation must match the standards set by the U.S.; Canada can capitalize on its close proximity to the U.S. to position itself as an important HPC software developer.

- Two technology trends—downsizing to workstations and growing acceptance of parallel systems—have been understood much more slowly in Japan than in North America; this is having an impact an Japanese exports of HPC hardware and software.

- Japan has made very effective use of HPC to support the intensive R&D programs that fuel the growth of their strategic industries, especially in high-technology segments.

- The most important lesson that Canada could learn from Japan is an understanding of Japanese success in transferring HPC technology effectively to small and medium-size businesses in industries that do not use HPC in this country, e.g. chemical, pharmaceutical, construction, consumer-product manufacturing and others.

- As in the U.S. and Europe, government participation in HPC programs has been an essential factor in the successful development and application of this technology.

- Japan uses collaboration on a grand scale, even to the extent of involving major competitors in joint HPC development programs.

## CONCLUSIONS

### 1. Summary of the Present HPC Situation

There is a consensus that Canadian research and industry continues to suffer from the absence of a national high-performance computing program. Countries that have actively promoted the diffusion of HPC have benefited from large pay backs in terms of both scientific research breakthroughs and improved economic competitiveness.

One quantifiable measure of a country's commitment to HPC is the relative per capita investment in HPC by government, since no country has embraced the technology without infrastructure support from government. Federal and provincial government support in Canada averages less than 25 cents *per capita* per year, while most of our economic neighbours in the G-7 are exceeding that amount by 10 to 20 times. One individual, who has been an HPC user since the 1970s, questioned the logic of why, when a program like the Canadian Space Agency is initiated, there could not have been even a 1% financial allocation for HPC.

However, what is particularly disconcerting is the lack of awareness of the benefits and industrial opportunities that result from the utilization of HPC. Canadians need to be exposed to success stories of companies that have improved their global competitiveness by shortening design cycles, getting products to market sooner, reducing design and product-manufacturing costs and improving quality, all through HPC modelling and simulations. With new emerging database and multimedia applications supported by HPC technology, Canadians need to become more aware of how HPC technology can provide the underpinnings of improved health-care services, the support for lifelong learning and better dissemination of government information.

### 2. Canada Compared with Other Industrialized Nations

The study has looked at what has driven the high-performance computing and networking (or communications) programs in Europe, Japan and the U.S., in an attempt to see why Canada has not developed its infrastructure or knowledge base in HPC at an equivalent pace. What has been concluded is that the U.S. superpower position necessitated the advancement of supercomputing in that country. Subsequently, the U.S. research community's concern with the lack of available computing resources stimulated the development of the National Science Foundation HPC centres and the network infrastructure base. Lastly, quality of life and international competitiveness has crested the HPC wave with support at the current level, which now exceeds one billion dollars per year of federal funds. Though confirming figures are not readily available, this amount is likely doubled with the inclusion of State supported programs.

In Europe, some elements of "keeping up with the U.S." have occurred, but science has led the demand for supercomputing. Qualify-of-life programs, especially weather and related atmospheric research, have been amalgamated across Europe for several years. The hardware design and manufacturing of HPC devices has fallen into the hands of U.S. or Japanese interests. Also, individual member countries of the EC find it difficult to sustain their individual HPC programs. Consequently, the EC members have united to establish the HPCN program of networked experts, facilities and collaborative projects, with emphasis on HPC applications and software. Like the U.S., an expenditure of one billion dollars per year is being recommended by the EC. However, with the inclusion of expenditures by the member states of the EC, this could easily rise beyond the $2 billion level.

Japan's traditional challenge with technological competitiveness and its dependence on a strong world trade position has led to the nation's emphasis on HPC. As the world's largest R&D investor per capita, Japan has also posted the highest HPC expenditure per capita of any country. The penetration of HPC in industry is the highest in the world. Computer developers and manufacturers have not only seen sales opportunities but have realized the "trickle-down" benefits of using the advanced technological principles of HPC in other newer products.

With a population comparable to that of many member states of the EC or American states, plus a territory that exceeds either of these, Canada cannot afford to maintain an equal

HPC position. However, like many of our national programs, such as communications and transportation, we can achieve the same relative benefits by proceeding more intelligently.

## 3. HPC Support of R&D

Certainly the major factor that explains the low level of acceptance of HPC in Canada is the relatively low R&D spending. Furthermore, most of Canada's R&D does not require HPC solutions, because the large scientific projects that drive HPC usage internationally are conducted elsewhere, often by parent companies of Canadian corporations.

Total spending on R&D in Canada, currently at about $8 billion per year, in no way approaches that of Japan, the U.S. or several of the member states of EC. Even when expressed as a percentage of the GDP, Canada spends only 1.5% on R&D, compared to 2.75% in the U.S., ranking Canada fourteenth among the 24 nations of the OECD and considerably lower than that of most G-7 countries.

Furthermore, major Canadian industry is often partnered with a parent firm in the U.S., Europe or Japan. The automotive and pharmaceutical industry present the best examples. Whatever they might require in HPC is served more cost-effectively by using the corporate HPC facility outside of Canada. Furthermore, the relative cost saving of moving data the short distance to the U.S. border versus moving it across Canada encourages north/south usage and prohibits active consideration of use of an equivalent Canadian capability whenever it is too remote.

Even though Canada does not have the R&D base of its major trading partners. Canada's industries must still find ways to compete in the global economy. If we examine the top 100 R&D industrial performers in Canada, we discover that 53 of these, representing $2 billion in R&D expenditures, are involved in sectors where HPC solutions have provided a significant competitive boost in other countries, i.e. aerospace, automotive, biotechnology, chemicals, pharmaceutical, electrical products, energy, metals, nuclear energy and petroleum.

The real question is, How does Canadian industry, which still invests approximately $5 billion per year in R&D, become more competitive by selectively applying HPC techniques where it will be most effective?

## 4. A National HPC Program for Canada

A culture of high-performance computing must be instilled in Canadian thinking. This awareness cannot be achieved simply by formal education over a short period of time, but instead must result from experience in the application of HPC in imaginative ways, empowered by availability of the devices, the software, the tools and the access mechanisms. Also, an essential dialogue must be encouraged between experienced and inexperienced or potential users. This meshwork of needs to enhance the Canadian knowledge base in HPC, combined with the geographic spread over which it must be dispersed, is clearly a unique Canadian challenge.

*HPCNet, metacentre and national virtual machine room* are three terms that have been suggested to represent the combination and collaboration of all aspects of Canadian high-performance computing, from education and support services through provision of facilities of all technological types, accessed via standard, affordable telecommunications media. It does not represent a full funding commitment by any single group, but instead represents a sharing of associated resources under the structure of a long-term (minimum ten-year) Canadian HPC plan. All players will have a role and a say in the direction HPC will take with existing resources providing an immediate base from which Canada's future in HPC is leveraged.

The presence of high-performance computing facilities within Canada cannot be perceived as an end in itself and certainly cannot be a primary goal of the network of facilities and resources discussed above. Instead, where facilities in Canada are economically reasonable and complement the Canadian infrastructure base, they must be encouraged. It is much more important to a Canadian HPC position to ensure the availability of a variety of modern technologies, accessible and supported from the user's desktop.

The U.S. has linked all of its major HPCC centres in a broadband network and is effectively implementing a shared human-resources network to provide education, training, support services and industry outreach. In fact the U.S. has recently announced a Metacentre Regional Affiliates Program to broaden the reach of the concept to geographical areas that heretofore have been inadequately served. A distinctive Canadian metacentre would have to be conceived on a much more modest scale of course. Furthermore, linkage of a Canadian network of centres to the U.S. metacentre would even further enhance the opportunities for Canadian users. For shared research programs, the cost and availability of U.S. centres is likely attractive. However, one potential pitfall is that U.S. centre access could become overly expensive and less available for Canadian *commercial* users, particularly where the end product (e.g. a software package) of a research venture might be perceived by the U.S. as a competitive threat.

As Environment Canada is by far the largest user of HPC in all of Canada, it is reasonable to expect that this department may wish to play a role in the development of a Canadian metacentre strategy. In conversations with upper management at Environment Canada, especially Atmospheric Environment Services, there is a willingness to explore possibilities.

The CITI applied research laboratory of Industry Canada focuses on the development of collaborative work environments. As the HPC environment is characterized by a wide variety of expensive computing resources and technical specialists, there is an interesting opportunity to apply the CITI methodologies and systems to establishing a Canadian version of the U.S. metacentre (a pan-Canadian virtual machine room with a network of support services). CRIM in Montreal has also expressed interest in becoming a partner in such a program.

Lastly, it is important that the federal government agencies that are committed to the advancement of science, e.g. NSERC, NRC and MRC, be encouraged to participate in a national program.

## 5. Canadian HPC Software Program

What is expected to be uniquely Canadian in a national HPC program is Canadian-developed software tools and a support methodology.

With a few exceptions, according to the findings of this study, Canadian excellence in software development has not yet discovered HPC, as only a few developers concentrating on HPC applications were identified. It is fully expected that a general increase in HPC awareness and education, along with the experience gained from renewed access to HPC facilities and services, will significantly expand the number and variety of Canadian individuals and organizations developing HPC software. What remains is to find a methodology to stimulate HPC software-package development, to ensure the availability of resources to support such a methodology, and to aid in the development of national and international markets.

Industry Canada is interested in promoting the development of the strategically important software industry within the information technology sector. In HPC, software is presenting greater opportunities than hardware, especially with new parallel systems for which designing software is more complex and time-consuming than developing the hardware.

The Software Human Resources Council, Inc. (SHRC) was set up with a contract from Human Resources Development Canada to provide top-quality training, improve the definition and image of software workers, maximize the supply and quality of new entrants into the software workforce, track changes in labour and skill demand, integrate the education and training continuum, and promote certification and national standards.

SHRC has launched a three-year program that aims to retrain 15,000 software workers across the country, bringing their outdated skills into the 1990s. The $12 million program with funding split 50/50 by private industry and the federal government will fuel a number of projects. The goal is to establish about 150 partnerships between business and education. HPC should be identified as a specific goal within this program to satisfy the software development objectives defined in Section 3.9.

## 6. HPC and the Canadian Information Superhighway

The implementation of "information servers" on Canada's information superhighway is not a new concept. Implementing well-supported HPC information servers, using advanced HPC parallel technology, could offer Canadians unique application services (content) that were previously unavailable and unaffordable. These could also target requirements of international users and be available through shared "highway" links, thus creating new opportunities for Canadian entrepreneurial ingenuity and new Canadian jobs.

One example of such a centre currently operating in the U.S. is "Network Entrez", an Internet-based client-server system that allows for integrated searching of DNA and protein sequences and the medical literature associated with these sequences.

Canadian examples could include a national climate database provided by Environment Canada; extension of the British Columbia protein and DNA sequencing database service to a national level; development of a national database for satellite observations including data to be downloaded from Radarsat, to be launched in 1995; a geographical information system (GIS) database for resource companies, utilities, municipalities; and a test-bed for new multimedia information services.

Through CANARIE, which already receives substantial support from Industry Canada, there is an opportunity to develop synergy between HPC and HPN (high-performance networking), similar to what can be observed in the U.S. and Europe.

When the consultants visited the Communications Research Centre (CRC) and Telesat in Ottawa, there was encouragement to identify HPC applications that could be pilot projects for the broadband (ATM) communication facilities being introduced by these two organizations.

CRC has a research and development facility at Shirley's Bay in Ottawa, designed to demonstrate and test innovative information highway applications over high-speed ATM fibre-optic networks, with interconnection capability and extension via Telesat-operated satellites. The CRC will collaborate with research institutes and private industry on programs of potential public benefit, in fields of interest to the HPC community, including:

- Science and Engineering—the high-speed transfer of data and high-resolution imaging, coupled with full-motion video consultation between professionals in different locations.

- Government Services—the high-speed transfer of information using ATM technology for government applications such as movement of large weather and climate data files, GIS information (e.g. maps) and massive data from Earth-observation satellites; these are applications where HPC parallel information servers could be most efficient.

Such possibilities of collaboration with the HPC community opens the doors for several interesting opportunities to test and demonstrate new HPC applications requiring the high-speed networking of computer facilities and technical specialists.

HPC users constitute the one user group that needs the highest possible bandwidth. The risk of not maintaining high-speed networks at international levels is that Canada will lose its network service business to U.S. bypasses and, furthermore, the users will then make more use of U.S.-based HPC facilities in the process, leaving Canada with a crippled HPC infrastructure. Canadian tariffs are, up to an order of magnitude, more expensive than equivalent long-haul high-speed circuits in the U.S. We can only look forward to deregulation and the positive impact it has had on tariffs in other countries.

## 7. HPC Support and Associated Services

Canadian suppliers of HPC services are few in number and thus cannot provide convenient, universal, desktop support to a user community. This, too, is the case for remote use of U.S. Centres. Thus, there is a continuation of a tendency to use only those HPC tools with which the user is familiar and which are readily accessible, i.e. the local workstation. A goal of a national HPC program must be to enhance the knowledge and use of HPC by ensuring desktop access to support services.

We also have noted the emergence of tools which could support a unique Canadian HPC program. Consider, as a sample scenario, an individual HPC user or a professor in a HPC class wanting to determine the best solution to a large-scale problem, but being uncertain how to decide which HPC facility or software support tool might best serve the need. A video link could be established with an expert who may be under the employ of any one of the participating facilities or perhaps an associated consultant group. An interactive dialogue could be supported by the individual's displays of problem definitions, associated software programs, set-up routines, and the visual or numerical results from an HPC system execution. The impact on the new or inexperienced HPC user would be most encouraging and perhaps dissolve some of the HPC mystique the currently exists.

Then there is the user who has decided to try an execution, obtained a little help in moving the application to a specific, remote facility and then finds an error in the results. In the current Canadian HPC environment, it is likely that the entire program, data file and results must be sent to the remote centre's expert, who recreates the problem almost in isolation, perhaps guesses at a solution and then returns the results to the frustrated user several days later. A Canadian networked HPC facility could support an interactive visual dialogue with an immediate, joint resubmission of the problem program and shared review of the result. This participation will not only increase familiarization of the user but will ease the frustration of use of distant facilities.

Such capability for user support would have to be extended to the use of U.S. facilities if this becomes a part of the Canadian strategy.

The availability of documented knowledge of HPC must above considered. Individuals can certainly commence the development and maintenance of their own libraries. They can access the reams of papers available in the U.S. and Europe. Again, using the multimedia tools described above, a Canadian on-line library could be established, containing information on users, applications, examples and tools.

## 8. Role of Government

In these times of government downsizing to bring the deficit under control, it is unrealistic to expect significant new investments in a national HPC program, even though excellent arguments could be made, by pointing to examples in other countries, that this will have a major positive impact on the economy. The approach at this time should be to better coordinate the expenditures within existing programs and to ensure that HPC projects are presented in a positive light.

The acquisition, installation and use of HPC facilities in Canada is limited by cost factors, including the costs of purchase, operation and replacement. Less obvious is the cost of conversion of applications and software packages to such new devices and the learning time necessary to migrate to new devices.

One solution to the financial dilemma being expressed could be the simplified access and use of shared facilities, such as that being suggested under this national HPC program. Several government departments at both the federal and provincial levels should collaborate on a national HPC program, in a similar fashion to the way in which agencies are cooperating in the U.S. HPCC program.

In general, provincial governments have been more proactive in supporting HPC than the federal government over the past decade. In Ontario, the government provided approximately $18 million to the previous OCLSC centre and has now earmarked $29.4 million to new facilities presently under final review. The Alberta government provided $5 million over five years to the HPC Centre in Alberta. The Quebec government has provided $12 million over six years to CERCA, a consortium of industry and university researchers that investigate new algorithms primarily in computational fluid dynamics. Also the Quebec government has been the largest financial contributor to CRIM, which has a group of 10 to 15 persons developing software for parallel computing. Other smaller provincial government grants have been provided for specific HPC projects that are considered strategic to the stimulation of economies.

While the provincial funding of HPC has significantly exceeded federal funding to date, the total amount of support is inadequate and too sporadic to establish even the basic infrastructure upon which a national HPC program can be built. One of the principal recommendations received during the interview process of this study is that these provincial programs be integrated with a national HPC program and that efforts be made to encourage inter-provincial collaboration for the sharing of scarce computational resources and personnel.

Thus, government encouragement and leverage of the situation is necessary, at least for the first years of a national HPC program. Once the understanding and familiarization process is well underway, new methods of continuing the program will have to be developed. As well, in recognition that new jobs for Canadians are developed at a much faster pace within the small-business sector than with traditional large businesses, these new enterprises need added support, especially with the acquisition and use of HPC devices and software.

In closing, it is recommend that the reader request form CITI the companion document *A Study on High Performance Computing in Canada: Executive Summary*. This document reviews the issues that were deliberated by the HPC Advisory Committee and summarizes the resulting recommendations.

# APPENDICES

## APPENDIX A

## Reference List of HPC Reports and Previous Studies

The following is a list of high-performance computing and communications reference material that was used as one of the information sources for this study.

### Canadian Documentation

1. "Feasibility Study, High Performance Computing Facilities in Canada"; initiated by the Ontario Centre for Large Scale Computation and supported by NSERC, Nov. 1992.

2. "National Initiative for Scientific Computing (NISC), A Proposal for Implementation"; NSERC June 1991.

3. "Report of the ad hoc Committee on Research Computation (CORC)"; NSERC, Sept. 1991.

4. "Supercomputing Research: Issues and Options", CRIM, Sept. 1988.

5. "A Supercomputer Policy for Alberta, Report of the High Performance Computer Task Force of the Premier's Council on Science and Technology"; Alberta Government, Nov. 1991.

6. "HPC Inc., Business Plan"; the original business plan of the Alberta HPC High Performance Computing Centre, March 1992.

7. "HPC High Performance Computing Centre, Vision Document", HPCC, May 1994.

8. "A Proposal for an NSERC Collaborative Special Grant" (to support High Performance Computing and Communications in Western Canada); WurcNet, Nov. 1993.

9. "An ATM Trial Network an Five Application Projects, Proposal to CANARIE" (includes projects in high performance computing and communications); WurcNet Inc. Nov. 1993.

10. "White Paper, OilNet Calgary High-Speed Network" (for ATM access to HPCC); Fujitsu, Feb. 1994.

11. "A Proposal for the Extension of Provincial Support for the Ontario Centre for Large Scale Computation"; Management Board of OCLSC, Jan. 1991.

12. "A Large Scale Computation Evaluation Study for the Council of Ontario Universities"; Gillespie Folkner & Assoc., Sept. 1991.

13. "Reference Plan, Ontario High Performance Computing Facility, Enabling the Future; Competing in the World"; Advisory Committee on High Performance Computing formed by the Ontario Ministry of Education and Training, July 1993.

14. "Supercomputing Symposium Proceedings"; published by Super*Can for the past 8 years, containing hundreds of technical articles on high performance computing in Canada.

15. "Applications of Supercomputers in Industry and Government"; Ron Crossan, Feb. 1992.

16. "Reaction Draft Proposal for a Research Initiative in the Computational Arts, Sciences and Engineering"; by The Collaboratory Inc., on behalf of Memorial University of Newfoundland, regarding a high performance computing facility for the province, April 1994.

17. "A Report on the Ontario Research and Education Network (OREN)", the Ontario Ministry of Economic Development and Trade, June 1, 1993.

18. "Design for the Future ONet", ONet, March 1993.

**U.S. Documentation**

1. "Promoting High Performance Computing and Communications"; a Congressional Budget Office Study, June 1993.

2. "High Performance Computing and Telecommunications Infrastructure Development Activities in the U.S."; Art Monk & Associates, May 1994.

3. "High Performance Computing and Communications: Investment in American Competitiveness"; Gartner Group, March 1991.

4. "The National Science Foundation Supercomputer Centres"; NSF annual publication.

5. "Supercomputing and the Transformation of Science"; by William Kaufmann and Larry Smarr, Sept. 1992.

6. "Grand Challenges 1993: High performance Computing and Communications"; the US Committee on Physical, Mathematical and Engineering Sciences Federal Coordinating Council for Science, Engineering and Technology.

7. "Numerically Intensive Computing, Center Management Study"; RCI Inc., Third Quarter, 1993.

8. "RCI HPC User Survey: Multi-processing to Massive Parallelism"; RCI Inc., Second Quarter, 1993.

9. "HPC Industry Program Study"; RCI Inc., 1994.

10. "Trends in High Performance Networks"; RCI Inc., 1993.

11. "State of the Superperformance Computing Industry", Palo Alto Management Group, Inc., Oct. 1993.

12. "High Performance Computing and Communications, Technology for the National Information Infrastructure", supplement to the President's fiscal year 1995 budget.

Note: Detailed documentation is also available on most of the individual U.S. NSF centres and national laboratories.

**Japanese Documentation**

1. "A summary of High Performance Computing in Japan"; U.S. Office of Naval Research and German National Centre for Computer Science, June 1992.

2. "Supercomputer Trends in Advanced Economies", Ron Crossan for Fujitsu, Jan. 1992.

3. "Japanese Supercomputer Market"; Nikkei Computer, Jan. 1992.

**European Documentation**

1. "Report of the EEC Working Group on High Performance Computing"; Commission of European Communities, 1991.

2. "1992 RCI European HPC User Survey, Digest of Results"; RCI Europe, Fourth Quarter, 1992.

3. "Report of the High Performance Computing and Networking Advisory Committee, Volume 1"; Commission of the European Communities, Oct. 1992.

4. "Report of the High Performance Computing and Networking Advisory Committee, Volume 2"; Commission of the European Communities, Oct. 1992.

**International**

1. "National R&D Programmes for New Computer-Communications Networks and Applications"; OECD Directorate for Science, Technology and Industry, Feb. 1994.

2. "High Performance Computing and Communications Week"; Kings Communications Group, Washington D.C., weekly issues Aug. 93-Aug. 94.

3. "An assessment of High Performance Computer Technology in Japan and Europe"; Computer Research and Applications Group, Los Alamos National Laboratories, U.S.A., 1994.

4. "HPC Newswire"; weekly Internet service, Jan.-94 to Feb.-95.

## APPENDIX B

## Major HPC Sites World-wide

This listing was extracted and abbreviated from the Internet HPCwire service on August 19, 1994.

No guarantee of accuracy is given; and information has been collected from many diverse sources.

The sites are ordered by total levels of installed performance, starting with the largest site.

Ratings at the start of the first line of each site represent the ratio of the total installed HPC computer performance compared to a Cray Y-MP/1, based on NASA NPB BT Size A benchmark. For example, the first listing, the National Aerospace Lab in Japan has a total performance rating of 380.82 times the Cray YMP/1.

It appears that only systems currently installed or planned for a 1994 installation have been used to calculate the total rating. The sites listed after ranking 166 do not qualify as of August, but they do have equipment on order that enable them to qualify.

On the first line of each site, the rating is followed by a list of computer vendors that have supplied the HPC equipment. The vendor of the largest machine is listed first.

Figures suffixed "?" denote relative guesses.

The current minimum performance for inclusion is a benchmarked rating of four for systems that were benchmarked, or 10.56 Gflops peak for unbenchmarked systems.

Note that Environment Canada (ranking number 23) is the only Canadian system that qualifies, out of a total of 176 systems world-wide.

1) 380.82 Fujitsu, Intel
National Aerospace Lab, Chofu-shi, Tokyo, Japan

2) 190.37 TMC, Cray
Central Security Service, National Security Agency Headquarters, Fort George G Meade, Maryland, U.S.

3) 146.41 TMC, Cray
Los Alamos National Labs, Los Alamos, New Mexico, U.S.

4) 142.08 Cray
National Security Agency, Dallas, Texas, U.S.

5) 123.38 IBM, KSR
Center for Theory & Simulation in Science & Engineering, Cornell Uni, Ithaca, New York, U.S.

6) 112.97 Cray
Cray Research Computer Network, Eagan, Minnesota, U.S.

7) 106 Intel, Cray
Sandia National Labs, Albuquerque, New Mexico, U.S.

8) 104.38 TMC, Cray
Minnesota Supercomputer Center, Minneapolis, Minnesota, U.S.

9) 100.96 Cray, IBM, TMC, Intel
NAS, NASA Ames Research Center, Mountain View, California, U.S.

10) 98.4 Fujitsu, TMC
Angstrom Technology Partnership, Tsukuba, Japan

11) 83.52 Cray
Pittsburgh Supercomputing Center, Pittsburgh, Pennsylvania, U.S.

12) 82.4 Intel, Cray
Caltech, Pasadena, California, U.S.

13) 80.75 Fujitsu, Tsukuba
Uni of Tsukuba, Ibaraki, Japan

14) 79.52 Cray, Meiko
Lawrence Livermore National Labs, Livermore, California, U.S.

15) 76.82 Cray
Ford Engineering Computer Center, Dearborn, Michigan, U.S.

16) 73.07 NEC, Hitachi, Cray
Tohoku Uni, Aramaki, Sendai, Japan

17) 63.8 IBM
Maui High Performance Computing Center, Maui, Hawaii, U.S.

18) 62.5 Alenia
ENEA, Rome, Italy

19) 62.5 Alenia
Italian National Institute for Nuclear Physics, Rome, Italy

20) 57.92 IBM
IBM T J Watson Research Center, Yorktown Heights, New York, U.S.

21) 55.97 NEC
NEC Plant, Fuchu, Japan

22) 48.8 TMC, Cray, Convex
National Center for Supercomputing Applications, Uni of Illinois, Urbana-Champaign, Illinois, U.S.

23) 48.25 NEC
Environment Service, Canadian Meteorological Center, Dorval, Quebec, Canada

24) 46.62 NEC, Hitachi
Institute for Molecular Science, Okazaki National Research Institutes, Okazaki, Aichi, Japan

25) 43.52 Cray
Waterways Experiment Station Information Technology Lab, US Army Corps of Engineers, Vicksburg, Mississippi, U.S.

26) 43.52 Cray
Bettis Atomic Power Lab, West Mifflin, Pennsylvania, U.S.

27) 43.52 Cray
Knolls Atomic Power Lab, Schenectady, New York, U.S.

28) 43.52 Cray
National Meteorological Center, National Weather Service, NOAA, Suitland, Maryland, U.S.

29) 43.32 Intel
Oak Ridge National Lab, Oak Ridge, Tennessee, U.S.

30) 43.2 TMC
Naval Research Lab, Washington, DC, U.S.

31) 39.64 Intel, Cray
San Diego Supercomputer Center, Uni of California, San Diego, California, U.S.

32) 35.52 Cray
Agency of Industrial Science & Technology, Tsukuba, Japan

33) 35.52 Cray
Systems Engineering Research Institute, Korean Institute of Science & Technology, Yoosungku, Taejeon, South Korea

34) 35.52 Cray
United Kingdom Meteorological Office, Bracknell, Berkshire, England

35) 35.52 Cray
European Center for Medium-Range Weather Forecasts, Shinfield Park, Reading, England

36) 35.01 NEC, Fujitsu
NEC Supercomputer Center, Daito, Japan

37) 33.46 Cray, IBM
National Center for Atmospheric Research, Boulder, Colorado, U.S.

38) 32 Fujitsu
Fujitsu Parallel Computing Research Facilities Center Lab, Nakahara-ku, Kawasaki, Japan

39) 28 Intel
Intel Supercomputing Systems, Beaverton, Oregon, U.S.

40) 28 Cray
Uni of Edinburgh, Edinburgh, Scotland

41) 27.99 NEC
Swiss Scientific Computing Center, Swiss Federal Institute of Technology of Zurich, Manno, Switzerland

42) 27.99 NEC
National Institute for Fusion Science, Nagoya-shi, Aichi, Japan

43) 27.84 Hitachi
Hitachi Works, Kanagawa, Japan

44) 26.64 Cray
German Climate Computing Center, Hamburg, Hamburg, Germany

45) 24.87 NEC, Fujitsu
Japanese Atomic Energy Research Institute, Naka-gun, Ibaraki-ken, Japan

46) 24 Cray
Swiss Federal Institute of Technology of Lausanne, Lausanne, Switzerland

47) 24 IBM
Argonne National Lab, Argonne, Illinois, U.S.

48) 23.38 Cray Computer
Cray Computer, Colorado Springs, Colorado, U.S.

49) 21.07 Matsushita, Fujitsu
Kyoto Uni, Uji-shi, Kyoto, Japan

50) 20.88 Cray
Center for Nuclear Studies in Grenoble, French Atomic Energy Commission, Grenoble, France

51) 20.03 KSR, Cray
Leibniz Computing Center, Munich, Bavaria, Germany

52) 18.94 U-Tokyo
Uni of Tokyo, Meguro-ku, Tokyo, Japan

53) 18.56 Hitacihi
Hitachi, Tokyo, Japan

54) 17.76 Cray
National Center for Scientific Research (CNRS), Paris, France

55) 17.76 Cray
US Navy Fleet Numerical Oceanography Center, Monterey, California, U.S.

56) 17.44 Cray, Fujitsu
Power Reactor & Nuclear Fuel Development, Tokyo, Japan

57) 16.46 Cray, SRC
Supercomputing Research Center, Bowie, Maryland, U.S.

58) 16.32 Fujitsu
Fujitsu, Numazu-shi, Shizuoka, Japan

59) 16.29 KSR
US Army Research Lab, Aberdeen Proving Grounds, Maryland, U.S.

60) 15.66 Cray, Intel
Research Center Juelich (KFA), Juelich, North Rhine Westphalia, Germany

61) 15.04 Siemans (Fujitsu)
RW Technical Uni of Aachen, Aachen, North Rhine Westphalia, Germany

62) 15.04 Siemans (Fujitsu)
Technical Uni Darmstadt, Darmstadt, Hesse, Germany

63) 14 TMC
Geco-Prakla, Schlumberger, Houston, Texas, U.S.

64) 13.99 NEC
German Aerospace Research Establishment (DLR), Goettingen, Lower Saxony, Germany

65) 13.99 NEC
Toyota Central Research & Development Labs, Nagakute, Aichi, Japan

66) 13.99 NEC
Osaka Uni, Osaka, Japan

67) 13.32 Cray
Electricity of France, Clamart, France

68) 13.32 Cray
NASA Center for Computational Sciences, Goddard Space Flight Center, Greenbelt, Maryland, U.S.

69) 12.13 Hitachi
Uni of Tokyo, Bunkyo-ku, Tokyo, Japan

70) 12.06 NEC
National Aerospace Lab, Marknesse, Netherlands

71) 12.06 NEC
Houston Advanced Research Center, The Woodlands, Texas, U.S.

72) 12.06 NEC
IBM Tokyo Research Lab, Yamato-shi, Kanagawa-ken, Japan

73) 12.06 NEC
National Institute for Environmental Studies, Japan

74) 12 Cray
Arctic Region Supercomputing Center, Uni of Alaska, Fairbanks, Alaska, U.S.

75) 12 Cray
Chrysler Technology Center, Auburn Hills, Michigan, U.S.

76) 12 Cray
National Cancer Institute, Fort Dietrich, Maryland, U.S.

77) 12 Cray
Center for Study in Limeil-Valenton, French Atomic Energy Commission, Limeil-Valenton, France

78) 11.88 NEC
NEC, Miyamae-ku, Kawasaki-shi, Japan

79) 11.38 Intel
Uni of Colorado, Boulder, Colorado, U.S.

80) 11.2 TMC
Paris Institute of Global Physics, Paris, France

81) 11.2 TMC
American Express Information Processing Center, Phoenix, Arizona, U.S.

82) 11.2 TMC
Mobil Exploration & Producing Technical Center, Dallas, Texas, U.S.

83) 11.2 TMC
MIT, Cambridge, Massachusetts, U.S.

84) 10.44 Cray
Interuni Computing Center, Bologna, Italy

85) 10.32 TMC, KSR
Advanced Telecommunications Research Institute International, Souraku-gun, Kyoto, Japan

86) 9.25 Intel, Cray
Swiss Federal Institute of Technology of Zurich, Zurich, Switzerland

87) 8.88 Cray
Uni Stuttgart, Stuttgart, Baden-Wurttemberg, Germany

88) 8.88 Cray
Stichting Academic Computing Center Amsterdam, Amsterdam, Netherlands

89) 8.88 Cray
Meteo-France, Toulouse, France

90) 8.88 Cray
George C Wallace Supercomputer Center, Huntsville, Alabama, U.S.

91) 8.88 Cray
Daimler-Benz Information System House, Stuttgart, Baden-Wurttemberg, Germany

92) 8.88 Cray
DuPont de Nemours Experimental Station, Wilmington, Delaware, U.S.

93) 8.77 IBM
Queensland Parallel Supercomputer Facility, Griffith Uni, Brisbane, Queensland, Australia

94) 8.7 KSR
Pacific Northwest Lab, Richland, Washington, U.S.

95) 8 Cray
AAOC, U.S.

96) 8 Cray, IBM
NASA Langley Research Center, Hampton, Virginia, U.S.

97) 8 Cray
UK Atlas Supercomputing Center, Rutherford Appleton Lab, Chilton, Didcot, Oxon, England

98) 8 Cray
POPS SuperComputer Facility, US Naval Oceanographic Office, NASA Stennis Space Center, Bay Saint Louis, Mississippi, U.S.

99) 8 Cray
NASA Lewis Research Center, Cleveland, Ohio, U.S.

100) 8 Cray
Atomic Weapons Establishment, England

101) 8 Cray
Ohio Supercomputer Center, Columbus, Ohio, U.S.

102) 8 Cray
Center for High Performance Computing, Uni of Texas System, Austin, Texas, U.S.

103) 8 Cray
NOAA Geophysical Fluid Dynamics Lab, Princeton, New Jersey, U.S.

104) 8 Convex
Jozef Stefan Institute, Ljubljana, Slovenia

105) 7.66 Intel
Hong Kong Uni of Science & Technology, Kowloon, Hong Kong

106) 7.66 Intel
Purdue Uni, West Lafayette, Indiana

107) 7.5 IBM
Japanese National Cancer Research Center Institute, Tokyo, Japan

108) 7 NEC
National Institute for Space Research, San Jose dos Campos, Brazil

109) 7 NEC
NEC Scientific Information, Japan

110) 7 NEC
Obayashi, Japan

111) 7 KSR
Georgia Institute of Technology, Georgia, U.S.

112) 6 Cray
Lockheed Information Technology, Denver, Colorado, U.S.

113) 6 Cray
NASA Marshall Space Flight Center, Huntsville, Alabama, U.S.

114) 6 Cray
Boeing Computer Services, Seattle, Washington, U.S.

115) 6 Cray
Nissan Motor Research Center, Yokosuka-shi, Kanagawa, Japan

116) 6 Cray
Taiwan Central Weather Bureau, Taiwan

117) 6 Cray
Toyota Motor, Toyoda-shi, Aichi Prefecture, Japan

118) 6 IBM
Royal Shell Lab Amsterdam, Amsterdam, Netherlands

119) 6 IBM
Center for Advanced Studies Research & Development in Sardinia, Cagliari, Italy

120) 5.6 TMC
Real World Computing Partnership Research Center, Tsukuba, Japan

121) 5.6 TMC
Japan Advanced Institute of Science & Technology, Hokuriku, Japan

122) 5.6 TMC
Uni of Wisconsin, Madison, Wisconsin, U.S.

123) 5.6 TMC
Boston Uni, Boston, Massachusetts, U.S.

124) 5.6 TMC, NEC
German National Research Center for Computer Science, Saint Augustin, North Rhine Westphalia, Germany

125) 5.6 TMC
Epsilon, Burlington, Massachusetts, U.S.

126) 5.44 Siemens (Fujitsu)
Uni Hannover, Hannover, Lower Saxony, Germany

127) 5.44 Fujitsu
Taisei Construction, Tokyo, Japan

128) 5.44 Fujitsu
Fuji Heavy Industries, Japan

129) 5.44 Fujitsu
Kyushu Uni, Kyushu, Fukuoka, Japan

130) 5.44 Fujitsu
Nagoya Uni, Nagoya, Japan

131) 5.03 Intel
Indiana Uni, Bloomington, Indiana, U.S.

132) 5 Cray
Exxon Upstream Technical Computing, Houston, Texas, U.S.

133) 4.64 Hitachi
Nihon Uni, Chiyoda-ku, Tokyo, Japan

134) 4.64 Hitachi
Senshu Uni, Tokyo, Japan

135) 4.64 Hitachi
Hokkaido Uni, Kita-ku, Sapporo, Japan

136) 4.64 Hitachi
National Lab for High Energy Physics, Tsukuba, Japan

137) 4.44 Cray
Bayer Central Research Facility, Leverkusen, North Rhine Westphalia, Germany

138) 4.44 Cray
UNI-C, Technical Uni of Denmark, Lyngby, Denmark

139) 4.44 Cray
National Institute of Meteorology, Madrid, Spain

140) 4.44 Cray
National Environmental Supercomputing Center, EPA, Bay City, Michigan, U.S.

141) 4 Cray
Merck Research Labs, Rahway, New Jersey, U.S.

142) 4 Cray
BMW, Munich, Bavaria, Germany

143) 4 Cray
National Office of Aerospace Computing Studies, Chatillon, France

144) 4 Cray
Toshiba, Kanagawa, Japan

145) 4 Cray
CSIRO Supercomputing Facility, Carlton, Victoria, Australia

146) 4 Cray
Max-Planck-Institute, Garching, Bavaria, Germany

147) 4 Cray
Peugeot-Citroen, Velizy, France

148) 4 Cray
Hyundai Passenger Car Engineering Center, Ulsan, Korea

149) 4 Cray
Samsung Advanced Institute of Technology, Yongin-gun, Kyung Gi-do, Korea

150) 4 Cray
Matsushita Electric Works, Moriguchi, Osaka, Japan

151) 4 Cray
Uni of Trondheim, Trondheim, Norway

152) 4 Cray
Konrad-Zuse-Center for Information Technology, Berlin, Germany

153) 4 Cray
North Carolina Supercomputing Center, MCNC, Research Triangle Park, North Carolina, U.S.

154) 4 Cray
Mitsubishi Motor, Minato-ku, Tokyo, Japan

155) 4 Cray
Defense Research Agency, Ford Halstead, England

156) 4 Cray
National Autonomous Uni of Mexico, Mexico

157) 4 Cray
German Weather Service, Offenbach on the Main, Hesse, Germany

158) 4 Cray
Supercomputing Research Institute, Florida State Uni, Tallahassee, Florida, U.S.

159) 4 Cray
Uni College Dublin, Belfield, Dublin, Ireland

160) 4 DEC (Maspar)
Parallab, Institute for Information, Uni of Bergen, Bergen, Norway

161) 4 DEC (Maspar)
European Institute for Genomic Research, Heidelberg, Baden-Wuerttemberg, Germany

162) 4 Maspar
Lockheed Missiles & Space, Palo Alto, California, U.S.

163) 4 DEC (Maspar)
US Army Theater High Altitude Area Defence, Northern Californian Ordnance Proving Ground, California, U.S.

164) 4 DEC (Maspar)
US Army Theater High Altitude Area Defence, White Sands Missile Range, White Sands, New Mexico, U.S.

165) 4 DEC (Maspar)
US Army Theater High Altitude Area Defence, United Technologies, SanJose, Cal., U.S.

166) 4 DEC (Maspar)
US Army Theater High Altitude Area Defence Site 4, U.S.

12 Cray
Phillips Petroleum, Bartlesville, Oklahoma, U.S.

6.25? Parsytec
Uni, Germany

67.2 Fujitsu
Institute of Physical & Chemical Research, Wako-shi, Saitama, Japan

24 Fujitsu
Communications Research Lab, Tokyo, Japan

16.8 Fujitsu
Institute of Space & Astronautical Science, Tokyo, Japan

8 Alenia
German Electron-Synchrotron, Hamburg, Hamburg, Germany

7.72 Archipel
Lab for High Performance Computing, ENS Lyon, Lyon, France

4.79 IBM
North Eastern Parallel Architectures Center, Syracuse Uni, Syracuse, New York, U.S.

6.24 Cray
Hawtal Whiting, Basildon, Essex, England

4.79 IBM
Uni of Southampton, Southampton, England

# APPENDIX C

## List of Advisory Committee Members

Dr. Andrew Bjerring
President
CANARIE Inc.
410 Laurier Ave. West, Suite 1120
Ottawa, Ontario K1P 6H5

Mr. David Burden
Senior Director, Information Services
Air Canada
ZIP-045, P.O. Box 9000
Saint-Laurent, Quebec H4Y 1C2

Dr. Arthur Carty
President
National Research Council
Montreal Road
Ottawa, Ontario K1A 0R6

Mr. Paul Davis
President & CEO
HPC High Performance Computing Centre
3408, 400-3rd Avenue S.W.
Calgary, Alberta T2P 4H2

Dr. Kirk Dawson
Executive Director
Canadian Institute for Climate Studies
650 Hastings Avenue, Suite 765
Penticton, British Columbia V2A 7N1

Mr. Gordon W. Gow
Assistant Deputy Minister
Ontario Government, MEDT
Technology and Communications
56 Wellesley Street W.
Toronto, Ontario M7A 2E4

Dr. Inge L. Hansson
President
Ashurst Technology Centre, Inc.
2486 Dunwin Drive
Mississauga, Ontario L5L 1J9

Dr. Kevin Keough
Vice-President, Research
Memorial University of Newfoundland
St. John's, Newfoundland A1C 5S7

Dr. Cooper Langford
Vice-President, Research
University of Calgary
2500 University Drive
Calgary, Alberta T2N 1N4

Mr. Andrew Macdonald
Chief Informatics Officer
Information Management, Systems & Technology
Treasury Board Secretariat
Ottawa, Ontario K1A 0R5

Ms. Elizabeth Pearce
President, Super*Can
51 Sandringham Drive
Toronto, Ontario M5M 3G4

Dr. Denis Poussart
Vice-President
CRIM
1801 McGill College Avenue
Montreal, Quebec H3X 2H9

Mr. Hugh Stanfield
President & CEO
Pulsonic Corporation
510, 400-3rd Avenue S.W.
Calgary, Alberta T2P 4H2

Mr. John Taylor
Vice-President Engineering
Bombardier Inc., Canadair
P.O. Box 6087
Station Centre Ville
Montreal, Quebec H3C 3G9

Ms. Harriet Velazquez
Vice-President, Delivery Programs
Canadian Imperial Bank of Commerce
901 King Street West, 7th Floor
Toronto, Ontario M5V 3H5

Dr. Donald Weaver
Dept. of Chemistry and Neurology
Queen's University
Kingston, Ontario K7L 3N6

**DATE DUE**
DATE DE RETOUR

| | |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

1575, boulevard Chomedey, Laval (Québec), H7V 2X2 Téléphone: (514) 973-5700 Télécopieur: (514) 973-5757

1575, Chomedey Boulevard, Laval, Quebec, H7V 2X2 Telephone: (514) 973-5700 Facsimile: (514) 973-5757