HPC · CHP

ENGINES OF DISCOVERY: THE 21ST CENTURY REVOLUTION

THE LONG RANGE PLAN FOR HIGH PERFORMANCE COMPUTING IN CANADA

The development of this Long Range Plan was initiated and supported by the Executive and Board of the C3.ca Association Inc.

> *"As someone who is not doing grand-challenge computing, I asked "Does this plan offer anything to me?" Other researchers across Canada may wonder the same. I started writing down these points and was quickly convinced."*
>
> *Dr. Andrew Rutenberg*
> *Dept. of Physics and Atmospheric Science*
> *Dalhousie University*

The Benefits of the Long Range Plan

Advantages for the researcher

- **Access: clear opportunities for obtaining access to powerful national and regional computing resources**
  - streamlined and simplified resource application process
  - ongoing competitive resource allocation opportunities
- **Opportunity: support for ambitious, time-critical research projects (mitigates funding-cycle delays and uncertainties)**
- **Sustainability: access for ambitious projects on a long-term basis (allows planning for research and for personnel development)**
- **Coordination:**
  - ready access to appropriate regional- and national-level infrastructure
  - shared salary-support programs for skilled local technical staff
  - growth of a pool of talented technical support people in Canada
  - development of a national network of technical support available to all users
- **Autonomy: local control for powerful desktop and group computing facilities**

Advantages for the federal and provincial granting agencies

- **Clarity: clear planning and implementation**
- **Consistency: clear determination of resource needs and proposal quality**
- **Efficiency: coordinated use of computational hardware and personnel**
- **Competitiveness: preservation of an internationally competitive research environment**
- **Coherence: coordinated dissemination of techniques and results across Canada**
- **Transparency: appropriate and well-managed injection of resources**
- **Cost effectiveness: coordinated national plan eliminates the redundancy of independent local efforts**
- **Financial responsibility: financial and management accountability**

Advantages to industry

- **One-stop-shopping:**
  - access to significant computational resources for research
  - rapid scientific solutions for problems and for product development
  - a low-cost try-before-you-buy approach to HPC for the SME base
  - a well-maintained researcher knowledge base
  - a technology watch service with timely alerts
  - HQP: access to a guaranteed skilled workforce for the knowledge economy
  - Killer apps: today's leading-edge research builds tomorrow's leading-edge industry

Table of Contents

Table of Contents

Case Studies

Preface

High performance computing (HPC) is transforming research in Canadian universities, hospitals and industry. Computer simulations and models now supplement or even supplant traditional field or laboratory experiments in many disciplines. In addition, massive datasets from large-scale field experiments are being manipulated, stored and shared. Numerical laboratories open up otherwise inaccessible realms and provide insights that were inconceivable a few years ago.

Research worldwide has seen a dramatic increase in the demand for HPC in the traditional areas of science and engineering, as well as in the social sciences and humanities. In 1999, Canada occupied an inconsequential position in HPC based research, but that year saw the first funding of HPC by the Canada Foundation for Innovation. The subsequent combination of federal, provincial and industrial funding is enabling Canada to attract international experts, develop a strong foundation in HPC research, and train highly qualified personnel. The number of researchers using HPC has increased from a few hundred in 2000 to several thousand in 2004 (and that number is rapidly growing).

Today, HPC is ubiquitous. It affects our everyday lives in ways that most people cannot imagine. This document illustrates this in two ways: first, through a narrative that shows the many impacts of HPC in the day of an ordinary Canadian; and, second, through case studies that illustrate how research using HPC is shaping the future. These illustrations highlight the need for support if Canada is to be competitive in the "knowledge economy".

Despite the progress of the past five years, there is currently no clear plan to sustain research involving HPC. Our history clearly shows that *sporadic* HPC initiatives are not cost-effective, competitive or of lasting impact. C3.ca – the organization that coordinates HPC in Canada - recognizing the need for a long-term vision and stable funding, convened an expert panel to develop a Long Range Plan (LRP) for HPC. This document is the result. The work of the panel has been supported by the National Research Council of Canada, the Canada Foundation for Innovation, the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council, the Social Sciences and Humanities Research Council, and CANARIE Inc.

The panel began its work in December 2002. Since then it has held both regional meetings (Calgary, Toronto, Ottawa, Montréal and Halifax) and local town-hall meetings across Canada, it has widely distributed a summary report, and it has sought community input as the plan developed.

> *Preparing a Long Range Plan for HPC is especially challenging because it is an enabling technology for an extremely diverse set of researchers and it affects more researchers than any other major research infrastructure in Canada.*

Preparing a Long Range Plan for HPC is especially challenging because it is an enabling technology for an extremely diverse set of researchers and it affects more researchers than any other major research infrastructure in Canada. The industrial impacts are also widespread, ranging from the retail and natural resources sectors to very high technology areas such as aerospace and biotechnology. In this document, we cannot represent all the areas impacted by HPC. We have instead provided a cross-section of applications in the form of case studies: these studies represent the (broadly defined) scientific justification for sustained funding of Canadian HPC. The five chapters describe the technology, the human resources, the proposed management, the outreach structure, and the required funding for the plan. Sidebars throughout the document provide supporting information for the primary material. The report has been written recognising that many people will read sections of particular interest rather than the entire report and as a consequence the report does have some repetition of key points and examples.

It should be emphasised that this Long Range Plan focuses on a vision and general recommendations. HPC is a rapidly changing environment and the details of what will be needed at any given time need to be established in the context of the state-of-the-art at that time.

As the LRP was completing its final round of review, the need for a Canadian strategy was highlighted by two major announcements that demonstrated the importance other countries are placing on HPC. First, the US Congress recently passed H.R. 4516, the Department of Energy High-End Computing Revitalization Act of 2004, which will further U.S. supercomputing efforts by establishing a research and development (R&D) program within the Department of Energy (DOE) to develop new, record-breaking computing capabilities. The bill also authorizes the DOE to establish supercomputer user facilities that will provide U.S. researchers access to some of the world's most advanced computers. Second, the German Science Council recently recommended that three top-ranked high performance computers be created in Europe with one of them to be located in Germany. They noted that on the current Top500 HPC list, there are 128 supercomputers in European countries [compared to 7 in Canada]. The Science Council recommended that attempts to situate one such European supercomputer in Germany would cost the German government 30 to 50 million euros.

The Long Range Plan Authors Panel has benefited from the input willingly provided by a large and diverse range of researchers both inside and outside of Canada, including those who assisted with the preparation of the cases, the LRP Advisory Panel and the Board of Directors of C3.ca. We are very grateful for their time and their insights. We appreciate the enthusiastic support of the community as the LRP was refined. Feedback from community members has substantially influenced the form of this document and we are indebted to them. Others who have substantially contributed include Dr. Andrew Pollard, Ms. Cathy King, Ms. Leontien Storrier, Dr. Paul Lu and Ms. Claudette Tourigny.

Finally, I am indebted to the panel members who have devoted an enormous amount of time, energy, wisdom and scientific depth to the preparation of this document. I greatly appreciate their hard work, good humour and friendship throughout this challenging, educational and enjoyable process.

R. Kerry Rowe, P.Eng., FRSC, FCAE
Chair, Long Range Plan Authors Panel
Queen’s University, January 2005

On behalf of:
Jonathan Borwein, Simon Fraser University and Dalhousie University
Russell Boyd, Dalhousie University
Gilbert Brunet, Environment Canada
Hugh Couchman, McMaster University
Alan Evans, McGill University
Martyn Guest, Daresbury Laboratory, England
Ian Lancashire, University of Toronto
Jonathan Schaeffer, University Alberta

Foreword

Around the world, the application of ever-advancing information and communications technologies has revolutionized virtually every branch of enquiry, from the study of climate and ocean currents to the development of theories of protein crystalline structures and sub-atomic particle interactions. All such work is now inherently quantitative and computational: it requires both high performance computers to undertake the mathematical analysis and advanced networks to gather and share the increasingly massive data sets. What is at issue is no less than using these new technologies to invent better ways of conducting research.

Over the past decade in Canada, a new *infrastructure* has been built to support this critical transformation of the research enterprise. With strategic funding from the Canada Foundation for Innovation (CFI) and Industry Canada, as well as from provincial and institutional sources, Canadian researchers now have access to a range of regional high performance computing (HPC) resources that a mere ten years ago was the stuff of dreams, and to a pan-Canadian array of advanced networks to link them together. Both the regional HPC facilities and the networks are competitive with the best in the world, an achievement that should be celebrated. They also constitute cornerstones on which to build, since the revolution in how research is conducted - and in how economic and social benefits are derived from that research - is just beginning.

The Long Range Plan for HPC outlined in this report provides a superb platform on which to continue this revolution in Canada. Sustained funding for HPC infrastructure to advance what has been achieved through past investments, and on-going organizational development to facilitate and spearhead these collective efforts, are the obvious next steps. The members of the HPC community are to be applauded for having taken the initiative in developing this plan and focusing the debate with such well-argued recommendations.

Indeed, continued investment in *technological* innovation, exemplified by the computer installations and networks deployed in Canada, is the critical first half of a comprehensive strategy to maintain Canadian competitiveness in HPC areas. It is an expensive half, of course, for this infrastructure soon becomes obsolete and so must be constantly refreshed. But over time, such sustained investments will lead to an *integrated* "intelligent" or "cyber" infrastructure that will provide the research community with the essential tools needed for world-class work.

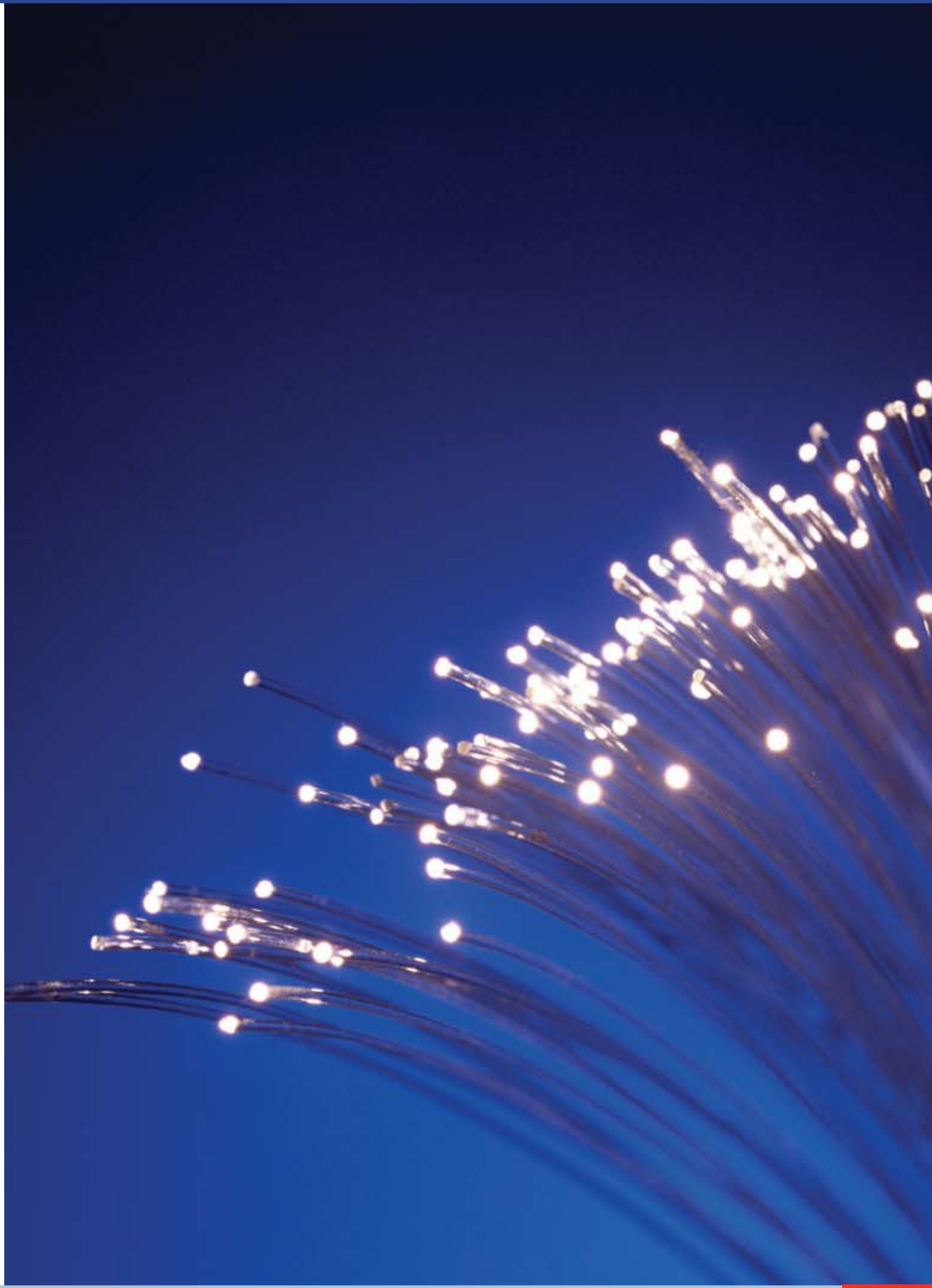
*For its part, CANARIE provides the national backbone, CA*net 4, that is so essential for HPC based research, and supports parallel development in areas such as e-learning, e-business and other network-enabled research. We also facilitate process and organizational innovation. In keeping with the role of facilitating innovation, we will continue to work with the HPC community and others to ensure the emergence of Canada's intelligent infrastructure.*

Also essential, however, is the second half of the strategy: continued investment in how Canadian and international research communities *organize* themselves to go about their business. In Canada, we have taken a consortium approach to addressing this challenge, as exemplified by the regional HPC consortia, C3.ca, Grid Canada and even CANARIE itself. The result has been a set of robust, multi-disciplinary, multi-sectoral collaborations built around strong institutional building blocks.

The linkage between these two aspects of a comprehensive Canadian strategy to guide and strengthen our research competitiveness – *technological* innovation and *organizational* innovation – is especially close, since building and using intelligent infrastructure requires new ways of collaborating. New arrangements must be created and new architectures designed to govern how data, software, networks, computers, sensors and other instrumentation, as well as facilities such as telescopes and synchrotrons, are to be shared and jointly controlled.

For its part, CANARIE provides the national backbone, CA*net 4, that is so essential for HPC based research, and supports parallel development in areas such as e-learning, e-business and other network-enabled research. We also facilitate process and organizational innovation. In keeping with the role of facilitating innovation, we will continue to work with the HPC community and others to ensure the emergence of Canada's intelligent infrastructure. Indeed, CANARIE and the HPC community have a critical shared goal: ensuring that Canada has both the infrastructure and the organizational arrangements to enable our researchers to not only be effective *collaborators* with the best in the world, but to be successful *competitors* as well.

Andrew K. Bjerring
President and CEO
CANARIE Inc.



One Day

One Day

High performance computing (HPC) affects the lives of Canadians every day. We can best explain this by telling you a simple story. It's about an ordinary family on an ordinary day: Russ, Susan, and Ceri Sheppard. They live on a farm 15 kilometres outside Wyoming, Ontario. The land first produced oil, and now it yields milk; and that's just fine locally.

Their day - Thursday, May 29, 2003 – begins at 4:30 a.m. when the alarm goes off. A busy day. Susan Zhong-Sheppard will fly to Toronto to see her father, Wu Zhong, at Toronto General Hospital; he's very sick from a stroke. She takes a quick shower and packs a day bag for her 6 a.m. flight from Sarnia's Chris Hadfield Airport. Russ Sheppard will stay home at their dairy farm, but his day always starts early. Their young daughter Ceri can sleep three more hours until school.

Waiting, Russ looks outside and thinks, *It's been a dryish spring. Where's the rain?*

In their farmhouse kitchen on a family-sized table sits a PC with a high-speed Internet line. He logs on and finds the Farmer Daily site. He then chooses the Environment Canada link, clicks on Ontario, and scans down for Sarnia-Lambton. Good, as he hoped:

Periods of morning rain. Risk of a thunderstorm. Wind south 30 km/h gusting to 50 diminishing to 20 in the afternoon. Sunny periods. High 14

One of a million who check that site every week, Russ uses HPC, although he doesn't know he does. Two decades ago, Russ found Environment Canada reliable only two days ahead. Now, the forecasting works pretty well for a week.

Russ thinks, *They've got some smart people at our Meteorological Service.*

But Russ worries. The farming Sheppards go back four generations to the thirties, times of hunger in the evening. A picture in the hall landing reminds him of 'those days'.

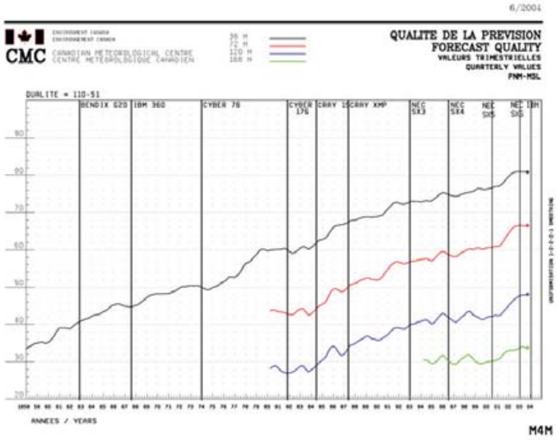
He sits a while, staring at the screen. Ten days ago Russ felt his world shake. BSE, bovine spongiform encephalopathy, appeared in one cow in Wanham, Alberta, and now seven countries have banned our beef. Russ gets tired thinking about this because he can do nothing about it.

What would he and Susan do if their cows had to be destroyed? And now there's her father, and he can't burden her with what he thinks.

Russ calls up *www.google.ca* and enters "bse alberta". The news describes the slaughter of a thousand suspect Alberta livestock. But an article from last March details BSE testing in the Animal

Weather Prediction

The accuracy of a five-day forecast in the year 2003 was equivalent to that of a 36-hour forecast in 1963 [REF 1]. The quality of daily forecasts has risen sharply by roughly one day per decade of research and advances in HPC hardware and software. Accurate forecasts transform into billions of dollars saved annually in agriculture and in natural disaster costs. Using a model developed by Dr. Keith Thompson at Dalhousie University, the Meteorological Service of Canada has recently been able to predict coastal flooding in Atlantic Canada early enough for the residents to take preventative action.



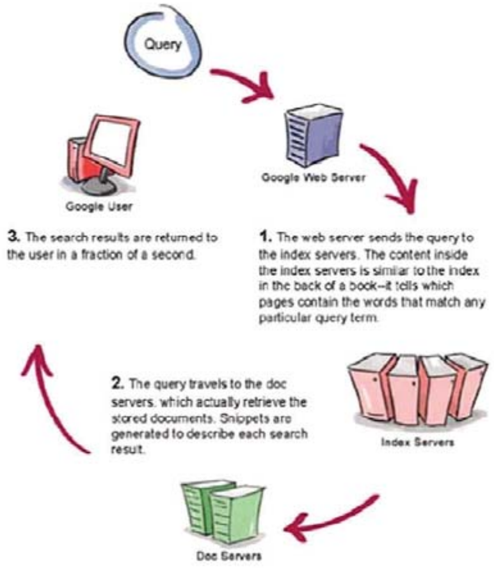
The figure charts the steadily increasing quality of weather prediction

Web Searching

Google [REF 2] is powered by over 30,000 PCs performing smart computer mathematics, and has three distinct HPC components which allow it to find and organize news, science and much else:

- Googlebot, a web crawler that finds and fetches web pages.
- An indexer that sorts every word on every found web page in a huge database.
- A query processor, which compares your search query to the index and extracts relevant documents.

Twenty-four hours a day, Google typically responds to a query in less than a second – thanks to HPC.



1. The web server sends the query to the index servers. The content inside the index servers is similar to the index in the back of a book--it tells which pages contain the words that match any particular query term.

2. The query travels to the doc servers, which actually retrieve the stored documents. Snippets are generated to describe each search result.

3. The search results are returned to the user in a fraction of a second.

Health Laboratory at the University of Guelph: no positives at all in 2002. A few tissue samples from dead animals went to a Canadian Food Inspection Agency lab in Winnipeg. He locates the CFIA site and its news release from yesterday. They're taking samples for BSE testing now in Saskatchewan, and a CFIA lab in Lethbridge is doing DNA analysis to trace that infected cow. These biohazard researchers rely on HPC, just as Russ unknowingly does in using Google.

Russ says to himself, *Ok, things are happening!* and then he calls out, "Susan, ready to go?" Russ clicks the programmable keyless entry system for his new SUV.

The drive to Sarnia's Chris Hadfield airport takes twenty minutes, and an EDR (event data recorder) in the driver's airbag tracks their SUV velocity, engine speed, and the status of their brakes. Russ taps the car's electronic compass. Susan gets out her cell phone to check departure time. After his "Give my best to Dad," a peck, a hug, and her "Pick me up at one, and don't let Ceri sleep in again!", Susan is out of the car and through the gates.

Belting herself into the 18-seater Air Georgian Beech 1900H, she thinks, *Oh my. The life of an only child, wife, and mother!* Outside the window, it's raining. Susan peers up at the air traffic control tower as the plane taxis out. She crosses her fingers. Since rising, she and Russ, in their rural world, have used three massive HPC-based systems.

Now she places her life in the tender mercies of another. The same company that built the Beech 1900H airplane, Raytheon, also made CAATS, the Canadian Automated Air Traffic System headquartered in Moncton. None of Susan's fellow passengers knows that NavCanada's flight data processing system ranks as the world's best. If anyone told them, most wouldn't be surprised.

A Canadian-made prop engine lifts Susan 20,000 feet into the sky. This Pratt & Whitney PT6A engine has been in continuous demand and development for over forty years. If you've been in an Otis elevator, you've been lifted up by yet another Pratt & Whitney product. The approach of Susan's flight to Toronto's Pearson Airport takes her over the offices of the HPC company that helps keep her airborne.

CATTS brings Susan down safely into what her friends are grimly calling "plague city" these days. Born in oil-rich Calgary, Susan heard Dad grumble about eastern Canada in the 1970s, but now he's an Easterner and she thinks, *Life's like that, isn't it? It takes you to strange places*. Like Russ, she worries, although not about an Alberta cow but about a new virus from China that rode the skies into Pearson Airport two months ago. Wu Zhong's an innocent in the thick of SARS (Sudden Acute Respiratory Syndrome). Last Monday, the World Health Organization listed Toronto as a city that had recently transmitted SARS locally, a blow to a big town already reeling in isolation.

Susan comes anyway. She has to talk to the doctor.

The plane taxis up to Terminal 2, and Susan walks down the stairs and across the tarmac in the rain. The (area code) 905ers were already heading *en masse* into downtown using the two huge highways, 401 and 427, that border on Pearson. The late spring pollution billows past, and Susan's throat catches. Asthmatic since childhood, she takes out her inhaler. The first puff of the day and her watch says barely 6:55.

Things look worse when Susan sits down at a Tim Horton's on the Terminal 2 Departures level. She buys a rimless coffee and looks at the morning's papers abandoned at her table. Yesterday, Toronto placed 1700 students and staff at Markham's Father Michael McGivney Catholic Academy in home quarantine until June 3, because just one student with SARS symptoms attended school last week. His mother works at North York General Hospital, one of the affected hospitals. Worse, the papers describe how Chinese Canadians, labelled as SARS carriers, along with their businesses and their children have been suffering from racism, boycotting and shunning. "This is ridiculous. What am I doing here?" Susan says to herself. And from the next table comes a voice, "Just what I was wondering, dear. . . . No one wants a person *like you* deplaning from SARS headquarters today, do they?" Susan answers, "What do you mean by that? I'm from Sarnia, not Toronto or China!" But the voice persists, "Oh, are you? . . . Well, if I were you, dear, I'd watch my step today."

With as much dignity as she can muster, Susan gets up and hurries down the terminal concourse, looking for the city buses bound downtown. *Right enough*, she thinks, *where are the frequent fliers today?* A few persons with rigid expressions walk briskly by, pointedly keeping their distance from her. Outside the automatic doors, a Toronto Transit bus waits. Susan gets on, pays, and goes right to the back.

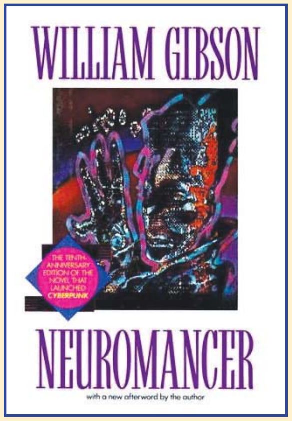
Sometimes, as at 7:25 a.m., when Susan's bus joins the 427 southwards highway, Toronto roads flow like refrigerated peanut butter. At other times, they are speedways. Drivers generally know what to expect according to when they drive, and that predictability, a function of Toronto's regulated traffic system, calms them. In mid-July of 1999, when a switchboard panel exploded at a Bell Canada building, traffic lights at 550 Toronto intersections failed. . . . This was a hint of what lays in store for this town on August 14, 2003, when all power will cease on the eastern seaboard. But this morning everything's OK; an HPC system, fed with information from magnetic sensors and fibre optic cables, controls the city's traffic lights.

Susan watches the city go by as the bus turns east onto the Gardiner Expressway and exits at Bay Street. A big yellow billboard with a large arrow pointing upward, advertising the Weather

Designing An Airplane

After Pratt & Whitney added HPC to its tools – a single computer grid governing 15,000 computers in North America – its products were engineered in half the time and cost. Platform Computing, the HPC firm that Pratt & Whitney hired, has its headquarters in Markham and offices in 15 cities from Boston to Beijing. Computational fluid dynamics research predicts how the Pratt & Whitney gas turbine engine burns fuel and so allows improvements in its engineering. [REF 3]

"Wingmating" – wing making in the Bombardier CRJ900



Canada's leading science fiction writer, William Gibson, coined the word "cyberspace" in his novel *"Neuromancer"* which won the Hugo, Nebula, and Philip K. Dick awards in 1984–85. *"Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city light . . ."* [REF 4]

William Reeves, Technical Director of Pixar Animation Studios, studied in Waterloo and Toronto, getting a B.S. in math and an M.A. and a Ph.D. in computer science. In 1980, he joined Lucasfilm and invented an image-synthesis technique called Particle Systems. It generates complex and detailed images through many (normally) simple geometric particles changing stochastically over time. After working on films like *The Empire Strikes Back*, Reeves and two partners created Pixar. Today his technique, taught everywhere in film courses, powers many of Pixar's groundbreaking animated films. As supervising Technical Director on *Toy Story*, the first feature film created entirely by computer animation, Reeves won four technical Academy Awards. Another Canadian, Darwyn Peachey (BSc and MSc from Saskatchewan), has played a key technical role with Pixar, receiving a technical Oscar in 1995 for his part in the development of *RenderMan*.

Patrick Coleman and Karan Singh, from the Dynamic Graphics Project, the University of Toronto research group of which Reeves is an alumnus, developed nonlinear projection software for Chris Landreth's "Ryan", which received the Academy Award for Best Animated Short film of 2004. The Canada Council for the Arts participated in the development of "Ryan". [REF 5]

Ryan, a victim of drug addiction and alcoholism, experiences states of mind that affect his perception of the space around him. Computer animation software normally uses the rules of linear perspective, but to show how characters like Ryan see their world, those rules had to be modified. The nonlinear projection system used to create Ryan allows animators to combine multiple points of view in various ways to achieve effects that distort linear perspective.

Pixlet (Pitxar + wavelet) is the first studio-grade algorithm for fimmakers. Pixlet provides 20-25:1 compression, allowing a 75 Megabytes/sec series of frames to be delivered in a 3 Megabytes/sec movie, similar to DV Data.

Network, has the caption, "(We told you so.)" It was right again. A wet bus pulls up to the Royal York Hotel. "Good for the farmers," the bus driver observes evenly, as Susan gets off. She enters the hotel and quickly steps downstairs to the underground walkway. It's 8 a.m. and 75 minutes until she meets with her father's doctor. Susan wants a gift for Ceri, so she'll browse in what Toronto says is the world's biggest subterranean complex, 27 kilometres of halls with a thousand shops. At an early-opening CD shop in the underground city she picks out *Finding Nemo* because Ceri liked *Toy Story*, another Pixar animation. Didn't a Canadian win an Oscar for that? Susan doesn't remember Bill Reeves' name and has no idea that Pixar uses eight supercomputing clusters and sixty terabytes of disk space, called a "RenderFarm", to make entertainment for folks on farms like hers. There's no reason she should remember. If digital processors were pixels on a North American screen, Toronto's underground city and the Bay Street markets above it would shine out like a beacon. But like most Canadians, Susan never sees the cyberspace world all around her.

At 9 a.m. Susan exits the underground at City Hall, turns west to University Avenue, and finds that Toronto General Hospital is in sight. In ten minutes, she's at the Peter Munk Cardiac Centre's entrance on Gerrard Street.

She meets a sign: "Patients Infectious Disease Screening. "Two gowned and masked nurses block

Susan's way. They give her a mask, squirt some antibacterial soap on her hands, and tell her to rub them well. Susan is still outside the door. Then one nurse asks, "Who are you and why are you here?" "I'm Susan Zhong-Sheppard from Wyoming near Sarnia, and I've an appointment with Dr. Christopher Andres in Neurosurgery about my father, Wu Zhong. He's a patient here." The nurse firmly says, "Didn't you read this sign? We have an emergency situation here! You can't visit your father. Don't you know about SARS?" "I'd hoped . . ." Susan says weakly. The other nurse consults an appointment sheet and says, "You're not listed here. Well, take a seat over there and we'll see."

Susan joins some silent, masked people seated just inside the door. Orderlies send most away. Forty-five minutes pass. Her return flight is scheduled to leave at noon, and she frets. At 10 a.m. a nurse tells her to come down the hall. She sits next to a computer and another gowned, masked nurse kindly asks, "Do you have a cough, fever, headache, muscle pains, or shortness of breath?" At this, Susan's asthma acts up, and the nurse becomes quieter. "No? Please, Mrs. Zhong-Sheppard, step into this room so that we can examine you." Researchers further east on highway 401, at Queen's University, use HPC to simulate how inhalers work so that people like Susan will get more effective relief. She could use that relief today.

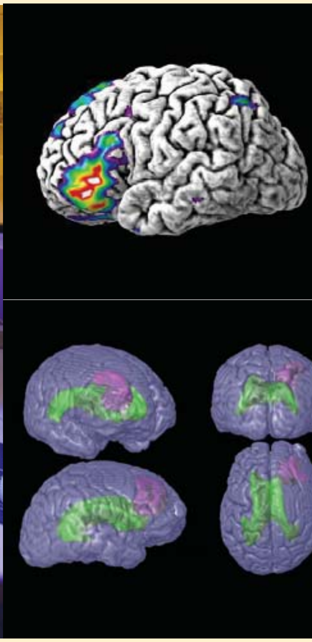
Susan shows the intern her inhaler. He takes her temperature and asks more questions. Twenty minutes later, a shaken Susan is sent out. The nurse at the computer workstation says, "I'm sorry, but this is the new normal. We've checked and Dr. Andres will see you now. Take this sheet and get everyone you meet to sign it. When you leave, hand in the sheet. Do you understand?"

Susan makes her way to a fourth-floor waiting room in the west wing. A quiet masked nurse signs her sheet. Ten minutes later, the nurse accompanies her to a small consulting room. Dr. Andres walks in, masked. "I'm sorry you find us this way, Susan, but I'm glad you came. Did you make an appointment?" Susan says, "Yes I did, really." "Well, all that matters is you're here now. Your father is stable, but he still can't talk. Susan, I need you to let us create dynamic pictures of his left hemisphere. They're called functional magnetic resonance images. We're well set up to analyze them . . . have you heard of our Functional Imaging Research Network?"

"Dr. Miner sent my father here. Please, Dr. Andres, whatever's best for Dad." "Susan, we work with the Heart and Stroke Foundation and the Universities of Ottawa and Toronto, and we can design a good therapeutic strategy for Mr. Zhong." Susan doesn't know what to say. She wants to take her father home *now*.

Medical Imaging

Modern HPC imaging techniques (such as PET using 'positrons' and SPECT using 'photons') provide non-invasive two- and three-dimensional real-time dynamic images for the brain, heart, kidney and other organs. They are revolutionizing research, surgery and disease management. A one-minute three-dimensional reconstruction requires enormous computing power to generate these images. [REF 6]



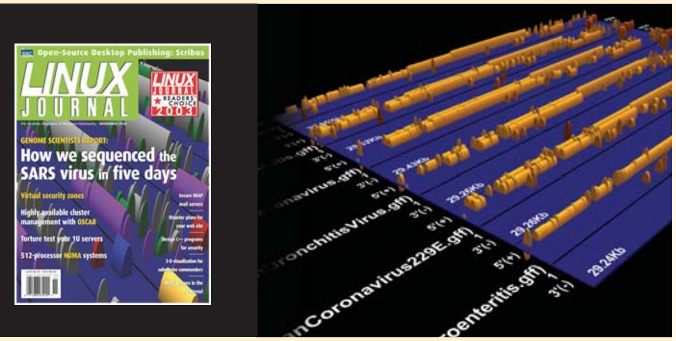
Sequencing SARS

"The rapid dissemination of the SARS genomic sequence through the Internet allowed groups all around the world to participate in the computational analysis of the genome and the 3D modeling of the protein sequences, ultimately resulting in an unprecedented number of scientific publications and vaccine initiatives just over one year after the discovery of this emerging disease. Such initiatives have typically taken many years."

Dr. Steven Jones

Head, Bioinformatics

Genome Sciences Centre, BC



The image shows a genomic (BLASTX) analysis of the protein complement of the *Nidovirales* family against build 7.4 of the SARS-suspected virus sequenced at Canada's Michael Smith Genome Sciences Centre. [REF 7]

He understands. "We can't move patients around these days. We're under siege here. I'll let you know soon, ok? But I've really got to go now, Susan. Are you all right?" She reaches for her inhaler but she nods. "Good. I'll call you next week, ok? I'm sorry . . . I am out of time!" Dr. Andres signs Susan's sheet and leaves her there.

When Susan reaches the Gerrard Street exit, it is 11:20 a.m. A masked orderly squirts disinfecting soap on her hands, tells her to put her mask in the bin, and takes her sheet. It has two signatures. Then she is escorted out the door. She pauses . . . a moment. Her face is wet, and not with rain.

She can't understand why no one knew of her appointment (aren't there computers?), why they all looked at her that way (I'm just from Sarnia, aren't I?), and why she had to wait so long. That morning, HPC and cyberspace failed Susan. But they were going to help her father. She believes that now, and she thinks, *Dad's going to get better*.

Susan uses her cell phone to call Air Canada. It forgives her the missed flight, but has cancelled all later flights to Sarnia. The new normal sinks in. She walks south under a clearing sky down University Avenue to Union Station. The only VIA train leaves at 5:40 p.m. and arrives in Sarnia at 9:40 that evening. Susan withdraws some cash from one of her bank's 4400 ATMs, buys her VIA ticket, and calls Russ on her cell.

Susan's and Russ' voices are converted to pulses and then transmitted, thanks to a hidden HPC infrastructure, 160 miles in less than a heartbeat. "I'll be at the Green Street station to meet you," Russ says. She's unknowingly hitched a ride on a global HPC-powered bank network to buy her way home. Not even the new normal of SARS prevents the bank from recognizing Susan Zhong-Sheppard, even without an appointment.

Six weeks earlier, on April 12, the DNA sequence for the new SARS coronavirus had been published by the Genome Sciences Centre working with the British Columbia Centre for Disease Control and the National Microbiology Laboratory in Winnipeg. Two days before that, scientists at the Hospital for Sick Children in Toronto published the completely sequenced human chromosome 7, including landmarks for diseases like cystic fibrosis, leukemia, and autism. Both projects heavily used HPC resources.

Outside in downtown Toronto, on the afternoon of May 29, no one wears masks, and Torontonians go pretty much where they want to. Many carry day bags like Susan's. She visits a big bookstore and wanders through the Eaton Centre. No one asks her what her business is. No one looks at her bag with interest. This is Canada, and today is the International Day of United Nations Peacekeepers. Yet electronically, whenever Susan steps into cyberspace to do banking or charge purchases, her avatar grows, date-stamped with events. Two hundred years from today, archivists will know she visited Toronto, bought a CD in the underground city, saw Dr. Andres, made a phone call home, and bought a VIA rail ticket. They will also know she stops at an Internet café at 3:45 p.m. and writes Ceri:

I can't be home tonight to tuck you in bed Ceri but tomorrow I've got something for you ok so DO do your homework tonight oh are you getting along at school all right? hugs and kisses xxx mom

Twenty minutes later Ceri checks her email and picks up the note. Digitally at least, the "Susan" avatar might be, well, anyone because anyone can send e-mail that appears to be from her address. So yes, maybe an HPC security system is listening in, although why bother listening to unencrypted, non-image texts such as Ceri receives? Canadian firms like Entrust in Ottawa and Certicom in Waterloo sell encryption tools that are now fundamental to secure bank transactions and much else.

Ceri arrives home from school by bus at 4:15 p.m. Her health teacher asked her to explain to class tomorrow why her father doesn't let her drink raw milk from the cows. So Ceri asks her father. Russ explains that until milk is pasteurized, or quickly heated and cooled, it can be contaminated by bacteria from the milking. They send their milk to a plant to be pasteurized. Ceri asks, "What bacteria?" and Russ says, "It's called *E. Coli*. Check out the stack of *Scientific American* magazines or look online."

Ceri sits down at the big kitchen table and searches the Web for mentions of *E. Coli*. At a University of Alberta website, she finds Project CyberCell, which simulates an *E. Coli* bacterium as a virtual cell. By using HPC to discover how proteins interact in a virtual cell, scientists can help biopharmaceutical companies tailor-make drugs for individual patients.

Ceri calls out, "Dad, they look like hot dogs!" Ceri stares at the HPC-simulated cells, twisting in their strange world. In ten minutes, she learns that real *E. Coli* cells live in her own intestines.

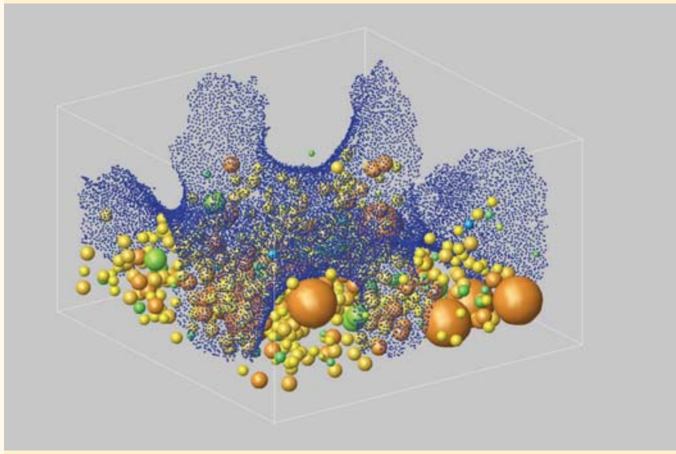
She calls out, "Dad, why can't I drink our milk if we already got *E. Coli in us?*" Russ patiently answers. "Ceri, it's a mutation of the normal bacteria that's bad. Search some more." And so she finds a *Nature* website image of *E. coli* 0157:H7 and its sinister flagella, and descriptions of bloody diarrhea and kidney failure. She gets up, glad that she doesn't drink raw milk.

Once on the train, Susan falls asleep, to dream of windswept springtime grasslands that make no one with asthma sick, of fields where cheering faces a lot like Ceri's throw their inhalers up into the antigen-rich Canadian rain.

When Susan steps off the train in Sarnia, Russ is there, smiling. Outside the station, they look up at a new moon and a shockingly starry sky. High in the south is Mars, bright and orange-coloured. She

Project Cybercell

Project Cybercell [REF 8] aims to model "the structures of almost all of the 4,000 E. coli proteins and then to simulate their interactions and functions in a computer," says Dr. Michael James, a member of the research team. James says once the E.coli proteins are "defined" and stored digitally, researchers can virtually interact with their digital versions and see how they would interact with antibiotic molecules. CyberCell, a huge cross-country project, will "take a long time to achieve – perhaps 20 years – but with more HPC resources this time could be significantly reduced."



Simulation of the dynamic interplay between a cell's interior and its environment through the plasma membrane.

OECD/OCDE

In Paris on January 30, 2004, 34 governments (including Canada, China, the European Community, the Russian Federation, and the United States) from the Organisation for Economic Co-operation and Development (OECD) agreed to establish ways to give the public open access to all publicly-funded digital research. [REF 9]

wondered if Chris Hadfield, who lent his name to the airport where Susan took off 16 hours before, was looking up too. He grew up in Sarnia and was the first Canadian ever to float freely in space. He also served on the crew of the Space Shuttle Endeavour that two years ago installed the Canadarm 2. Susan says to Russ, "Will we ever see families like ours on Mars, do you think? Will Ceri's children?"

And will there be cows on Mars too? Looking up, Russ wonders, *Are we the only part of this universe that is conscious of itself?* "Well, love, if they can simulate the first minutes of the life of the universe, getting to Mars can't be so hard, can it?"

Quietly, Russ and Susan drive home together. They find Ceri still up and waiting at her window, singing to herself:

"Star light, star bright,
First star I see tonight,
I wish I may, I wish I might,
Have the wish I wish tonight."
. . . and what *was* Ceri's wish that night?

Susan and Russ know, but only much later, long after a biographer has combed through their family papers, will others know . . . others like a mother younger than Susan, and a daughter taller than Ceri, who both will one day ask that same question.

“Selay,” says the mother yet unborn to her ten-year-old, “have you found her yet? Your homework’s due tomorrow, isn’t it?”

“The Library’s looking, Mom,” the girl answers, standing before a hovering plasma image. Her teacher had given her a strange question to answer: “What made a famous 21st-century woman from Ontario named Ceri Zhong-Sheppard want to become a scientist?” Selay turns to her constant friend, the online Library.

It shows her Ceri’s bio, a few pictures, and then a record from the online science archive. The screen dissolves into

2020.05.06:12:06 K.Z-S@IntSpaceStation2 to zhong-sheppard@sympatico.ca

Dad! . . . you can see our satellite in the NW sky about 23.30 your time. We’re in space monitoring Pacific carbon uptake and phytoplankton distribution now that the deep-ocean ‘mote field’s in place – those remotes feed us a quarter-petabyte a minute! What a far cry from our wee kitchen PC!
love Ceri

“Do you understand, Selay? Ceri was helping save our oceans.” But Selay asks, “What’s a wee kitchen PC, Mom?”

Her mother answers, “I think . . . yes . . . it was a little computer. It couldn’t listen, like the Library, but it knew things.” The plasma then resolves into an antique facsimile image of a real-paper printout.

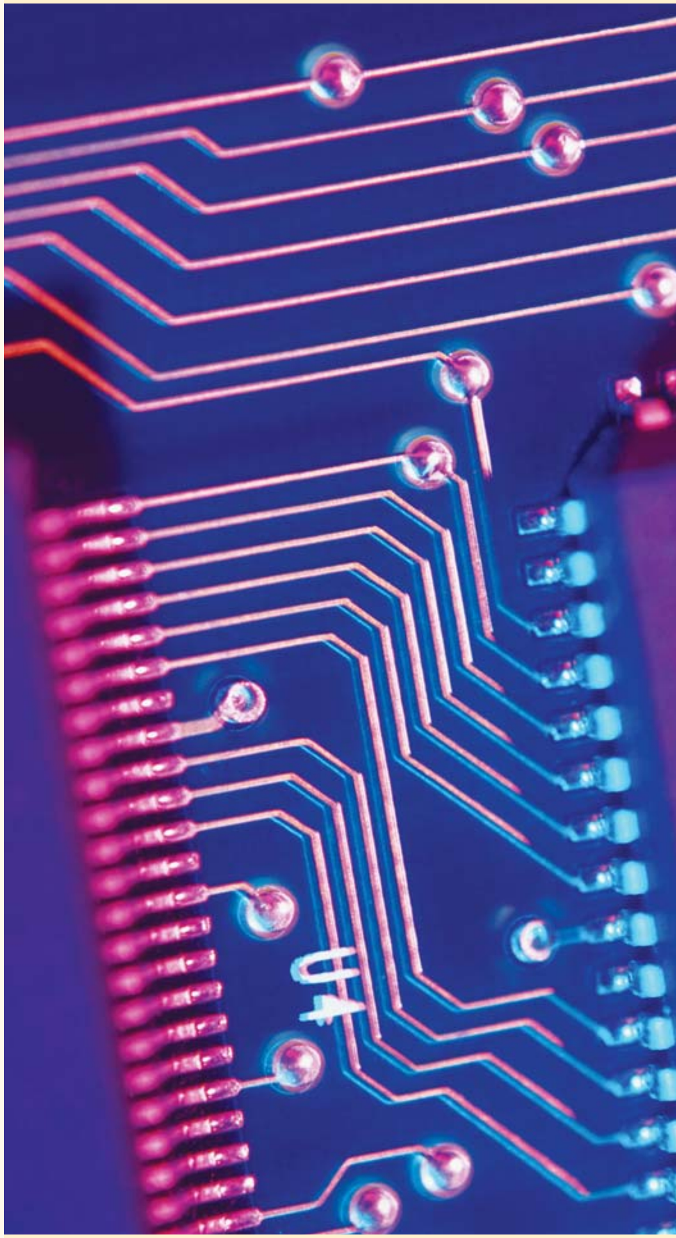
2003.05.29:15.51 zhong-sheppard@sympatico.ca to zhong-sheppard@sympatico.ca

mom are you there
I got PC pictures of ecoli that’s why we cant drink milk from the barn! can I make pictures like those when I grow up
xxx Ceri

Selay stands quietly a few moments before what Ceri’s mother, long ago, had secreted in her bedroom drawer.

“Mom, her PC at home . . . that made Ceri want to be a scientist!”

“Little one, wishes can come true.”



Executive Summary

Executive Summary

21st century computing involves a well-connected digital world where state-of-the-art computation enables research across all the sciences, including the health, environmental, applied, physical, life, social and human sciences. Canada needs sustained funding of computational infrastructure to build on the enormously successful investments made to date, to capitalize on Canadian researchers' proven ability to produce world-class research, and to take advantage of Canada's world-leading networking and telecommunications infrastructure.

Our vision is to create a sustained and internationally competitive high-impact, multi-institutional physical and human infrastructure for computationally based research that addresses the needs of industry and society for innovation, training and outreach.

This vision will be achieved through the sustained funding of high performance computing (HPC) infrastructure that builds upon the more than $250 million invested or committed over the past five years by the federal government, the provinces, the universities and industry. To increase Canada's competitiveness, and to gain new opportunities for research and development in support of our national economy and quality of life, the Long Range Plan Authors Panel makes the following overarching recommendations (where all recommendations are numbered according to the chapter in which they are presented and discussed, and hence the following recommendations appear in Chapter 1):

1.1 That a long-term funding envelope be developed (a) to provide sustained support for *mid-range advanced* computing facilities such as regional HPC consortia, and (b) to establish a pan-Canadian high-end computing facility that would have a sustained presence in the world's top 30 computing facilities.

1.2 That a national initiative be established to provide outreach, leadership, coordination, and oversight for HPC in Canada.

Impacts, Outcomes, Opportunities and Benefits of the Recommendations

Research has undergone a paradigm shift in which computational methods have assumed a pivotal role in driving the missions of the health, environmental, applied, physical, life and human sciences. In many fields, computation is now the primary driver of innovation. Our recommendations will ensure that Canadian research, industry, economic activity and society will, over a horizon of 15 years, be positioned to reap the internationally competitive benefits provided by HPC, benefits vital to our economic and social success.

There are no prizes or patents awarded for being second to solve a research problem. Our plan will have major impacts: it will establish a technology roadmap for computing-enabled research in Canada, it will provide the environment necessary for major new discoveries, and it will generate 21st century technological literacy via new research and training programmes. Having the best possible computing facilities and highly qualified personnel will significantly enhance research productivity, reduce time to manufacture and market, facilitate knowledge discovery and accelerate innovation.

Access to the best possible computing facilities leverages and maintains a pool of skilled personnel. Expertise is not an incremental benefit. These highly qualified people are the crucial resource needed to enable enhanced research and manufacturing productivity, accelerated discovery and industrial innovation. Canada's future opportunities are in the hands of these experts. Our country has many world-class researchers who now rely on HPC to make innovative discoveries at the international level. A sustained investment of public funds, growing from $76 million per year in 2006 to $97 million per year in 2012 and beyond, will allow Canada to remain competitive in this dramatically evolving technological future.

Competing in the Digital World

Recent progress in computing, information and communication technology has given rise to a revolution in computational research. With this technology, we can now do research on issues of national and global importance in new ways and with increased efficacy. Exemplary applications include understanding and protecting our natural environment, applying genomics, proteomics and medical imaging to human health, studying global climate change, developing nanotechnology, maintaining national security, finding cost-effective ways of dealing with the aging infrastructure of cities, and improving the efficiency of essential services, as well as predicting, protecting against, and recovering from natural and human disasters. We can also formulate innovative ways of answering such fudamental questions as how the universe was formed and what constitutes matter itself.

Talented Canadian researchers and technicians were able to understand the SARS virus quickly because they had access to state-of-the-art computational infrastructure.

(For more details, see sidebar on page 14)

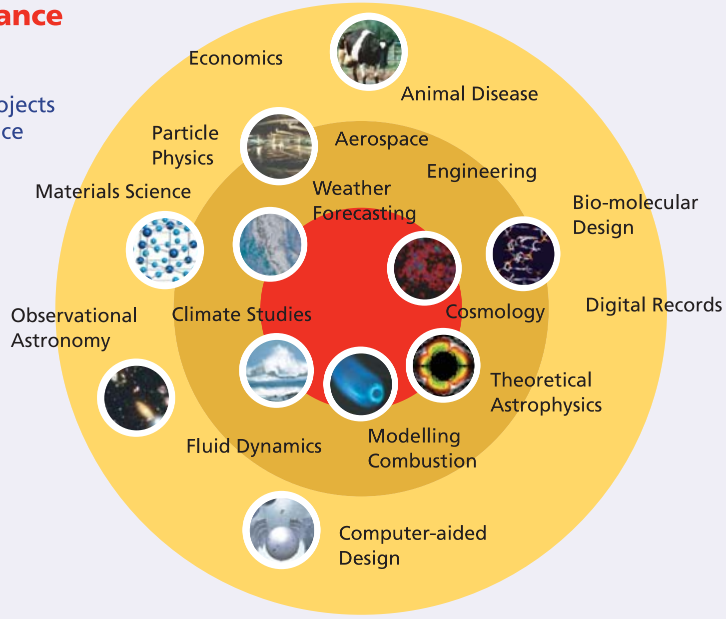
Canadian High Performance Computing Needs

The array of canadian research projects each have unique high performance computing requirements.

🟨 **Ring 1**
Desktop Computers / Small Clusters

🟧 **Ring 2**
Mid-Range Systems
(in the top 500 worldwide)

🟥 **Ring 3**
Supercomputers / Terascale System
(in the top 30 worldwide)



Canada can be a world leader in computationally based research in the 21st century, but this will require sustained, generation-long investments that enable it to compete on a global basis. In the United States, Europe and Asia, annual investments are orders of magnitude greater than those in Canada. For example, the U.S already spends about $2 billion (US) per year in support of HPC, and the recent passage by Congress of H.R. 4516, the Department of Energy High-End Computing Revitalization Act of 2004, will establish supercomputer user facilities that will provide U.S. researchers access to some of the world's most advanced computers. In addition, a Blue Ribbon Advisory Panel recently recommended that the U.S. National Science Foundation (NSF) invest a similar annual amount in Advanced Cyberinfrastructure – distributed computing, information technology, and communication technology. Canada has an enviable computer network (established by the federal and provincial governments through CANARIE Inc. and the associated regional networks), but it does not have a long-term strategy to support these networks and the HPC resources that they host. Without such a plan, known HPC initiatives in other countries pose a significant threat to Canada's future economic competitiveness.

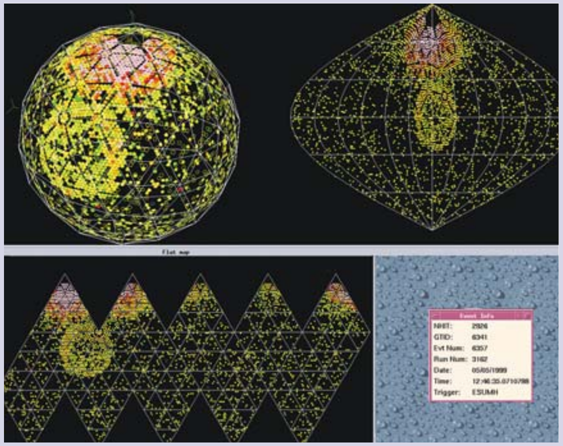
Partnerships between Canadian industry, universities and research institutes are growing, and will continue to grow as the investment in HPC hardware and expertise (people) continues. Industry must have the assurance that it can hire the talented people needed for both long and short term product development and that it will continue to benefit from a substantial pool of innovative research and researchers. To educate and retain these experts, Canada needs advanced HPC infrastructure.

HPC Needs

Computer-based simulations of our physical world open up new vistas. Traditional laboratory experimentation and the theoretical analysis of many complex problems in isolation are often no longer adequate. Computational modelling of very complex, multi-scale phenomena is becoming the key tool for answering difficult questions. For example, the design cycle and thus time-to-market of an aircraft engine or even a whole airplane can be reduced by HPC (see the sidebar on page 11). Computer-aided design is now considered essential to the development of the products of tomorrow.

In our Long Range Plan (LRP), we represent the range of computing capability as a series of concentric rings. The outermost ring represents the enormous aggregated capacity of *desktop machines* and *workstations* each capable of billions of calculations per second.

Sudbury Neutrino Observatory



Different projections of the Sudbury Neutrino Detector (SNO) photomultiplier array with two rings clearly present. This event is the result of a high energy neutrino produced from a cosmic ray interaction in the atmosphere on the opposite side of the earth. After passing through the earth, the neutrino interacted in SNO, scattering an electron, which can be seen as the less intense secondary ring. The colour and density are proportional to the energy of the event.

Where are the missing neutrinos? This question had for decades been puzzling scientists trying to understand the Universe. Canada gained considerable international recognition for leading the team to answer the question and hence better understand the world around us. Without the availability of advanced computing facilities to interpret data collected deep in a mine in Sudbury, important aspects of solar neutrino production would have remained an enigma. These and other visionary studies would not have been possible without the significant direct and indirect investments in HPC. (For more details, see case study on page 41)

> *"Canada has emerged as a world leader in the field of particle astrophysics. Several exciting new international projects which probe the structure of the universe and the properties of the particles within it are currently being launched. The characterization and analysis of data from these extremely large and complex detectors requires enormous computing resources. A high performance computing facility is essential for Canada to remain at the forefront of this rapidly developing field."*
>
> ***Dr. Tony Noble***
> ***Director, Sudbury Neutrino Observatory Institute***
> ***Canada Research Chair in Particle Astrophysics***
> ***Queen's University***

They provide essential "foundational" computing and visualization tools. The two inner rings represent smaller numbers of progressively more powerful computers: these are systems capable of running applications on much larger numbers of processors. The middle ring – *the mid-range computing facilities* – are those computers that have the power of many hundreds to thousands of workstations. At the centre are the *high-end* or *tera-scale computers* with the capability of tens of thousands of workstations.

Many general research problems can be solved with desktop computers. This plan recommends that the granting councils (NSERC, SSHRC and CIHR) continue to provide comprehensive funding for this outer ring. However, access to mid-range computing facilities is essential for an increasing number of key research areas as illustrated by the majority of the case studies contained in this Long Range Plan. These facilities are also enabling interdisciplinary research, such as the marriage between nanotechnology and biology, engineering and medicine, and materials science and chemistry. At the centre of the plan, access to high-end computers will allow Canadian scientists to tackle the "grand-challenge" problems, whose solutions will have a fundamental impact on the evolution of high-end technology and, by extension, on Canada's position as a world leader in health care, science and culture

We recognize two critical needs. The first and greatest need is for sustained, generational support for the mid-range computing facilities so that Canada can derive maximum benefit from the funds already invested. This will address the needs of most of the HPC-dependent research community.

A pivotal component of sustaining credible HPC capability at the mid-range and higher is the requirement that we develop and sustain the pool of HPC support personnel. HPC researchers rely on the expertise of the talented individuals who operate the systems and provide computational support. There is neither a current sustainable funding mechanism nor an established process for developing this human infrastructure. Retention of the research base that has been increasingly attracted to our Canadian university/hospital complexes is directly related to our HPC support personnel capacity. Current international practice – which is not yet followed by Canada – is to allocate about 25% of total expenditures to HPC support personnel. Canada must move towards this level of investment if it is to maximize the capabilities or potential of its computational scientists, its HPC infrastructure and its overall research capacity.

The second critical need is for a sustained pan-Canadian high-end computational facility that has a sustained presence in the world's top 30 most powerful computing facilities. This facility must also include the crucial corresponding support personnel to maximize the research potential of the computational hardware. We anticipate that an expert national procurement group, which represents the needs of the various scientific communities, will determine the acquisition, evolution, location (which may change over time) and precise form of this high-end facility. This group would likely be composed of members of key funding agencies and client groups.

These needs lead us to the following recommendations (in Chapters 2 and 5):

2.1 In addition to the continued funding of foundational research computing infrastructure (desktop, workstation and small system) by the granting councils, we strongly recommend that the research computing infrastructure in Canada moves towards a model that supports regional mid-range computing facilities and a world-class high-end facility. By 2012, this will require the following (in order of priority and including in each case capital infrastructure, human infrastructure and operational costs):

   a. $76 million/year in sustained funding and ongoing support for mid-range computing facilities. This support will (i) build on the present HPC regional consortium model (where universities in a region cooperate to provide shared multi-purpose HPC resources, as well as the human infrastructure to operate the system and to train and support the researchers), (ii) support major special-purpose in-house computational facilities for exceptional cases where it is impossible or inefficient to use a shared multi-purpose facility, and (iii) promote collaboration and create synergies that extend beyond computation.

b. $19 million/year to establish and sustain a multi-purpose pan-Canadian high-end computing facility and to fund appropriate support personnel. We anticipate only one such facility in Canada.

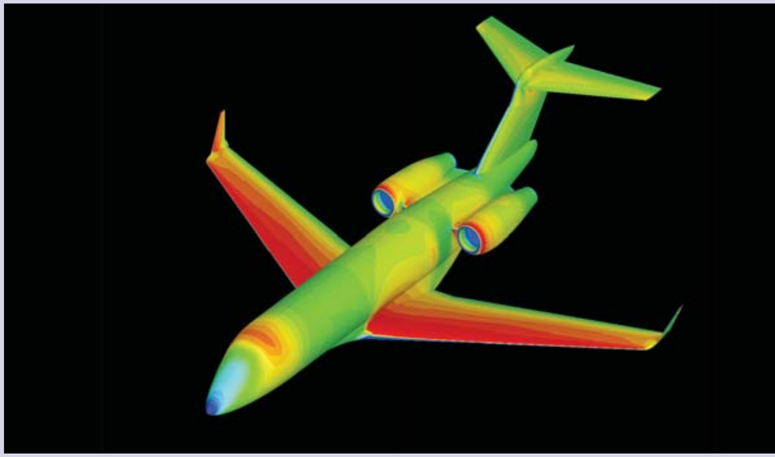
In addition $2 million per year will be needed to fund IMPACT Canada to be discussed below.

Recognising that computationally based research and training are essential, implicit, components to the long-range strategic research plans of most universities in Canada, we have several strong recommendations relating to human infrastructure (in Chapter 3):

3.1 That the granting councils, in consultation with the universities, develop funding opportunities which strengthen research programs both in the foundational technologies of HPC and in HPC application development;

Bombardier

"Engineers designing and developing a high performance aircraft are faced with several competing issues, such as how to achieve optimal cruise performance, how to design flaps and slats for optimum performance at take-off and landing, how to optimize the wing's structural integrity and how to best install engines for minimum interference. At Bombardier, CFD aerodynamic computations are significantly enhanced with access to HPC and provide reliable answers to many of these questions. The judicious use of CFD methods, ranging from panel methods to 3D Navier-Stokes solvers for complete aircraft configurations, allows the aerodynamicist to cost-effectively reach optimal configurations, of highest safety. This more thorough aircraft design optimization focuses wind tunnel testing and considerably reduces the costly and time-consuming number of experimental iterations. At Bombardier, we firmly believe our competitive edge is critically dependent upon use of, and access to, HPC. HPC was crucial to design our first Regional jet, and it will continue keeping us flying."



Computational fluid dynamics (CFD) simulation used in the design of Bombardier Challenger 300 at Mach 0.70 [REF 10]

Fassi Kafyeke
Manager, Advanced Aerodynamics
Bombardier Aerospace, Montreal

3.2 That universities develop training opportunities at the fundamental and advanced levels to enhance HPC skills in the research and development environment; and

3.3 That an amount equivalent to at least 25% of the annualized HPC hardware expenditures be allocated annually to fund HPC support personnel (adopted when establishing the funds required for recommendations 2.1 and 5.1).

Present Organization for Delivering HPC

There is currently no single organization responsible for delivering HPC, although multiple agencies are investing in its development, the recruitment of its researchers, and in the training of its personnel. It is essential that one coordinating mechanism be identified: HPC expenditures are enormous and cost-efficient HPC development will require the development of effective national strategies.

The result of existing investments has been six geographically distributed regional HPC consortia (WestGrid in western Canada, SHARCNET and HPCVL in Ontario, CLUMEQ and RQCHP in Quebec, and ACEnet in the Atlantic Provinces). These facilities knit together researchers, technical analysts and computational infrastructure at more than thirty universities. In addition to these consortia, some university/hospital complexes (such as the University of Toronto and its associated hospitals) and single universities (such as the University of Victoria) have developed significant in-house HPC facilities.

Centralized HPC resources, in the form of general-purpose consortia or domain-specific HPC facilities, have grown across Canada because the benefits of large shared resources are greater than those of several smaller, single-user dedicated systems. The consortium installations provide a wide range of researchers from across Canada with access to a range of different computing platforms, since each facility reserves on average up to 20% of its resources for suitable research projects by scientists from other regions. Domain-specific facilities are appropriate where large communities of scientists within a particular discipline require extensive HPC support for common computational problems. In either setting, the use of a shared facility allows increased aggregate throughput and can satisfy a larger user community, while still meeting the needs of the individual user. The sharing model provides clear economies of scale in personnel, space utilization, power consumption, back-up support, system administration, technical support, system up-time, software licensing, and hardware support. This leads us to recommend the following (in Chapter 5):

5.2 We recommend that funding requests for HPC computer resources be evaluated in the context of what can be provided by existing HPC facilities, either general-purpose consortia or domain-specific facilities, so that the most cost-effective use can be made of investments in HPC. Requests that genuinely cannot be reasonably satisfied using this existing infrastructure should compete on their merits for existing sources of funds such as conventional CFI envelopes or equipment funds from the granting councils.

The physical network for an umbrella HPC organization already exists. High-speed, high-bandwidth networking permits researchers to share large volumes of data, creating "communities of practice" among geographically dispersed experts. They can access resources and data thousands of kilometres away as if they were located next door. This is the basis for what is known as "grid-based computing". Its impact on the scientific community is comparable to that of the Internet on society at large. GRID Canada, which is a partnership between CANARIE Inc., the National Research Council of Canada and the C3.ca Association Inc., is leading the effort to move Canada in this direction. These advanced networks provide the essential connectivity between the HPC computing rings, but they do not have long term funding (for example, the funding for CANARIE lasts only until 2007).

This leads us to the following recommendation (in Chapter 2):

2.2 We recommend the continued support of CANARIE Inc., the optical regional advanced networks, and Canadian grid initiatives, and that these networks be maintained and enhanced in concert with new HPC investments.

Proposed National Initiative: IMPACT Canada

In recognition of the need to coordinate HPC facilities across Canada, to advise Canadian research funding agencies on HPC issues, and to develop and implement outreach and training programs, we make the following recommendation (in Chapter 4):

4.1 We strongly recommend the establishment of a body, IMPACT Canada, to provide national leadership, coordination, and oversight for HPC in Canada.

The name IMPACT Canada captures the essential elements of a well-connected, highly competitive digital world: **I**nnovation, **M**anagement, **P**eople, **A**pplications, **C**ommunication and **T**echnologies. IMPACT Canada will foster innovation. By providing management and coordination, it will enable people to reach their research potential. It will also assist in the development of the computer applications that will lead to scientific breakthroughs; it will foster the movement of ideas and interaction through broad communication strategies, and it will oversee and coordinate the acquisition and operation of the enabling technologies.

This initiative will provide an ongoing assessment of the rapidly changing field of HPC to identify emerging areas of strategic importance to Canada. It will create strategies for enhancing HPC research in the health, environmental, applied, physical, life, social and human sciences. It will also coordinate existing HPC resources across the country and help allocate new HPC resources to maximize the competitiveness of Canadian research. It will have full-time staff to facilitate program development, training, and co-ordination of symposia. This team will also provide advice and specialized assistance to HPC centres across the country.

It is envisaged that IMPACT Canada will play a key role in outreach and training. For example, it will work closely with universities, hospitals, colleges and consortia to deliver an early introduction to computational science to young Canadians via the World Wide Web. This medium is now recognized as a vital tool for promoting awareness of science. The innovative outreach programs of IMPACT Canada will be designed to demonstrate to our next generation of young researchers the central role that high performance computing can play in both learning and research.

An excellent example of a spin-off that would not have been possible without HPC is the Montreal-based biotechnology company Neurochem Inc., which grew out of research conducted at Queen's University. The company is publicly listed (TSX: NRM) and now has a market capitalization of over $1 billion.

(For more details, see case study on page 29).

IMPACT Canada will also develop training strategies in HPC foundational technologies (such as data-mining tools, meta-computing, performance evaluation tools, advanced visualization and automatic parallelization) to help balance Canada's acknowledged strengths in computationally based research.

The Prime Minister's response to the 2004 Throne Speech called for a Canada at the leading edge of the 21st century economy. We offer an HPC strategy that can make Canada ***the*** country that most fully realizes the potential of HPC. Initial HPC investments within our universities and hospitals have sown the seeds for new and exciting economic gains for Canada. Momentum has been created over the last several years by the establishment of excellent mid-range computing infrastructure. Much more action will follow once Canadian small- and medium-sized enterprises (SMEs) gain access to the resources that will be available through IMPACT Canada. While SMEs are recognized as Canada's economic engine, they do not have the resources to invest in HPC based pre-competitive research. Funding of IMPACT Canada will be one step towards closing this gap.

We envisage IMPACT Canada working with groups such as the National Research Council (NRC) / Industrial Research Assistance Programme (IRAP) to facilitate access of the SME base in Canada to HPC, to highly qualified personnel, and to advanced research outcomes and equipment. This includes linking SMEs with researchers and research opportunities.

Funding

Sustained funding will be a prerequisite for meeting the burgeoning needs of current and future computationally based research, and for ensuring that the present Canadian investment is fully exploited. As a result of the past four CFI competitions (1999, 2000, 2002 and 2004), the public (CFI and Provinces) has invested over $250 million in computational infrastructure and this has been supplemented with funding from industry and the universities. The total capital investment currently averages more than $40 million per year. NSERC now provides over $1 million annually in operating funds for these facilities. CFI currently provides a portion of the operating funds for *some* of these systems (about $6 million per year), but this funding may not be available for CFI competitions after 2004. There is no other system in place to maintain the funding of the existing infrastructure after CFI operating support runs out.

Unlike most other types of infrastructure, HPC becomes outdated very quickly: Moore's Law (see the sidebar on page 32) dictates that HPC equipment be upgraded about every three years if Canadian facilities and research are to be sustained at an internationally competitive level. To sustain the current infrastructure, our community must submit multiple applications to multiple agencies with multiple reporting and assessment mechanisms (and the timings of these are highly variable). The uncertainties engendered in the current Canadian funding environment – in contrast to the firm commitments to HPC made by our major international competitors – place Canadian research and development at a significant competitive disadvantage; this results in failures to capitalize on many opportunities.

Nonetheless, CFI made significant investments in HPC in 2004 that will roll out in 2005, and support is in place for a period of time through CFI; therefore,

Year	**2006**	**2009**	**2012**
	(Funding in $ millions[1])		
HPC capital infrastructure:	44+13*	49+14.5*	54+16*
Consortia			
High-end (tera-scale) facility	10+3*	12+3.5*	14+4*
HPC operations:			
Human infrastructure	13+4*	15+4.5*	17+5*
Facilities	8	9	10
IMPACT Canada	1	2	2
Total public contribution:	**76**	**87**	**97**
Total industrial contribution*	20	22.5	25

[1] In 2004$ * Industrial contribution

this plan sees the need for a phased build-up of funding to maintain the consortia and initiate the development of a high-end computing facility that will have a sustained presence in the world's top 30 most powerful computing facilities. We recognize an initial need for $76 million per annum from public funding in 2006 and this will grow to $97 million in 2012. It is anticipated that there will be an additional contribution from industry ranging from $20 million in 2006 to $25 million in 2012. The total annual budget (2012) required to provide and to maintain a nationwide world-class HPC infrastructure will be $97 million from public sources.

This leads us to the following recommendation for funding (in Chapter 5):

5.1 We strongly recommend that a long-term funding envelope be developed to support HPC in Canada. We anticipate that this funding could be phased in over a period of six years as follows:

Funding beyond 2012 will be based on future planning. The numbers are difficult to predict considering the changes in technology costs, increases in human resource costs, and other factors outside of our control or not yet known.

Based on this steady-state annual budget requiring public funding by 2012, of $97 million required annually:

- $76 million annually is for ongoing support for mid-range advanced HPC, including capital infrastructure ($54 million), human infrastructure ($14 million), and facility costs ($8 million);

- $19 million annually is for the high-end facility ($14 million for capital, $3 million for human resources and $2 million for facility costs); and

- $2 million annually is for IMPACT Canada (including people and operating expenses).

This budget presupposes that the granting councils continue to fund research projects made possible by access to the HPC infrastructure.

This long-term funding initiative will need to be re-evaluated every five years, perhaps as part of a national review of Canada's international research competitiveness. We suggest that the following performance evaluation criteria should be reviewed:

- Academic excellence (publications, awards, retention and recruitment of faculty);

- Qualified personnel produced (graduate students, postdoctoral fellows, technicians and research associates);

- Societal/economic impacts (patents, spin-offs, industrial partnerships, technology transfers, improved infrastructure, health outcomes, outreach activities); and

- Effect of the investments in HPC on Canada's international competitiveness.

For Canada to remain competitive with the rest of the world, HPC demands ongoing investment. Today's computational infrastructure supports an increasing number of researchers (almost 2000 in 2004) who are sowing the seeds for explosive growth in future computational needs. A stable HPC environment supports the activities of many of Canada's key technology clusters as well as of all the federal National Centres of Excellence (NCEs). These activities will continue to grow.

As well, the next large generation of computationally literate scientists is in the second or third year of their undergraduate studies. Their entrepreneurial and intellectual aspirations are being forged in an environment of expanding infrastructure, and their eventual research infrastructure needs will be well beyond the current HPC environment. On average, a student currently in an undergraduate program will require 15 years to become established in one of our many research professions. Their expectations are that the requisite infrastructure will be available to them as they move on to graduate degrees and then establish themselves between now and 2020.

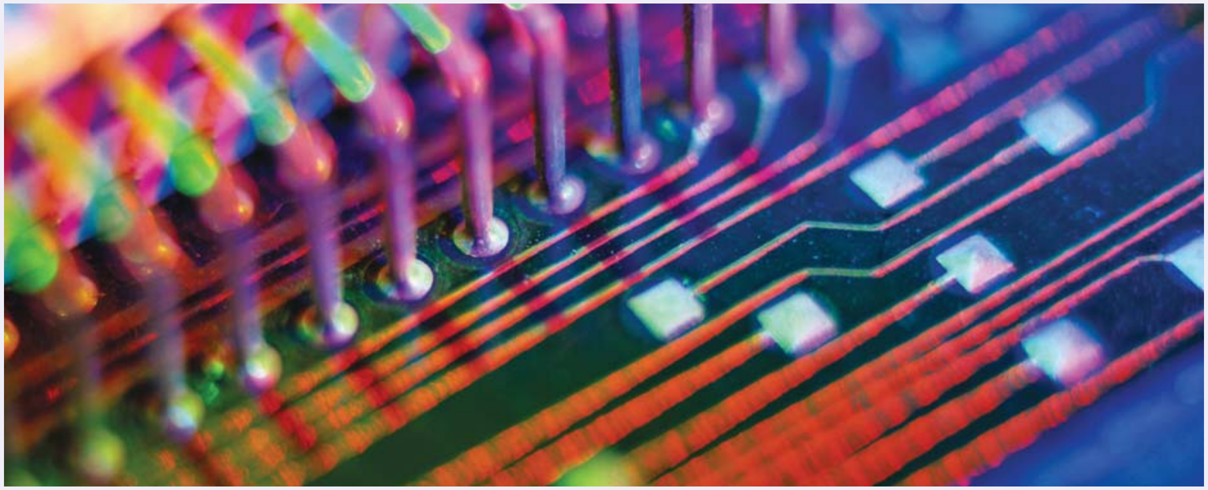
It often takes a decade or more to build strong research teams and to have their technology evolve from ideas to commercial products. It takes at least as long to train the next generation of people with advanced research capabilities. The creative sparks ignited in today's graduate students using the current CFI-funded infrastructure will not realize full potential for Canada unless these students have access to appropriate resources throughout their studies and on into their academic, public sector, or industrial careers. Failure to invest now will seriously damage Canada's innovation agenda. Long-term support for HPC will allow Canada to derive maximum benefit from its current funding initiatives.

At present, the CFI is structured to fund new infrastructure for innovative research. The CFI has played a pivotal role in establishing Canada's strong position in HPC. However, the CFI's present mandate and future funding to 2010 is not well adapted to ensuring the long-term viability of major multi-institutional infrastructure such as the regional HPC consortia. Once infrastructure is established, however, it is vitally important that – subject to review – it is kept up-to-date and that the researchers using the infrastructure are supported on a predictable long-term basis.

There are various ways that this new program could be implemented using existing mechanisms, including designated new funds within the CFI or through a special tri-council initiative (as for the Canada Research Chairs directorate). Alternatively, it could be implemented through a separate funding mechanism employing the review criteria outlined above. The implementation of the Long Range Plan will require leadership and substantial financing from the federal government. However, it would be desirable for provincial governments also to participate in this strategy. This would allow them to augment the national HPC infrastructure in ways that address specific provincial HPC priorities. Addressing these individual perspectives was beyond the scope of this plan and will require discussions between the two levels of government.

Conclusion

Over the past 5 years, Canada has established a strong international position with respect to mid-range HPC facilities. This has attracted many outstanding researchers to Canadian universities and helped Canadian industry to grow. However, if we are to keep these people and gain the full benefits of the investment, this HPC infrastructure must be sustained at a competitive level. If we are also to address the grand-challenge problems, it will also be necessary to establish a high-end computing facility. At present, Canada currently ranks 14th in the world in terms of HPC capacity relative to GDP (Chapter 5); below all countries with whom we regularly compare ourselves. The adoption of the recommendations in this plan would move us to 6th place (between Germany and Mexico). With this investment and these facilities, Canadian researchers can be leaders in discovery and innovation, and Canadian industry can be internationally competitive in a new and lucrative range of fields.



Neurochem and a Treatment for Alzheimer's Disease

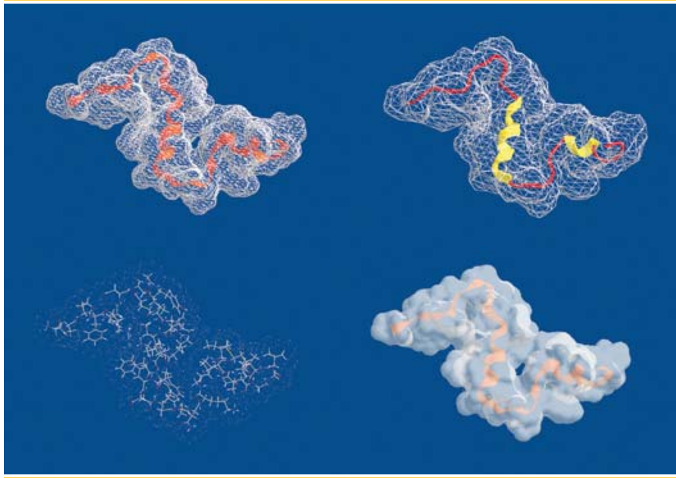
By supporting spin-off companies, Canada is not only supporting the sustainability and self-sufficiency of leading-edge Canadian research, it is investing in home-grown economic success.

Neurochem has been heavily involved with HPC developments, and was initially spun out of Queen's University. Having grown into a mid-sized Canadian drug company with about 100 employees, it is now a Canadian success story. HPC played a leading role in this success, acting as the central tool during the commercialization of university-developed technologies.

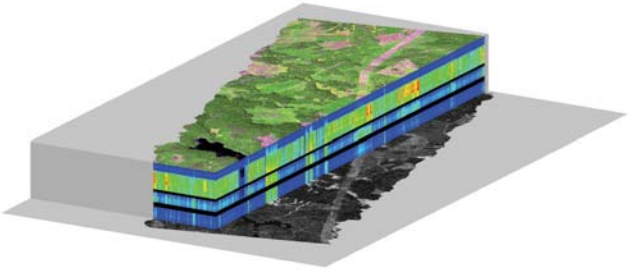
The pharmaceutical sector is an important component of the economies of developed countries. As the world's population increases and confronts the ever-expanding health problems of the modern world, the discovery of new chemical entities (NCEs) as therapeutics for human disease will become one of the major achievements of the 21st century.

The ability of HPC resources to produce and analyze data in a timely manner helps to push research forward. There are numerous examples of the importance of HPC to drug discovery. One critical discovery used calculations within an HPC environment to enable the design of novel molecules for binding to beta-amyloid, a peptide involved in Alzheimer's disease. Scientists estimate that up to four and a half million people in North America currently suffer from Alzheimer's disease. This Canadian discovery may lead the way to prevention of the disease, which currently has an incidence rate of approximately 360,000 new cases each year. The resulting compounds have now entered phase III human clinical trials in Canada, with a potential cure only a few years away.

> *"Canada needs more Neurochem success stories. The availability of HPC will be important to this goal. Moreover, the pay-offs will be just as great: an effective drug for Alzheimer's disease or a single new antibiotic drug with widespread usefulness will be a "billion-dollar molecule"."*
>
> ***Dr. Donald Weaver***
> ***Canada Research Chair in Clinical Neuroscience***
> ***Professor of Chemistry***
> ***Dalhousie University***



One of the leading approaches to drug design for Alzheimer's disease is to engineer molecules that can bind to the beta-amyloid peptide (shown here). Since this peptide has never been effectively crystallized, such structural studies are totally dependent upon molecular modelling and simulation studies within a high performance computing environment. These computer-aided studies are crucial in enabling the design of drug molecules that interact selectively with discrete disease-causing molecules.



The data cube above was created from a hyperspectral image, which was taken by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) in 2002 over the Greater Victoria Watershed District. The image is 1988 pixels long by 1331 pixels wide, each pixel covering 4 metres on the ground.

Hyperspectral images typically contain hundreds of narrow spectral bands, whereas digital camera images only contain three: Red, Green and Blue. Each band covers 10 nanometres in the electromagnetic spectrum. This image contains 204 spectral bands in the visible and infrared wavelengths. Each layer in the cube cut-out is one band.

The top layer shows a false colour composite, which consists of two bands in the infrared and one band in the visible wavelengths. Pink represents areas with no vegetation, light green represents areas with young forest, dark green represents areas with more mature forest and black represents areas containing water.

The greyscale part of the cube shows the image at a single band, 2400 nm, the largest wavelength detected by AVIRIS.

Maintaining Canada's Environmental Beauty

Canada has over 418 million hectares of forests, representing 10% of the forested land in the world! With this 10%, Canada is a global leader in some of the areas of forestry research such as remote sensing research in forestry and forest monitoring and carbon accounting.

The Canadian Forest Service (CFS), a branch of Natural Resources Canada (NRCan), is a leader in remote sensing research related to forestry and forest monitoring. CFS use large volumes of remotely sensed data to create products for forest inventory, forest carbon accounting, sustainable development monitoring, and landscape management. With a remote imaging library of over 2200 images the size of remote sensing data is getting larger and larger. In order to provide high spectral and spatial resolution imagery for accurate analysis HPC is critical to process and transport these images from remote sites nationally to be housed in a fully accessible repository. For example, a typical AVIRIS image can be about 4 Gigabytes. Multiply this by 2200 times and you can begin to appreciate the significant impact that HPC and high-speed networks have on processing these high quality remote sensory images.

CFS not only relies on high performance computing to gather, process and store remote sensory images; it needs HPC to facilitate its collaborative research and product distribution across Canada and allow the public to access the research results in a timely fashion. Without HPC the Pan-Canadian network of forest research centres would not be able to process remotely sensed data collected from the field every day. HPC also significantly reduces the computational time required for time-consuming applications and modeling.

The impact of this research not only plays a pivotal role in the timely analysis of results for forest management, forest inventory, forest industry, and public information, it provides next generation forest measuring and monitoring systems that respond to key issues related to climate change and to report upon sustainable forest development of Canada's forests both nationally and internationally.

> *"Economically, HPC not only facilitates this new method of research in remote sensing and forestry, it provides a return on investment that could generate a 1% improvement in forest product sales, which would amount to a benefit of $700 million annually in Canada alone."*
>
> ***Dr. David Goodenough***
> ***Chief Research Scientist***
> ***Natural Resources Canada***
> ***Canadian Forest Service***
> ***Pacific Forestry Centre***

Chapter 1: Introduction

The pace of scientific advance is accelerating. In the past decade, research areas such as nanotechnology, bioinformatics and googling have come to the scientific forefront. Their very names are trendy buzzwords only recently introduced to the language. These and other exciting new scientific frontiers offer the promise of new knowledge and technological breakthroughs that will support improvements in our economic and social well-being. When combined with new perspectives in the basic sciences – including changing views on the origins of the universe, the secrets of DNA and the foundations of chemical reactions – we have an unprecedented scientific revolution, more profound than the changes that fuelled the first industrial revolution. The pace of innovation is accelerating, and the race is intensifying to be the first to reap the benefits.

At the heart of this revolution is the computer. The computer is essentially a cost-effective shared research laboratory. It can be used to simulate the dawn of time, to calculate the expected behaviour of a new aeroplane wing, to sift through billions of pieces of biological data, and to help design a machine that is one-millionth of a meter across. Every area of research, including the sciences, engineering, medicine, social sciences, and humanities, has been profoundly affected by computing resources. As computing capabilities increase – computers get faster, memories and disks get larger, and network capacity grows – we are tackling problems that seemed impossible only a few years ago. Our intellectual curiosity also pushes this envelope continually further, and we can truly say that

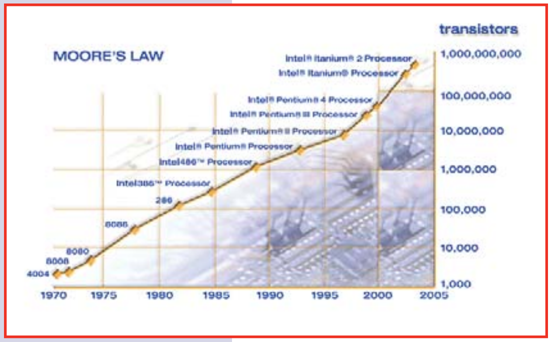
" . . . a new age has dawned in scientific, medical and engineering research, pushed by continuing progress in computing, information and communication technology, and pulled by the expanding complexity, scope and scale of today's challenges." *[REF 11]*

While the computational power of a state-of-the-art desktop computer continues to grow at a rapid pace (see the sidebar on page 32), for some applications these advances are not enough to do leading-edge research. For the British Columbia Cancer Agency to be first to sequence the SARS virus, 400 computers worked full-time to analyze the biological data. Even then, first place was won only by a photo finish (and there was no prize for second place) [REF 12]. Canada's Bombardier Inc. and the Canadian aerospace industry in general are major success stories. Massive computing resources are used to simulate the next generation of airplanes. This work previously required the construction and testing of expensive physical infrastructure, including massive wind tunnels, but can now be largely done in the confines of a computer. For Environment Canada to produce real-time weather forecasting, it needs a computer that has hundreds of times the capability of a desktop machine (a 960-processor IBM pSeries 690 that ranks 74 on the November 2004 world TOP-500 list). The Hospital for Sick Children in Toronto is using a computer with over 200 processors to analyze hundreds of billions of pieces of genomic data in its search for the genetic causes of diseases. Large Canadian scientific investments, such as the Sudbury Neutrino Observatory, the Canada Light Source, and TRIUMF all have massive high-performance computing (HPC) needs – far beyond that available in the ordinary desktop. As we move beyond familiar theoretical and experimental research methodologies, and delve into computational research – where we address problems that are too big, too expensive, too dangerous, or just impossible to do using any other technology – HPC becomes a cornerstone for progress.

The advent of the computer, and in particular HPC, has fundamentally changed the way research is done. To be competitive in the 21st century, researchers and their research and development partners must use the tools of the 21st century, and state-of-the-art high performance computational resources are now essential to the advancement of research across the applied, physical, and human

Moore's Law: Computer speed doubles every 18 months

In 1965, Gordon Moore (one of the co-founders of Intel) observed that the number of transistors (switches) that could be put on a computer chip was increasing at a constant rate. Since there is a correlation between the number of transistors on a chip and processor speeds, Moore's "Law" has been extrapolated to be suggestive of computer performance. The modern version of Moore's Law is that computer speed doubles every 18 months. Amazingly, Moore's prediction has held for almost 40 years, and the pace of technological innovation shows no signs of slowing down. For example, a year-long computation done on a personal computer in 1980 would take roughly 5 seconds to do using a 2004 workstation!



The implications of Moore's Law for computational research are challenging. A computer bought today will be half the speed of one purchased a mere 18 months from now. For many areas of research, giving up a factor of two in computational power to the competition is a serious disadvantage. There are few prizes awarded for being second to achieve a new result. Hence there is a continual need to buy the latest technology – a situation that is unprecedented in the history of scientific research. [REF 13]

sciences. Businesses and governments also require access to this infrastructure to implement new research knowledge and to take advantage of new opportunities in sectors as diverse as health, environment, energy, transportation, tourism, and manufacturing. Sobeys Inc. uses a 512-processor computer to meet its business needs. Oil companies in Calgary use computing clusters, some with tens of thousands of processors, to do seismic analysis to help find the best places to look for oil. Many other businesses are dependent on extensive computing resources to ensure reliable daily operations in industries such as banking, telecommunications, security, and internet searching. These computing resources are hidden behind the scenes, a silent partner in our economy, and few customers are aware of the enormous computing infrastructure that supports seamless day-to-day operations.

High performance computing infrastructure represents resources far beyond that of the ordinary desktop machine, and is the key technology fuelling today's knowledge revolution. This research infrastructure will be essential if Canada is to make key scientific discoveries and to reap the economic and life style benefits that accrue from innovation. Access to this technology and the benefits derived from it will be crucial to our competitiveness as a nation. We stand at a pivotal point in history, where we can become either the originators and vendors of the tools, services and benefits that flow from working with HPC resources, or the buyers of them.

The C3.ca Association (www.c3.ca) is Canada's national voice for high performance computing. Our vision is to build an internationally competitive, high-impact, multi-institutional platform for computational research that addresses the innovation, education, training and outreach needs of industry and society as a whole. Through a major two-year effort, the Association has formulated a national strategy for creating a sustained and sustainable research computing environment in Canada. Within this context, the Long Range Plan Authors Panel conducted town-hall meetings across Canada, sought input from all the key user sectors of HPC (including relevant industrial users) and reviewed international trends. This report is the culmination of these efforts and presents our conclusions and recommendations.

What is High Performance Computing?

As a fundamental tool in modern science, medicine, engineering and business, the computer has no equal. A physical microscope has a finite range of magnification, but a computer can mathematically magnify a problem with near-arbitrary scale.

Machines and instruments have to be retooled or reconfigured for different purposes, but a computer may be quickly loaded with new software for a new computation. Humans can skillfully consider many different economic scenarios, but a computer can exhaustively enumerate all interesting scenarios and then generate a graph to summarize the results. The computer extends the scientist's senses, mind, and models to consider problems that are otherwise impractical to solve. A remarkable property of computers is their ability to be used to study problems on a variety of scales: time can range from a billionth of a second to an eon, and size can be that of an atom or of a solar system. A computer model is a powerful and flexible tool that can simulate physical phenomenon, reducing or eliminating the need to build costly physical models such as wind tunnels. The computer is as important to biologists as it is to geotechnical engineers, physicists, and medical researchers; it is a cost-effective shared research tool that can be a virtual laboratory for almost every area of scientific exploration.

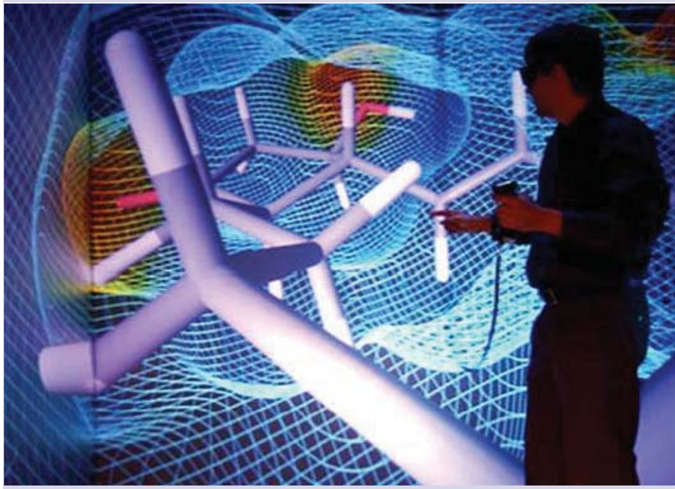
Researchers and developers, whether in academia or industry, typically test their theories using diverse sets of input parameters, including those created as a result of a previous computation. The longer the scientist has to wait for an output, the slower the process of working through various "what if" scenarios and consequently, the longer the wait before new knowledge is attained. Research productivity, time to manufacture and market, information acquisition, knowledge discovery and innovation can all be significantly enhanced through HPC.

Loosely speaking, HPC can be defined as a computing infrastructure that is at least 100 times more powerful than a state-of-the-art desktop machine and might be, by today's standard, tens of thousands of times more powerful. Powerful is the operative word. A computation that might take a year on a desktop machine is reduced to a lunch-break calculation on a world-leading HPC machine. The time required to reach an answer is often critical: it is as important to exhaustively testing a theory as it is to bringing new products to global markets. Thus, HPC has an important role to play in both scientific achievement and economic competitiveness.

Although one usually equates HPC merely with powerful computers, in reality "supercomputing" is only a part of the overall package. HPC includes the entire continuum from data acquisition to manipulation, storage, analysis and output. At one end of the research pipeline, data acquisition devices gather the information to be processed. These data can come from diverse sources such as seismic sensors, weather balloons, telescope observations, financial transaction databases and logs of World Wide Web activities. The data is prepared, stored and then fed into the HPC computers for analysis. The resulting outputs may be as simple as

"The human body is undoubtedly the most complex machine ever created. Genome researchers are undertaking the challenging task of unravelling how it is organized and how it functions. High performance computing plays a dominant role in this research. Without extremely sophisticated computational infrastructure, genomics research would not be possible."

Christoph Sensen
Director, Visual Genomics Institute
University of Calgary



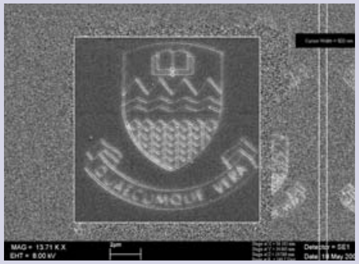
Christoph Sensen interacting with a 3-dimensional misoprostol molecule in the University of Calgary Visual Genomics Center CAVE.

Nanotechnology

Manipulating matter at the nanometre level (one billionth of a metre) promises to create a new industrial revolution. Canada is working towards becoming a leader in nanotechnology with initiatives like the recently created National Institute for Nanotechnology (NINT).

Scientists are rearranging individual atoms to build new materials, devices and structures that have unmatched, unique and invaluable qualities such as high strength and durability.

Physical experiments at this level of precision are difficult and expensive. Scientists study the physics of particle interactions using computer models that can identify which laboratory experiments would be a waste of time and resources and which ones would likely be successful. This new strategy is being adopted in many research areas.



The University of Alberta coat of arms on the end of a pin (10,000-fold magnification). The image was produced using electron beam lithography.

a Boolean verdict (a true or false answer) or as complex as a high-precision multi-dimensional image. Productivity benefits enormously from having access to a variety of forms of output to assist users in further interpreting and presenting their data, such as databases, screen displays, three-dimensional immersive environments (a CAVE: see the sidebar on page 33) and three-dimensional printers.

Although computers can be used to answer questions about a large range of problems, it is in the area of simulation that high performance computing is most influential. In fact, many leading-edge research questions can only be answered using correspondingly leading-edge computing technology to simulate the relationships between phenomena, with complex mathematical models providing the means to discover new information. For example, the development of a realistic model of the impact of global warming likely should integrate a large variety of interactions between the atmosphere, land and ocean. Groups of experts within the supporting disciplines can work together to develop this model and then, using HPC resources, begin to answer questions such as “What are the potential impacts on global shorelines if the rate of melting of the polar icecaps increases?” or “What are the consequences if measures under the Kyoto Accord are not implemented?” The answers will be crucial to assess the impact that climate change will have on humans, wildlife, property and economic activity.

A cycle of scientific and technological advancement drives the potential and the need for HPC resources. Faster processors necessitate more powerful computers, enabling the elucidation of larger and more complicated problems, driving intellectual curiosity and mandating even more powerful processors. HPC simulations of our physical world open up new vistas, and the traditional approaches of laboratory experimentation and theoretical analysis are often no longer adequate to achieve the knowledge we are seeking. Storage capacity (both memory and disk), network accessibility, and high-speed bandwidth availability must all be increased as scientists expand their boundaries and tackle more research questions as parts of multi-disciplinary teams. Indeed, the computational modelling of complex multi-scale phenomena is becoming the key technique for providing answers to difficult questions in nanotechnology, climate change, financial analysis, environmental science and transportation.

Canada and HPC

Canada has a long history of research into HPC and of developing HPC applications (see the sidebar on page 36). The premier HPC site in Canada for many decades has been at Environment Canada, where the technology is used for national weather prediction. An academic computing resource was greatly needed, however, and in the 1980s and early 1990s there were three attempts to establish such a facility. All three facilities survived for a short time, but ultimately closed because of a lack of sustained funding. Accordingly, Canada had a low profile in the international HPC research community until the late 1990s.

In 1997, the Government of Canada created the Canada Foundation for Innovation (CFI) to fund research infrastructure. Together CFI and its provincial, industrial and university partners have invested over $250 million in computational infrastructure since in 1999. This has rejuvenated computational research across the country and has stopped the drain of faculty, graduate students and skilled technology personnel from Canada in HPC-related areas of study.

With the advent of CFI, universities across the country were encouraged to collaborate to create the critical mass necessary for large investments in HPC. The result has been the creation of regional consortia to acquire, manage and promote computational-based research. This unique Canadian model of sharing has led to unprecedented cooperation between institutions, more powerful HPC facilities, and higher rates of utilization and shared expertise. The result is greater than the sum of the parts that the individual universities could have achieved.

This is internationally acknowledged as a major success story for Canada.

These major consortia are WestGrid (Western Canada), SHARCNET (South-Western Ontario), HPCVL (Eastern Ontario), CLUMEQ and RQCHP (Quebec), and ACEnet (Atlantic Canada). In addition, there are major HPC installations at the University of Toronto and the University of Victoria, as well as smaller facilities across the country. Every major university in the country either has its own HPC facilities or is part of an HPC consortium.

To be internationally competitive however will require further investment. An oft-cited metric of national commitment to HPC-related activities is the so-called Top 500 List of Supercomputer Sites (see the sidebar on page 37). As with other single metrics like “total research funding”, there is some debate over the methodology and interpretation of the rankings. However, there is no debate that

From Kilo to Mega to Giga and beyond

Not so long ago, computer hardware was measured in mega-something: megahertz of CPU performance and megabytes of memory. Times have changed. Yesterday’s 800-megahertz computers are today’s four-gigahertz machines. Gigabyte memories are common, and high-end machines now support terabytes of memory. Disk capacity has seen explosive growth from a few gigabytes of a few years ago to high-end terabyte storage capacity, with petabyte systems coming soon. As technology improves, we need new names to characterize the growing capabilities. Here is a quick reference guide to the common units of measure: [REF 14]

Prefix	Origin	Number of units		
Kilo	Greek *khilioi* (1,000)	$10^3 \approx$	$2^{10} =$	1,024
Mega	Greek *megas* (“great”)	$10^6 \approx$	$2^{20} =$	1,048,576
Giga	Latin *gigas* (“giant”)	$10^9 \approx$	$2^{30} =$	1,073,741,824
Tera	Greek *teras* (“monster”)	$10^{12} \approx$	$2^{40} =$	1,099,511,627,776
Peta	Greek *pente* (“fifth”)	$10^{15} \approx$	$2^{50} =$	1,125,899,906,842,624
Exa	Greek *hex* (“sixth”)	$10^{18} \approx$	$2^{60} =$	1,152,921,504,606,846,976
Zetta	Latin *septo* (“seventh”)	$10^{21} \approx$	$2^{70} =$	1,180,591,620,717,411,303,424
Yotta	Latin *octo* (“eighth”)	$10^{24} \approx$	$2^{80} =$	1,208,925,819,614,629,174,706,176

A brief history of high performance computing in Canada

1952	First research computer installed in Canada.
1958	Meteorological Service of Canada starts numerical weather prediction.
1964	The first computer science department in Canada is formed.
1985–1991	Cyber 205 supercomputer made available to researchers (Calgary).
1988–1992	Cray XMP-4 supercomputer made available to researchers (Toronto).
1993–1996	Fujitsu VPX240 supercomputer made available to researchers (Calgary).
1993	Atmospheric Environment Service (AES), Environment Canada, peaks at number six on the Top 500 list with one third of the computing capacity of the top entry.
1993	CANARIE Inc. is formed with a mandate to deliver a leading-edge national networking infrastructure to Canadian researchers.
1995	HPC researchers from across Canada meet to discuss common computing needs. This leads to the formation in 1997 of C3.ca, a national organization to promote the interests of the HPC community.
1997	Creation of the Canada Foundation for Innovation (CFI) is announced.
1998	Walter Kohn (B.Sc. and M.A. from the University of Toronto) and John Pople (former National Research Council employee) receive the Nobel Prize in Chemistry. Kohn invented density functional theory, an application that requires extensive HPC resources. Pople's contribution was GAUSSIAN, a simulation package which likely uses more research computing cycles around the world than any other research application.
1999	First CFI high performance computing grants awarded.
2003	Five CFI-funded research facilities appear on the Top 500 list.

many governments around the world, especially in the G7, have made large and sustained HPC investments and that this is reflected in the list. Canada occasionally makes a minor splash on the Top 500 List, but we have never had a continued presence on the list other than Environment Canada. Whereas the United States and Japan have invested a great deal of pride and resources into gaining and regaining dominance in the list (see the sidebar on page 38), the best Canadian system over the 11 years of the list has averaged less than 10% of the computer power of the best system (in the November 2004 list, it is 5.4%). Other regions of the world have also committed more resources to HPC than Canada, even when considering the relative sizes of the countries and their economies. Our global competitors have made HPC a priority by creating at least one world-class HPC centre each, in addition to a number of smaller centres. Canadian HPC sites only represent roughly 1.4% of the computing power on the list – down from 2.2% in 1993. During the period from 1993 to 2004, the computing power of the best system on the list increased by a factor of 700, while the best Canadian system only increased by a factor of 226. Canada is falling further behind in HPC capability despite large investments in this area.

Canada needs to establish a sustained presence on the Top 500 list if it is to support a competitive research environment. The infrastructure will be used to solve some of the most challenging computational problems in academia and industry. A sustained Canadian presence on the Top 500 List will also help to attract and retain HPC expertise, thereby increasing the potential and the usefulness of this technology for Canada. As many sites across Canada have discovered, talented HPC programmers and qualified system administrators are a rare breed. These experts are very mobile and constantly seek to work on high-profile world-class problems, and must be offered a challenging work environment and the facilities need to do the work if they are to remain in Canada.

People, Applications, Communications and Technology (PACT)

A well-connected highly competitive digital world will be one of the dominant features in the lives of Canadians in the 21st century. For this reason it will be vital to have (a) well educated and computationally literate *people* in our universities and research institutes; (b) a vibrant economy and an innovative environment to foster the development of new digital *applications*; and (c) a national *communications* strategy that includes access to broadband networks and to the leading-edge advanced computational *technology* that will be the engine. A successful HPC program is multidimensional, and all aspects must be properly addressed:

People. The effective use of HPC facilities can only happen when strong research teams (faculty, postdoctoral fellows and graduate students) are supported by a team of highly qualified technical people (programmer/analysts and system administrators). Highly-qualified personnel (HQP) with the requisite set of skills are currently in short supply in Canada. This is, in part, a reflection of the historic lack of HPC infrastructure in the country.

Applications. Extensive research and development is needed to develop the applications that can effectively use HPC facilities. Writing a large program that runs correctly on a single computer (sequential program) is still a difficult task. Adding parallelism to an application can enormously complicate the software development challenge. The design of a program that works well on a computer with 10 processors will likely differ significantly from one that is effective with 1,000 processors.

Communications. Resources – people and computers – must be able to communicate effectively. Geography no longer needs to be an impediment to collaboration. With the proper communications infrastructure and tools, virtual communities of researchers can work together, computers across the country can be harnessed as a team to work on a computational problem, and shared data repositories can ensure that all researchers have the latest up-to-date results.

Top 500

Over the years, many non-academic Canadian institutions have appeared on the Top 500 list.

- Companies: AMOCO, BC Tel, Bell Canada, C.O.R.E. Digital Pictures, DST Canada, Ernst & Young, IBM Canada, Hydro-Quebec, Petro-Canada, Pratt & Whitney, Sears, Silicon Graphics Canada, Sobeys, Telus, and the Toronto Stock Exchange
- Government agencies: Canadian National Defence, Defence Research Establishment, and Environment Canada
- Research institutes: Centre de Recherche en Calcul Appliqué, Hospital for Sick Children, and the National Research Council

For the past 20 years, the trends in high performance computers have been monitored through the "Top 500" list of the most powerful computers in the world (www.top500.org). The figure shows the trends in high performance computing. The green line shows the most powerful computer from the Top 500 list, the purple line the bottom entry on the list, and the red line indicates the middle entry (note that the vertical axis is logarithmic). Over the last 20 years, the top machines have become almost 500 times more powerful, while the standard to just get on the list has risen by a factor of over 1000. [REF 15]

Size matters

In 1997 a team of Japanese engineers dared to imagine a computer so powerful that it could keep track of everything in the world at once – steaming rain forests in Bolivia, factories in Mexico belching smoke, the jet stream, the Gulf Stream, the works. What's more, they dared to build it. On March 11, 2002, when they turned it on, the engineers did something no mere mortal had ever done before: they created the Earth. Or at least the next best thing.

The Earth Simulator, the most powerful supercomputer ever built, was designed for a single purpose: to create a virtual twin of our home planet. Before the Earth Simulator arrived, the fastest computer in the world was an American military machine that can perform 7.2 trillion calculations per second. The Earth Simulator runs at more than 35 trillion calculations per second, almost five times faster. In fact, it's as powerful as the next 12 fastest supercomputers in the world put together. Located at a vast, newly built facility in Yokohama, the Earth Simulator is the size of four tennis courts. The price tag? Around $500 million.

It was worth every penny. By plugging real-life climate data from satellites and ocean buoys into the Earth Simulator, researchers can create a computer model of the entire planet, then scroll it forward in time to see what will happen to our environment. Scientists have already completed a forecast of global ocean temperatures for the next 50 years, and a full set of climate predictions will be ready by year's end. Soon, instead of speculating about the possible environmental impact of, say, the Kyoto accord, policymakers will be able to plug its parameters into the virtual Earth, then skip ahead 1,000 years to get a handle on what effect those policies might have. That kind of concrete data could revolutionize environmental science. By digitally cloning the Earth, we might just be able to save it. [REF 16]

Technology. To get research results as quickly as possible, researchers need HPC technology. They need not only multiple processors, but also access to large memories, large disk storage and sufficient network capacity (bandwidth). To achieve high performance, HPC applications use facilities that provide *capacity* (thousands of processors is not uncommon today) and/or *capabilities* (machines with a terabyte of memory or a petabyte of disk space are emerging) that are significantly beyond what can be found in a high-end desktop machine. Furthermore, obtaining the results of computation is only one step in the HPC research process; the data must also be interpreted. Visualization technology is critical to translating the raw computational output (possibly very large data sets) into something visual and amenable to insightful analysis.

People, applications, communications and technology – PACT – are all essential to any HPC initiative. Weaken any one of them and the effectiveness of the overall plan will be seriously undermined.

IMPACT Canada

HPC is foundational to 21st century research. Canada's HPC history clearly shows that sporadic HPC initiatives are not cost-effective, competitive or useful in the long term. A long-term sustained plan is needed for acquiring the infrastructure, nurturing the people and applications, and moving forward in this rapidly changing area. Canada therefore needs an initiative, here dubbed IMPACT Canada, to build and coordinate HPC facilities, to do short-term and long-term planning, to initiate technological outreach, to perform education and training, and to manage the financial, human and HPC technology resources in the country. This initiative would also provide an ongoing assessment of the rapidly changing field of HPC to identify emerging areas of strategic importance to Canada. It would create strategies for (i) enhancing HPC in education and research in the health, environmental, applied, physical, life and human sciences; (ii) coordinating existing resources across the country; and (iii) allocating new resources to maximize the competitiveness of Canadian research. It would have full-time staff to assist with developing programs, training, co-ordinating symposia and providing advice and specialized assistance to HPC centres across the country. It would also interface with HPC users in Canadian government and industry. The initiative would become the primary adviser to Canadian research funding agencies on HPC issues.

IMPACT Canada will play a key role in education, outreach and training. It will work closely with universities, hospitals, colleges and consortia to deliver an early introduction to computational science to young Canadians, now recognized to be a vital component of science education. IMPACT Canada's innovative outreach programs will develop awareness in the minds of our next generation of researchers and workers of the central role that computing plays in both learning and research.

"Over the next decade, the strategic intent of the Meteorological Service of Canada R&D is to provide the science capacity in support of risk-based decision making involving atmospheric and related environmental change and variability which affect Canadians' security and health, economy and the environment on the scale of few hours to centuries. Significant and sustained investments in human and high performance computing infrastructure will be pivotal to attain our strategic objective."

Dr. Michel Béland
Director General
Atmospheric and Climate Science Directorate
Environment Canada

The name *IMPACT* captures the essential elements of a well-connected, highly competitive digital infrastructure: Innovation, Management, People, Applications, Communication and Technologies. IMPACT Canada will foster *innovation*, enable *people* to reach their research potential, assist in the development of the computer *applications* that will lead to scientific breakthroughs, foster the movement of ideas and interaction through broad *communication* strategies, and oversee and coordinate the acquisition and operation of the enabling *technologies*. IMPACT will have a key role to play in bridging the technical and application aspects of HPC, and will thereby increase Canada's overall competitiveness.

Council on Competitiveness study

The U.S. Council on Competitiveness conducted a study of industrial HPC usage in the United States. The report was funded by the Defense Advanced Research Projects Agency (DARPA). The report, released in July 2004, surveyed 33 companies, over half of which had in excess of $1 billion in revenue. The following quotes are from that report: [REF 17]

"High performance computing is not only a key tool to increasing competitiveness, it is also a tool that is essential to business survival. Nearly 100% of the respondents indicated that HPC tools are indispensable, stating that they would not exist as a viable business without them or that they could not compete effectively."

"Companies described a range of impressive competitiveness benefits realized from using high performance computing. . . . Strategic benefits included gains such as shortened product development cycles and faster time to market (in some cases more than 50% faster), not to mention the resultant reduced costs, all of which can improve a company's bottom line."

"Respondents noted a range of reasons that HPC is not used more aggressively. The largest single factor is the lack of computational scientists – human experts (internal or external) who can apply HPC tools to the problems in question – and the budget to hire them."

"Despite the often proven returns from using high performance computing, respondents noted that upper management often does not appreciate the value of HPC hardware and software tools. As a result, HPC is often viewed as a cost instead of an investment. . ."

The National Research Council Canada

NRC is the Government of Canada's leading resource for science and technology development and commercialization. It works across the innovation spectrum from scientific discoveries at the very frontiers of knowledge to the development and commercialization of products for the world's marketplaces.

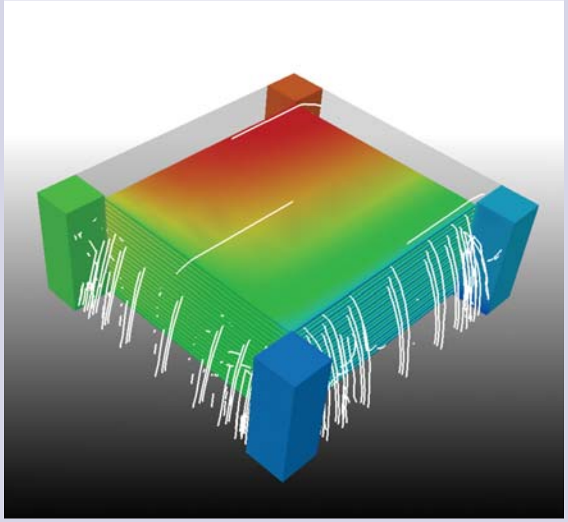
HPC is a major foundation of NRC's innovation programs. It is used for problems that span the computational research domain: large dataset and data analysis work in bioinformatics and astronomy, real-time three-dimensional visualization, finite element and computational fluid dynamics (CFD) for engineering, and *ab initio quantum* mechanics.

Biology is seeing an explosion in new discoveries and tools and there are many emerging applications for HPC. NRC manages bioinformatics data repositories for internal and external clients, develops data analysis algorithms, and is moving into computational biology.

In CFD, NRC's long-term development is oriented to the simulation of integrated, multi-disciplinary designs, and physics-based models for complex flow phenomena. NRC must continually expand resources to meet client needs as these firms move to more powerful and affordable clusters.

In materials research and nanotechnology, the challenge is to understand novel materials and their properties. One early success is an algorithm for protein-ligand bonding affinity, representing important progress towards computer-assisted molecular design.

NRC performs most of its research on internal, department-scale HPC systems, although some jobs are run on external University systems. As needs grow, access will be required to national-scale HPC systems. NRC has extended its research to explore how the grid and middleware tools can better serve its distributed locations and diverse user base.



Computational fluid dynamic simulation of a solid oxide fuel cell stack. (Institute for Chemical Process and Environmental Technology, National Research Council)

Overarching Recommendations

An evaluation of the scientific need (illustrated by the many case studies distributed throughout this document) and the current state of HPC in both Canada and worldwide (see Chapters 2 – 4) leads us to the following overarching recommendations:

1.1 That a long-term funding envelope be developed (a) to provide sustained support for *mid-range advanced* computing facilities such as regional HPC consortia, and (b) to establish a pan-Canadian high-end computing facility that would have a sustained presence in the world's top 30 computing facilities.

1.2 That a national initiative be established to provide outreach, leadership, coordination, and oversight for HPC in Canada.

Understanding the Power of the Sun: The Sudbury Neutrino Observatory

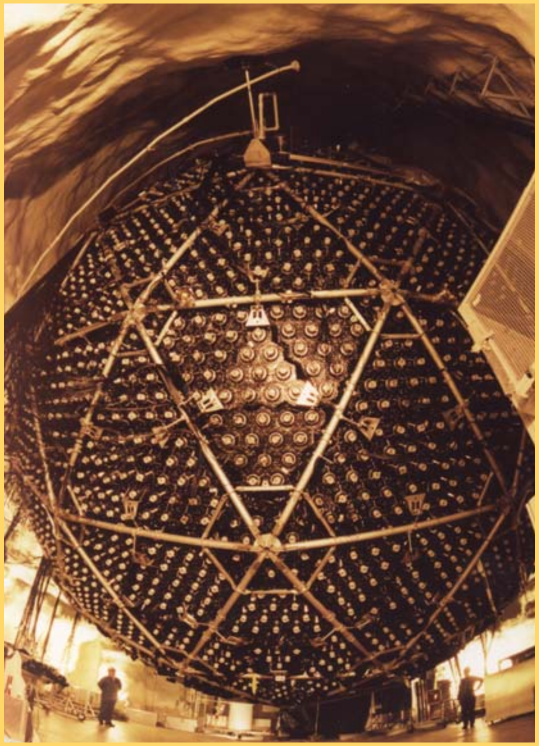
Have you ever thought about how the Sun produces energy? How solar power really works? Physicists have. For more than 30 years, they have been baffled by a basic physics question known as the "Solar Neutrino Problem". The number of neutrinos being observed was far less than predicted by long-held theories about the physics of the Sun; what was wrong?

To address these questions, Canada, the U.S. and the U.K. established the Sudbury Neutrino Observatory (SNO), which has the unique ability to detect and measure interactions among all flavours of neutrinos. Led by Dr. Art McDonald of Queen's University, this international team of researchers measured the reaction rates of all three flavours on neutrino. In June of 2002, the particle physicists working at SNO announced an amazing discovery – they could explain the puzzle of the Solar Neutrino Problem and reveal new neutrino properties. SNO found that neutrinos oscillate as they travel from the core of the Sun to the Earth and that they have mass!

Of course, such a major endeavour had a major HPC requirements for HPC for: data storage, large scale data reprocessing, complex multivariate analysis, advanced stochastic calculations, and CPU intensive Monte Carlo simulations. To continue these globally important discoveries, HPC infrastructure must be sustained. Technical advances made at the SNO laboratory in establishing ultra-low background environments have opened new windows of opportunity for exciting new measurements in particle astrophysics and have led to the establishment of a major International Facility for Underground Science (SNOLAB) funded by the Canada Foundation for Innovation (CFI). SNOLAB will open a window on three major research opportunities:

- Using neutrinos to probe the inner workings of our Sun and thus determine the nature of our primary energy source.
- Understanding dark matter, which manifests itself in the gravitational fields of the universe but is apparently not in the form of any kind of matter that we have so far identified.
- Studying basic neutrino properties via a very rare nuclear decay mode called double beta decay.

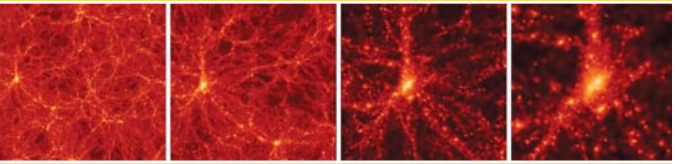
> *"We need high performance computing to support Canadian collaboration in the international quest for an understanding of the most basic properties of space, time, and matter. Advanced computational facilities are a necessity for Canada to continue as a front-rank participant in Particle Astrophysics."*
>
> ***Alain Bellerive, Canada Research Chair in Particle Physics, Carleton University***



The Sudbury Neutrino Detector probes the inner workings of the Sun to better understand the Sun as an energy source. The 10,000 photomultiplier tubes, which view the light produced by neutrino interactions in SNO, are held in place on an 18 metre sphere surrounding 1000 tonnes of heavy water.

"Simulating the growth of cosmic structure, from the star clusters and gas within galaxies to the large superclusters containing tens of thousands of galaxies and beyond, is a huge computational challenge (indeed, it is has been called one of the "Grand Computational Challenges" along with global environmental modelling and protein folding for example) and Canada's continued leadership in this field pivots on access to the appropriate large-scale computational resources."

Dr Hugh Couchman
Professor of Physics and Astronomy
McMaster University



Massive galaxy clusters contain many hundreds of galaxies, many much bigger than our Milky Way galaxy. The bottom right panel shows a simulated galaxy cluster. The remaining panels zoom out by factors of two to show the cosmic environment of the cluster. The top left panel is 500,000,000 light years across. [REF 18]

Understanding our Universe

Astrophysics seeks answers to the most fundamental questions of existence: What is the nature of our Universe, what is it made of, how did it begin, and what is its future? It is a field that studies both the illimitably large (the cosmic) and the vanishingly small (the quantum). The significant discoveries that have been made in the last few years are starting to reveal the astonishing complexity of process and structure that makes our universe.

In recent years, Astrophysics has undergone a transition from a field in which there was a relative dearth of data to one in which there are now staggeringly large data volumes. This dramatic increase comes from rapid advances in detector technologies, from satellite missions delivering high-resolution images of the solar system to telescopes surveying millions of galaxies, neutrino detectors buried deep in northern Ontario bedrock or under Antarctic ice investigating the ghostly signatures of these particles, and huge interferometers scanning for ripples in the space-time fabric that may reveal the existence of gravity waves. The challenge we face is to build, manage, and scientifically mine these enormous databases (tens to thousands of terabytes of data), and to create theoretical models upon which to base predictions that will test and refine/reject models describing cosmic populations and the universe. Because of the physical complexity of these systems, detailed numerical modeling is crucial. The simulations, which generate enormous quantities of data, are completely dependent on sophisticated, high performance computing.

The leading-edge skills acquired by Canadian numerical astrophysicists, both here in our own robust research community and abroad, have produced flexible personnel capable of making significant contributions in a number of fields, including remote sensing, environmental modeling, disease etiology, drug design, signalling and other intracellular systems, medical imaging, as well as the financial and banking sectors. In most of these cases, the value that astrophysicists bring is their ability to undertake large-scale computation and to 'push the computational envelope' in addressing leading contemporary challenges.

The pace of change in computing is now so rapid that leading computations very quickly become routine, and to stay competitive Canadians must invest in leading-edge computers and in the highly qualified computational experts capable of driving the ground breaking research and helping mentor others in widely disparate fields.

Chapter 2: Technology and Communications

Introduction

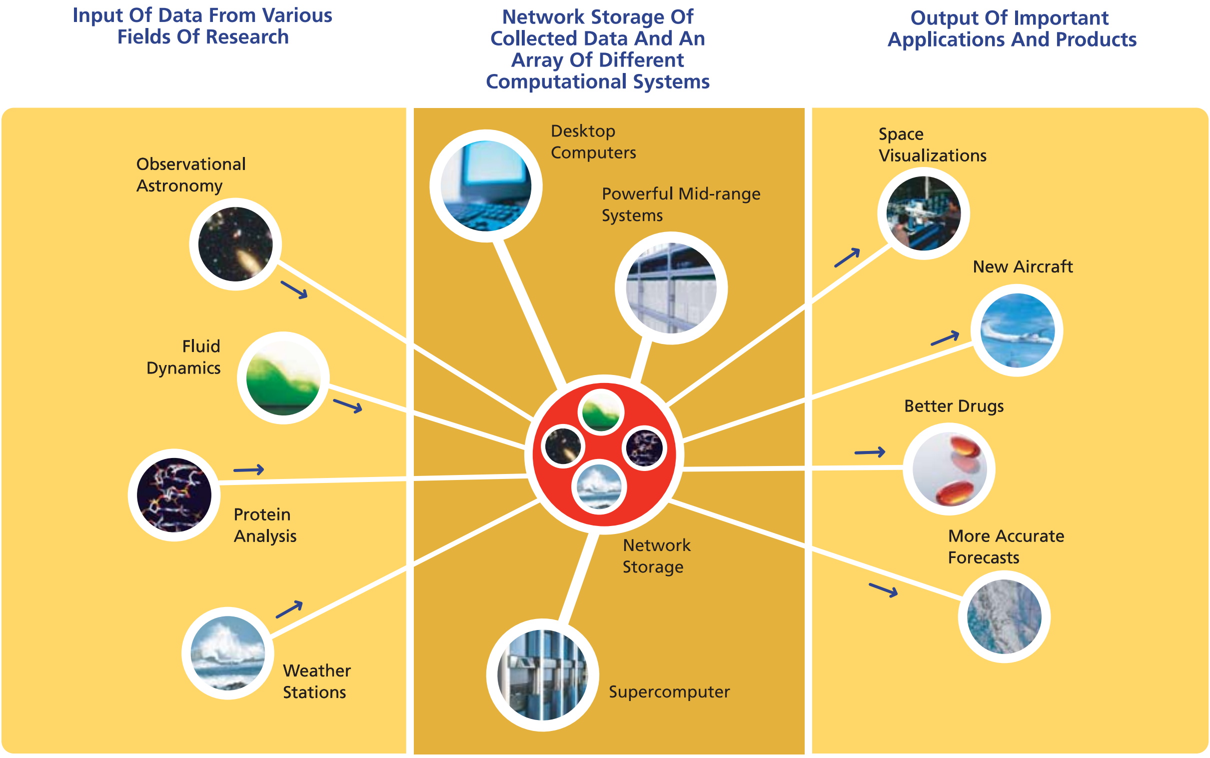
Fewer than 30 years, ago high performance computing (HPC) involved users interacting with a mainframe through a simple, directly connected text terminal – similar in many respects to a slow dial-up connection – or even using stacks of punched cards.

Today, problem solving using high performance computing immerses users in a networked environment that is rich in powerful and interconnected technologies. HPC inputs could be financial data residing at multiple data repositories distributed across continents. They could come directly from particle physics experiments, astronomical observatories or medical imaging systems. They could be entirely virtual, such as a model protein created in the computer or a model of colliding black holes obeying Einstein's equations. These data or models are processed or evolved on powerful computer systems that may consist of many hundreds or thousands of processors in a single system, or be a part of a vast array of networked computers that provide the power to solve the problem without the user ever knowing where the work was done.

The outputs of these powerful systems can, in many cases, only be understood or represented using sophisticated techniques able to handle unimaginably large data sets. For example, visualization techniques enable users to see the active site of enzymes in three dimensions using stereo goggles. Special "haptic" devices in virtual reality environments allow them to move inside, manipulate, touch and even feel 3D images.

These environments allow researchers to understand complex processes rapidly, and thus to refine their models and to control calculations in real time. Such systems can also allow medical doctors to see patient organ structures in 3D, foresters and geologists to visualize landscapes as if they were standing in them, engineers to view and refine their designs before having to manufacture them, and home buyers to choose their home designs before building them.

A Rich Computing Environment



In the research context, users from different continents can interactively collaborate, discuss and modify the same distributed calculation using a network of workstations (some of them wireless). These cooperating technologies provide the underlying infrastructure that supports a dramatically different and accelerated way of doing research – a way that improves our understanding of the world around us, connects us to this world, and allows us to achieve and sustain international research leadership.

The next sections provide an overview of modern HPC technologies, of their capabilities and of the challenges in using them effectively. This chapter inevitably involves somewhat more technical material than elsewhere in this report. The body of the chapter contains the primary discussion, while the sidebars provide information for readers interested in pursuing a deeper technical discussion. (Sidebars marked with ◆ contain the most technical material.)

The Technology

The range of hardware used for computational research is staggering: from single PCs on researchers' desktops to huge systems with thousands of connected processors, massive storage devices and advanced visualization. Computer technology is an enabling tool for most research, and researchers can choose the kind that is most suitable (and most cost-effective) for their projects. Innovative and internationally competitive research is currently performed on a wide range of systems.

The dramatic evolution of high performance computing technology enables researchers to address increasingly sophisticated computational problems. The consequence of Moore's Law (which states that computer speed doubles every 18 months) is to rapidly reduce computing times for specific research problems while simultaneously bringing more challenging problems within reach. An equally dramatic change since the early 1990s is the burgeoning usage of parallel computing. This technology provides a powerful divide-and-conquer strategy for reducing the computing time for many problems, and sometimes is the only feasible means of addressing massive computational problems. For individual research applications, the best technology to use depends on a variety of factors.

Computer Applications and Architectures: the Right Tool for the job

Characterizing high performance computers would require a graph with multiple axes: processor speed, the number of processors, the amount of memory, the various properties of the hardware connecting memory and processor, and the various properties of the hardware connecting individual processors would all have to be represented. Computational research overlays this range of computational characteristics with the specific needs of the particular research application. Just as a straw is useful for drinking a can of soda but is of no use for emptying a swimming pool, so a PC is ideal for computing a household expenses spreadsheet but is of no use for generating tomorrow's weather forecast. In this section, a simple model is presented for understanding how to match computers to applications.

For a large class of applications, perhaps the majority, a key technical constraint is the ability to move data at an effective rate into, or between, contemporary high-speed processors. Indeed, there is a hierarchy of "data pipes" over which information may have to flow. The process may start with data flowing over a geographical network from repositories distributed on the Internet, then moving onto a local disk, and then through various levels of local memory or networking internal to the computer before finally feeding into the processor (or processors). This hierarchy generally has very uneven performance, both in terms of latency (the time for data to start flowing) and of bandwidth (the volume of data flow). Significant bottlenecks frequently "starve" the processor of data and dramatically reduce the efficiency of some applications.

Parallel computers may usefully be categorized by the capabilities of the network that couples the processors, in effect by describing how "tightly coupled" the processors are. Tighter coupling implies a higher performance (lower latency and

◆ "Capability" computing versus "Capacity" computing

Modern commodity processors available in a $1500 PC are in many respects at the leading edge of performance. This is a dramatic shift from the situation at the end of the 1990s, when the leading processor was typically available only in systems costing ten times as much. Today, only specialized vector processors still command a per-processor performance lead, and even here, the vector instructions present in many commodity chips that are designed to support video gaming often provide competitive performance.

The impressive performance of modern processors compared with the rate at which data can be moved over networks has lead to a dramatic simplification of the characterization of computer systems. For many applications running on multi-processor parallel computers, the key element that distinguishes performance is the properties of the network connecting the processors. A high performance network (high bandwidth and low latency) is required to support "fine-grained" parallel applications; that is, those applications which perform only modest amounts of computation before a processor must communicate with its peers, perhaps to get more data. This is called *capability computing*. An example of a fine-grained application needing a tightly coupled, or capability, computer is the simulation of fluid flow over an aircraft wing. Here the change in fluid flow on one part of the wing affects all the neighbouring parts of the wing.

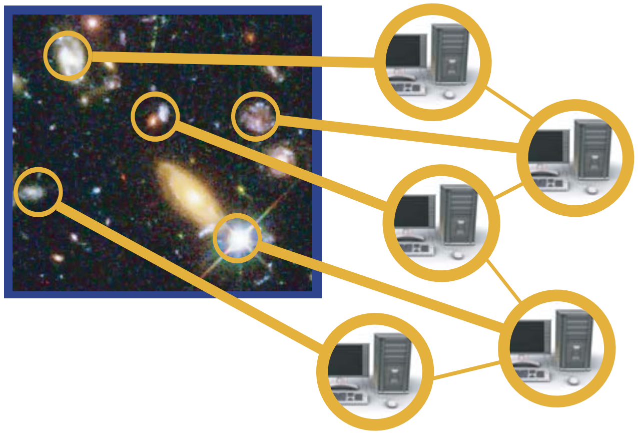
Conversely, a lower performance network can support "coarse-grained" highly parallel applications; that is, those which have very modest communication requirements compared with the amount of computation performed. This is called *capacity computing*. At its most extreme, coarse-grained parallelism can be provided by a networked laboratory of workstations, or even computers networked over regions or countries. An example of this type of computing would be the analysis of light profiles in a million-galaxy survey. The computational task for each galaxy is independent and only the aggregate capacity of all processors is required.

The practical, and frequently dominant, feature distinguishing a large fine-grained parallel system from a collection of $1500 workstations is therefore the performance and cost of the network connecting the processors.

Coarse-Grained, Loosely Coupled Application

Example: Observational Astronomy

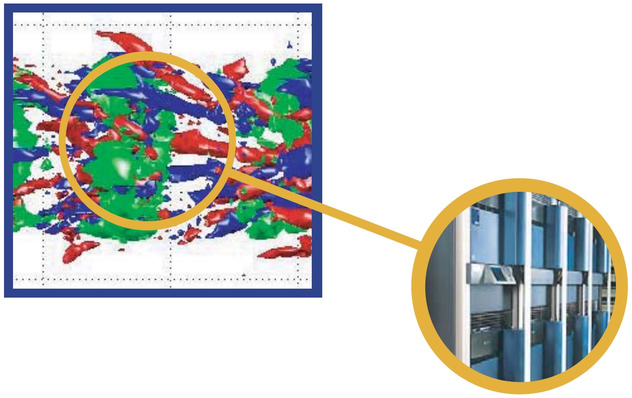
Each galaxy is separate and can be analyzed independently on different computer systems that have little or no connection to each other.



Fine-Grained, Tightly Coupled Application

Example: Fluid Dynamics

Different particles interact with one another, and the problem cannot be broken up into Individual components. Requires a computational system with strong connection between processors.



higher bandwidth) of the network that connects processors, as well as greater system expense. Parallel computer systems are manufactured with a range of network performances from tightly coupled *capability* computers to loosely coupled *capacity* computers (see the sidebar on page 45).

Parallel applications vary widely in terms of their need for interprocessor communication. Some applications are termed *fine-grained*, requiring substantial communication among the processors for a given amount of computation, while others are *coarse-grained* (highly parallel), requiring very little communication relative to the amount of computation undertaken. An ideal tightly coupled parallel computer could run any application, either fine-grained or coarse-grained. In the latter case, however, the expensive network connecting the processors would be underutilized: it makes more sense to use tightly coupled systems for those HPC applications needing that capability. The wide range of research applications demanding access to HPC in Canada requires that we establish the full spectrum of computing capability.

Computing Capability Rings

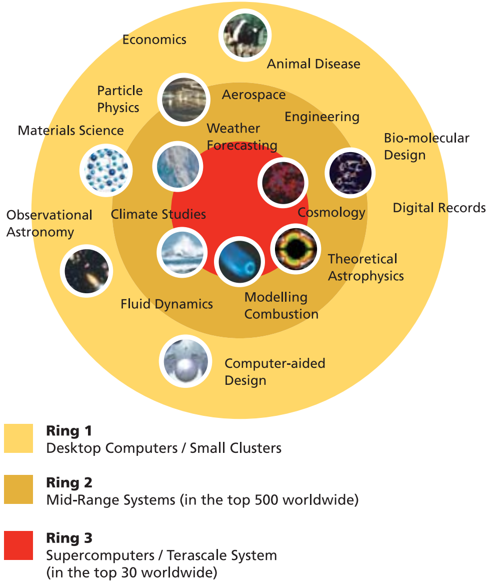
We have chosen to represent the range of computing "capability" that we are proposing to establish with a series of concentric rings.[1] The outermost rings represent the enormous aggregated capacity of desktop machines or loosely coupled systems. Inner rings represent smaller numbers of progressively more capable parallel machines – that is, systems with a greater number of tightly connected processors. Mid-range systems might today[2] consist of perhaps 64 to 512 processors; they would consist either of (i) processors connected by a high performance interconnect, (ii) "symmetric multi-processor" (SMP) systems where processors directly share the same memory, or (iii) a mix of these two ideas with small SMP systems tightly coupled with a capable interconnect. The progressively smaller rings towards the centre represent smaller numbers of the most capable parallel machines that today typically have several hundred to several thousand processors. The clear trend, as reflected in the Top 500 list (see the sidebar on page 47), is for these systems to be constructed from small SMP systems connected by a high performance interconnect. This approach for system development is in many respects the most straightforward and cost-effective when dealing with very large numbers of processors.

The properties of the HPC application dictate where in the ring diagram the appropriate computational resource will lie. Highly parallel, coarse-grained applications with little communication requirement will run on systems in the outer rings. In principle, these systems could be the desktops such as are found in a laboratory of workstations diverted for research use at night or a large, loosely coupled "farm" of processors. Mid-range configurations will be used by single applications that can effectively use perhaps 32, 64 or 128 processors, or by applications that require the total memory available on these systems (often far more than is supported on typical desktop systems). The most complex and computationally demanding fine-grained applications require the greatest HPC capability: the largest and most tightly coupled systems available. These are the resources represented by the central rings.

An important element missing from our picture is the rapid evolution that occurs over time, both in the capability of parallel systems and in the sophistication of the user community. The systems on the Top 500 list occupy the mid-to-central capability

Canadian High Performance Computing Needs

The array of canadian research projects each have unique high performance computing requirements.



range of the diagram.[3] Instead of quoting processor numbers, which change rapidly for each ring, it is convenient to classify a top 30 system as being in the central ring and the remainder as mid-range (see the sidebar on page 49). In the November 2004 list number 30 is a system with 1312 processors and a performance of 5.6 teraflops (Tflops); number 500 has 416 processors and a performance of 850 gigaflops (Gflops). Each year, the performance of these systems approximately doubles (the corresponding figures from the November 2003 list for number 30 were 1024 processors and 2.2 Tflops, and for number 500, 128 processors and 400 Gflops). This progression leads to a new population of central systems, the outward migration of the central systems to the mid-range and the migration of the mid-range systems towards the desktop.[4] Today's supercomputer really does become mid-range in two or three years.

HPC users display an opposite trend. As the sophistication of the user community increases, a greater

Top 500 trends

The Top 500 list was introduced in Chapter 1 (see sidebar on page 37). The historical trends present in these lists illustrate the challenges that must be met in order to maintain competitive parallel computational infrastructure. Here are some interesting trends:

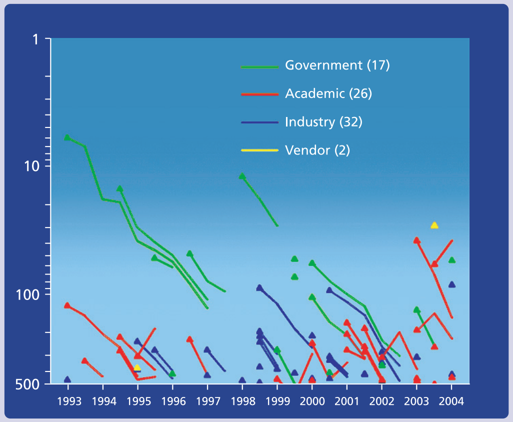
- The performance doubling time for systems on the Top 500 list is close to one year (this is faster than Moore's Law because the number of processors in parallel systems also increases).
- There is roughly a constant (over time) 100-fold performance difference between number 1 and number 500. A system entering the list at number 1 would take roughly 6 to 7 years to fall off the bottom of the list (its approximate yearly rank on the list would be: 1, 2, 5, 12, 30, 80, 200, 500).
- To stay on the list for 3 years (the approximate useful lifetime of many current systems), a system would have to enter the list around number 20. This is the typical entry point for the national facility of many countries.
- Entering the list at number 200 gives a system at most one year on the list. This is the entry point for most Canadian academic facilities (the red triangles in the figure).

These trends indicate very clearly the need to continue to inject funds on short timescales if we are to *maintain* the capability of Canada's parallel computing infrastructure at the national or regional level, and hence the competitive science that it supports.

The figure shows the progression through the Top 500 list of all Canadian systems since 1993. A system above roughly number 30 on the list would be in the central capability ring, the remainder would be in the middle capability ring. The highest Canadian system in the list history, number 6 in 1993, was at the Atmospheric Environment Service in Dorval, Quebec.

The Top 500 list also shows interesting trends in the types of systems populating the top 40. The large majority of these systems are now clusters (composed of nodes with 2, 4 or up to 32 processors). Although posing challenges for the application programmer, the ubiquity of this type of system is driven by the practical difficulty and large cost of otherwise constructing effective parallel systems with thousands of processors. A very recent trend is the appearance of true commodity clusters: inexpensive PCs connected by an inexpensive network. The tremendous cost advantage of these systems coupled with the availability of increasingly powerful and cost-effective interconnects is having a dramatic impact. There were no such systems in the top 40 in the June 2002 list, a year later there were 11 and in the November 2004 list there were 18. For such a system to be useful as a capability computer requires the use of a specialized high performance network; for such systems the cost of the network is frequently comparable to the cost of the processing nodes.

Canadian Systems In The Top 500 List Since 1993



---

[1] A similar categorization uses the idea of a "pyramid of computing capability."

[2] November 2004: As this report was in press, the June 2005 Top500 list was released. It confirms the trends we have described in this report. [REF http://www.top500.org/]

[3] The Top 500 list has become the *de-facto* standard for ranking large HPC systems, but the degree to which the benchmark tests computer *capability* is controversial.

[4] Dual-CPU systems are commonplace for desktops and there is now a push towards multiple-CPU chips.

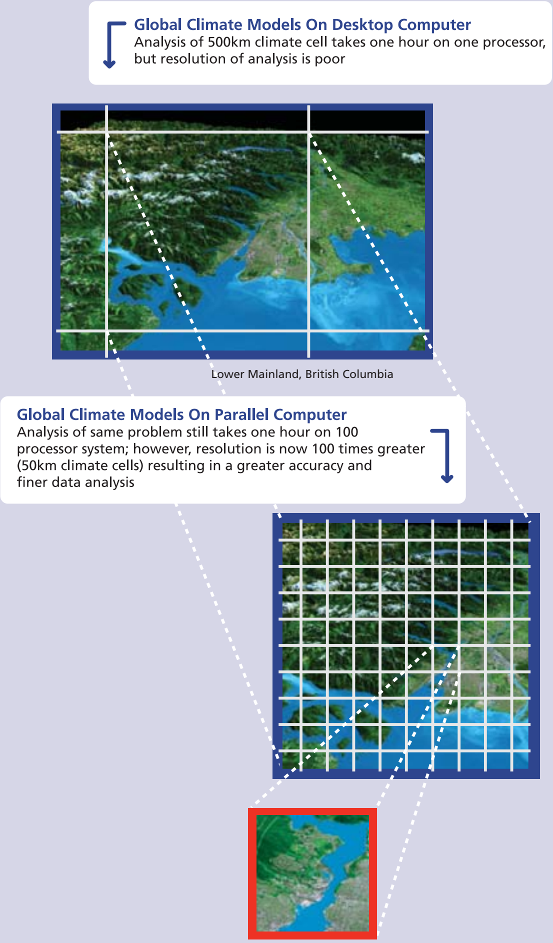
◆ Parallel Applications: strong & weak scaling

Perhaps the most immediate promise of parallel computers is the divide-and-conquer approach, in which a computational task is divided among many processors to reduce the overall execution time. If the application consists of a number of largely independent tasks, then this sort of approach can lead to significant improvements in performance. When the job runs on 100 processors and completes in 1/100th of the time it would take on a single processor, then the application has achieved perfect "scaling"; if it completes in only 1/70th of the time it has achieved 70% scaling. Occasionally one gets lucky and the application achieves super-linear scaling, completing in perhaps 1/120th of the time!

Many applications have sections that cannot be effectively distributed across a significant number of processors, and at some point using more processors will not speed up the application; it may even slow the calculation down. The scaling achieved by running a problem of fixed size on an increasing number of processors is known as *strong scaling*. The success of strong scaling is heavily dependent on the application and on the computer; nonetheless, it is a fair rule of thumb that if an application is well-conditioned, fits in main memory and runs on a single-processor system in a tolerable amount of time, it may scale reasonably to 8 or 16 processors and still show some speedup to 32 or even 64 processors. There are, of course, many highly parallel applications that will show scaling well beyond this, but at some point dividing a given problem into smaller and smaller pieces inevitably leads to each processor having too little data on which to operate efficiently. What, then, is the utility of systems with many hundreds or thousands of processors? A loosely-coupled "farm" of processors can handle many thousands, hundreds of thousands or millions of independent analysis tasks, such as those encountered when analysing events from particle detectors in high energy physics. The primary benefit of such systems compared with a distributed collection of smaller systems is their ease of management and use. Tightly-coupled systems with hundreds or thousands of processors, on the other hand, while they cannot run a particular problem arbitrarily fast, are making possible extraordinary progress for key applications by appealing to *weak scaling*. The idea behind weak scaling is that if a problem of fixed size cannot continue to scale beyond, say, 32 processors, then the additional processing power of several hundred processors can be used to run a larger – or higher resolution – version of the same problem in the same amount of time. This is much more than a way of occupying idle processors, however: it is driving a revolution in the modeling of physical systems.

The global climate model in the illustration is a good example of the idea of weak scaling. If the application models the atmosphere using a grid of 500km square cells and runs on a single processor in a month, then the same application using 100 times as many cells each with sides of 50 km should, if it scales perfectly in the weak sense, run in a month on 100 processors. Running the application with cells with 10 km sides would require 2500 processors to complete in a month. Of course, even weak scaling is not perfect for real applications, but the ability to even attempt such large computations means that we can develop models with higher resolution which are more accurate and have greater predictive power in a huge range of fields, from environmental science and engineering to molecular biology and astrophysics. The common characteristics of these applications are a strong coupling between different parts of the problem and a need for prodigious amounts of memory – hundreds or thousands of times more than on a high-end PC. These fine-grained, *grand-challenge* computations can only be attempted on tightly coupled systems towards the centre of the capability rings.

The Power Of High Performance Computing: Climate Modeling



Global Climate Models On Desktop Computer
Analysis of 500km climate cell takes one hour on one processor, but resolution of analysis is poor

Lower Mainland, British Columbia

Global Climate Models On Parallel Computer
Analysis of same problem still takes one hour on 100 processor system; however, resolution is now 100 times greater (50km climate cells) resulting in a greater accuracy and finer data analysis

number of users will develop and use applications that take advantage of more capable parallel systems. Over time, we thus expect a greater number of research applications to reach further towards the centre rings as expertise grows. Furthermore, the uptake of HPC by non-traditional disciplines such as social science and the humanities is increasing rapidly. These two trends – the increasing sophistication of the existing user base and the increasing number of users – are putting enormous pressure on our HPC infrastructure. Aggressive action to support up-to-date infrastructure is necessary to meet the research community's needs.

This capability ring picture, while highly idealized, nonetheless illustrates that a range of systems must be available to satisfy the community of HPC users. A particular strength of the Canadian HPC landscape is that the range of capability is currently well populated in both the outer layers and the mid-range, and is well supported in these layers by experienced staff. The central ring is not presently represented in Canada. The outer, middle and inner rings of HPC capability are discussed in greater detail in the next sections.

The Outer Layers: Capacity Computing

Leading computational research is being undertaken at all levels of capability and on all types of computing equipment, including desktop systems and small clusters existing in single research groups. Distinguishing what is, and what is not, HPC on such systems was not considered to be a fruitful way to determine if the outermost rings should be covered in this report. For our purposes, we have chosen not to consider systems that may reasonably be obtained by an individual or a small research group. Our definition of HPC thus depends upon cost and coordination. Mid-range facilities require coordination among diverse researchers, perhaps on a regional scale (e.g., the current consortia) and command financial sums that are frequently not accessible even to individual institutions. The central ring requires national coordination and a substantial financial commitment at the national level.

The outer rings are, of course, of crucial importance in enabling a wide variety of computational research as well as providing important connections to the interior rings. The outer rings also represent the aggregation of many loosely coupled systems including systems in labs that may be accessible outside working hours, dedicated serial farms, and even wide-area computational grids. These all make valuable contributions to computationally based research.

The Middle Rings: Mid-Range Computing

The computational capability in the mid-range is in many senses the bedrock of HPC. Here, many researchers' first forays into large-scale and parallel computing are fostered and facilitated by the

Mid-range facilities in Canada

Since 1999, CFI has made major investments in HPC infrastructure across Canada. The funding model used by CFI and the recognition that mid-range HPC supports a diverse research community has resulted in the growth of a number of regional consortia and resource centres. These are illustrated in the accompanying map.

Mid-Range Facilities in Canada

availability of appropriate hardware and expertise. Mid-range HPC supports a vast range of leading research – such as computational chemistry, brain function analysis, pollution modeling, computational fluid dynamics and financial mathematics – that is critical to economic and societal well-being. In addition, the knowledge gained at this middle level permits us to develop the expertise and vision needed to successfully address HPC problems using the central ring and to educate the community in the use of the high-end capability.

Canada's investment in mid-range facilities, primarily through CFI, has enabled computational research at most major research institutions in Canada (almost two thousand faculty, graduate students and research associates). The model of geographical cooperation that has evolved over the last four years has proved to be an extremely cost-effective and efficient way of kick-starting expertise in high performance computing. The success in providing developed, broadly accessible HPC has, in many respects, produced a visibility for Canadian mid-range computing similar to that enjoyed by Canadian networks. Indeed, the futures of these two vital components of the information revolution are closely aligned, and the challenge remains in both cases to maintain the infrastructure in the face of rapid technological advances.

A critical outgrowth of the emergence of mid-range computing has been the development of a highly collaborative environment, drawing together researchers from research institutions from across Canada. This has produced a web of HPC facilities and technical analysts that has created an effective and pivotal support network. Regional HPC consortia have emerged that capitalize on local strengths and needs. All of these consortia are intent on sustaining vibrant research environments that are increasingly dependent in every aspect upon HPC. These efforts have resulted in thirty of Canada's universities organizing themselves into eight major HPC resource centres and consortia, each leveraging the value of regional cooperation (see the sidebar on page 49).

These consortia are having a profound effect on many disciplines, fostering innovative research and the growth of new research communities. These HPC facilities enable universities to offer new programs in computational medicine, arts, engineering and science that are often inter- and multi-disciplinary. These programs include computational health and telerobotics, computational biology and drug design, computational nanotechnology, computational fluid dynamics, computational forecasts for weather and the environment, computational oil and gas reservoir modeling, and disaster remediation. The shared mid-range facilities act as magnets in their communities, attracting bright young academics and students to the participating universities and creating an environment for skill development that is critical to Canada's ongoing research capability. Furthermore, success at this level is a prerequisite for the effective use and implementation of high-end computing capability.

The Central Rings: Tera-Scale Computing

At the centre of the capability picture are the so-called "tera-scale" systems. These are tightly coupled systems (presently) having a thousand or more processors and associated storage, which are capable of sustaining teraflops (Tflops) performance on single user applications. The more useful definition (introduced earlier) of a tera-scale system that recognizes the inexorable performance increase of Top 500 computers, is a system that ranks in the top 30 of this list. Such systems now exist in national research centres in many countries, but this capability is not currently available in Canada.

Tera-scale systems are focused on problems that simply cannot be addressed any other way. Frequently these "grand challenge" problems require substantial amounts of dedicated time on systems with many CPUs and with very large total memory, and require specialist programming support. The facilities, rather than being available for general access, are focused on key peer-reviewed projects of exceptional scientific quality and significance. In many respects, tera-scale facilities are akin to other large scientific instruments such as particle accelerators and astronomical observatories.

Computer modeling and simulation of physical systems has achieved the status of a "third way" between traditional theory and experiment: in many cases, computation is the only plausible way to address key scientific questions and to link theoretical understanding with observed phenomena. In many natural systems, however, the range of scales present is dramatically greater than it is possible to model on any current computer. For example, it has been estimated that, at the current yearly increase in computing speed, the detailed modeling of airflow over a full aircraft will only be possible in 2020. This future computing system will need 30,000 times the performance of the computer that is currently the world's fastest! The differences between reality and our approximate models drives many computational researchers to use the largest computers available, striving to more closely and accurately model nature. The sidebar on page 48 examines in more detail this mode of operation of tera-scale systems.

Most of the disciplines requiring access to tera-scale systems are also heavy users of mid-range systems. Indeed, skills are often honed on mid-range systems and preparatory work for very large tera-scale calculations done at the mid range. Within each of these disciplines, however, there are a select few *grand-challenge* problems for which scientific understanding and progress depend fundamentally on the existence of tera-scale systems. Specific examples of such critical problems are discussed in the sidebar on page 55 and include the modelling of turbulence and combustion, coupled oceanographic and atmospheric climate systems, astrophysical supernovae and structure formation, protein folding, cell signalling systems and the electronic structure of materials. Not having access to tera-scale computing capacity has significant consequences in terms of the international competitiveness of the disciplines that require systems beyond the mid-range. It has implications for Canada's ability to attract and retain those scientists, professionals and students who work in these fields. Although it is certainly possible for researchers to use offshore large-scale computational resources, they are extremely costly[5] and this practice inevitably limits the growth of Canadian expertise. It also limits the health and vibrancy of the various research communities and severely degrades quality of the training of HQP in these disciplines. Many of the techniques learned in the use of current tera-scale systems generate knowledge and skills of direct relevance for competitive future mid-range HPC.

The fundamental decision to be made is whether or not Canada wants to generate and support internationally competitive research in key areas requiring tera-scale computing. A plausible goal for Canada is to maintain a system with performance within the top 20 or 30 worldwide to support competitive tera-scale applications. Achieving this goal is essential to our current and future research success.

The critical importance of maintaining a comprehensive HPC infrastructure spanning the middle and central capability rings motivates a central recommendation of our report:

2.1 In addition to the continued funding of foundational research computing infrastructure (desktop, workstation and small system) by the granting councils, we strongly recommend that the research computing infrastructure in Canada moves towards a model that supports regional mid-range computing facilities and a world-class high-end facility. By 2012, this will require the following (in order of priority and including in each case capital infrastructure, human infrastructure and operational costs):

a. $76 million/year in sustained funding and ongoing support for mid-range computing facilities. This support will (i) build on the present HPC regional consortium model (where universities in a region cooperate to provide shared multi-purpose HPC resources, as well as the human infrastructure to operate the system and to train and support the researchers), (ii) support major special-purpose in-house computational facilities for exceptional cases where it is impossible or inefficient to use a shared multi-purpose facility, and (iii) promote collaboration and create synergies that extend beyond computation.

---

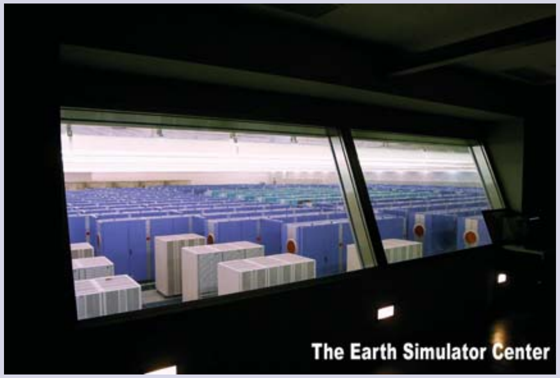
[5] At perhaps $1/CPU-hour, a large simulation of 200,000 hours would cost $200,000. One year on 1024-processor system is nearly 9,000,000 CPU-hours.

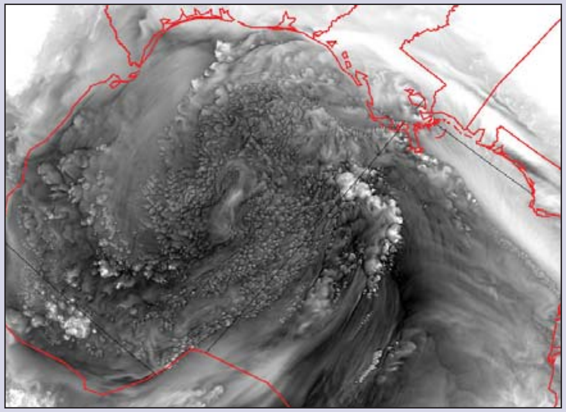
Meteorology

Canadian meteorological scientists who need to achieve very high-resolution simulations of severe weather precipitation systems seek high performance computer resources outside the country. To carry out a complete simulation of Hurricane Earl (1998), which developed into a system with heavy precipitation affecting the Maritime provinces, a Canadian team used the world's fastest computer, the Earth Simulator (ES) in Japan. This system which held first place on the Top500 list for two years now ranks third.

The team was composed of Drs. Brunet and Desgagné (Meteorological Service of Canada – MSC), Professors Yau and Gyakum (McGill U.), and Dr. Ohfuchi (ES Centre). The Canadian MC2 model achieved 10 Tflops on an 11000 x 8640 km domain with 50 vertical levels covering North America and the North Atlantic. This outstanding performance was attained with approximately 4000 vector processors on 500 nodes. This is 25 times more than can be achieved on the new MSC IBM cluster that currently ranks 74 on the Top 500 list.

The image shows the simulated specific humidity (altitude 325 m) on September 1, 1998, for a 1000 x 1000 km sub-domain over the Gulf of Mexico. It shows 1% of the total surface of the computing domain! The pixels are at the model 1 km horizontal resolution. The simulation shows realistic small weather disturbances evolving around the spinning hurricane.





b. $19 million/year to establish and sustain a multi-purpose pan-Canadian high-end computing facility and to fund appropriate support personnel. We anticipate only one such facility in Canada.

Data and Visualization: Making Sense of the Data

Large parallel systems, able to compute at 100 Gflops or several Tflops, can process or generate prodigious amounts of data. A typical system in the top 30 of the Top 500 list could have a terabyte (TB) of main memory or more; if this computer was used to run a large climate model, for example, it might generate 40 TB of data over the course of a simulation. This is more than 1000 times the data that a desktop PC with a 40 GB hard drive could store! Many HPC fields are now data dominated, a trend that will only increase over time. These data must be stored and in addition must often be made accessible to researchers located around the world.

With the ever-increasing size of the data sets comes the challenge of interpretation. In many fields, the ability to visualize data is a crucial element in extracting the full scientific value of the computation. Parallel computation has again produced a significant challenge: the ability to generate data has far outstripped our ability to easily visualize and interpret it, particularly in terms of interactive visualization. This situation is starting to change with the emergence of cost-effective parallel visualization engines that can handle these very large data sets. Generating the full scientific return from these data sets requires graphics programmers, data-mining experts, machine learning experts and even visual artists.

Networks and the Grid

For most HPC users, networks have traditionally provided the service required for access to distant HPC systems and for the transfer of data to and from these systems. The current view is that broadband networks will enable a "computational grid" by supporting and integrating a rich array of services and hosting diverse distributed resources. This grid supports – in a secure environment – many of the tasks associated with using HPC, including submitting jobs, managing data and visualizing results. Like electricity grids that provide utility power simply by plugging into the wall, computational grids provide utility HPC resources simply by logging in. Although it may be hard to easily incorporate all HPC usage into the grid, it is clear that a vast range of HPC services can be made available to a wide range of geographically dispersed users, significantly easing use and increasing research productivity. Canada is extremely well placed to pursue this course and to build on its leadership in network technology and on the forefront national network already established by CANARIE (see the sidebar on page 54). This national network, together with the consortia as natural regional components of a national grid, provides an ideal framework within which to pursue these opportunities.

Data, data, data……

Everywhere there are data, and the volume will increase as we rely more and more on electronic rather than paper records. These data require stewardship and many must be curated. Scientific research data can be obtained from experiments involving massive amounts of information from one location, such as the Sudbury Neutrino Observatory (SNO); the data may also be from many locations, such as those being collected from a huge array of sensors associated with the NEPTUNE project in the northeast Pacific Ocean. Huge amounts of data also are generated from simulations of complex physical and chemical phenomena. Data from publicly funded agencies are critical to many fields of research. All these data must be sifted to select those that need to be retained, and they must remain secure.

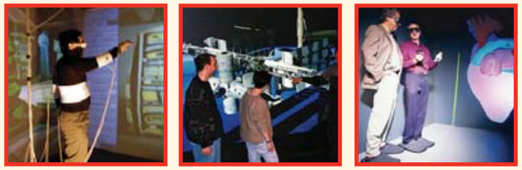
For example, consider the world-acclaimed SNO: its 9600 photomultiplier tubes collect data continuously and produce about a terabyte of raw data per year. These data are unique and therefore must be preserved for future generations of scientists to reprocess as new theories are developed. SNO-LAB is being established to broaden the experiments to answer questions about other topics such as "dark matter." These additional data will increase the pressure to provide data storage facilities.

The reliance on HPC to either generate or process raw data will increase. It is estimated that thanks to the Canada Foundation for Innovation (CFI), there are about 100 terabytes of hard disk storage installed at academic institutions, with many times that in tape backup. Researchers continue to scramble to handle the ongoing expectations for data archiving and stewardship.

> *"As a country that has invested heavily in high-speed communications, high performance computing, and is moving toward e-government, e-commerce, e-everything, the fragility of those data generated suggests that if we do not place these data into archives and curate them, we may face a generation of lost scientific effort."*
>
> ***Dr. David Strong***
> ***Chair***
> ***National Consultation***
> ***on Access to Scientific Research Data***

VISUALIZATION: Making Sense of the Data

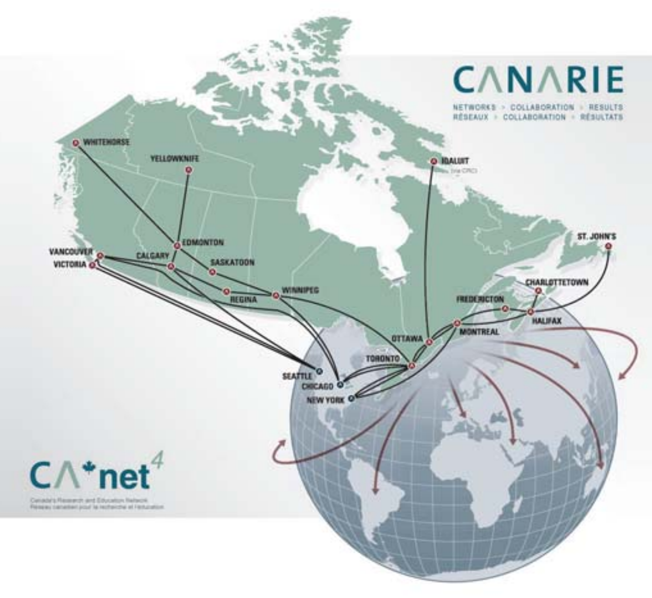
In many fields, the ability to visualize the increasingly complex data generated by high performance computer systems is a crucial element in extracting the full scientific value of the data.



CANARIE Inc

CANARIE Inc. is Canada's advanced networking organization. Since 1993, it has been funded by the Government of Canada to develop and operate the country's research and education network, currently CA*net 4, and to facilitate development of advanced applications and technologies relating to networking.

CA*net 4 embodies an innovative design that allows researchers to connect at very high speeds to distributed HPC sites as well as to remote sensors, experimental facilities and data repositories. Since research applications such as these place extreme demands on the network, portions of bandwidth have to be dedicated to them or else they would interfere with all other concurrent transmissions. Creating an automated way for researchers to establish and manage such dedicated pathways across the network is one of the key objectives of CA*net 4. This has been accomplished through software similar to that used to provide access to distributed HPC systems. The similarity in the approach to the two applications is leading to the concept of an integrated architecture and infrastructure for a wide range of related applications, variously called an "intelligent" infrastructure or a "cyber" infrastructure.



Within Canada, CA*net 4 currently consists of three 10 Gbps wavelengths and thirteen points of presence, with at least one in every province and one each in the Yukon and North West Territories. These are the points at which interconnections are made with the Optical Regional Advanced Networks (ORANs). Each ORAN is responsible for interconnecting institutions and other research and education sites within its respective province or territory.

Internationally, CA*net 4 interconnects with all the leading research networks in the world. Through its innovation of bandwidth dedication, CANARIE is one of the leaders in the development of international research networking: the software used to automate the process is being studied by research networks around the world for possible implementation as a global standard. CA*net 4 also offers transit services across North America to international peers and is becoming an international hub for research networking.

Clear competitive advantages will accrue to Canada by pursuing this objective, but the success of an integrated HPC network infrastructure will be crucially dependent on maintaining our leading broadband networks. This foundational requirement motivates the following recommendation:

2.2 We recommend the continued support of CANARIE Inc., the optical regional advanced networks, and Canadian grid initiatives, and that these networks be maintained and enhanced in concert with new HPC investments.

Grid technologies are widely used in areas such as finance, defence research, medicine, decision-making and collaborative design. A grid allows people from different organizations and locations to work together to solve specific problems. The US and Europe are accelerating funding for projects that develop and use grid architectures, and the expectation is that these paradigms will play a central role in future high performance research and infrastructure. The development of Canadian computing platforms to permit collaboration across organizational boundaries and among distributed members of communities of practice will (i) enhance Canada's ability to develop unique solutions to complex problems, (ii) allow researchers to capitalize on our existing expertise and network infrastructures, and (iii) maximize the benefits, both social and economic, that will be derived from the innovative and advanced research thus supported.

Computational grand challenges

A common element of many problems appealing to tera-scale computing is the very wide range of scales present in the problems – a range of scales that is typically substantially greater than can be accommodated on even the largest current systems. The following table identifies a number of these fundamental computational challenges and the capability in sustained Tflops for a *single application* estimated to be required to make progress towards their solution. In many fields the competitive edge is set by the ability to model systems at greater resolution, detail and accuracy. The ability of Canadian researchers to compete with their international peers in tackling these pivotal problems demands access to a sustained infrastructure providing world-competitive capability computing.

A tightly coupled system at number 10 in the Top 500 list might achieve close to 5 Tflops sustained on a highly optimized fine-grained application today (a somewhat more conservative estimate than the 10 Tflops of the tenth system on the current list as measured by the Linpak benchmark). At the present yearly performance doubling rate of Top 500 systems the two future performance milestones at 20 and 80 Tflops will be reached in 2 and 4 years respectively.

	5 Tflop (2004)	20 Tflop (2006)	80 Tflop (2008)
Turbulent combustion		Modelling of laboratory flames	Practical flames: simulations of environmental and industrial combustion; integrity of structures
Climate modelling	100 km atmospheric resolution; 50 km oceanic resolution; multi-century resolution	Include atmospheric chemistry and carbon cycle	Full coupling of atmosphere (50km) and ocean (15km); multi-decade resolution
Astrophysics	Accurate large-scale cosmic structure	Galaxy formation in cosmic volumes; Supernova simulation; full analysis of cosmic microwave background radiation	
Protein folding	Folding of 50 amino acids over micro-second intervals		Folding of proteins over milli-second intervals
Structural biology	Simulation of sub-cellular systems at the atomic level		Simulation of full bio-molecular complexes for micro-seconds
Nanoscience: Electronic Structure of materials		First principles calculation of electronic structure of a 10,000 atom system (interaction of a small of cluster of nano-particles)	
National-scale economic modelling			Pflop (1000 Tflops) computing: 2009 on the most powerful systems, 2012 on number 10 →
Multi-user immersive virtual reality			

[REF 19]

◆ Grid computing

Grid computing is an extremely active area of computer science research at present. Much of the research is involved with developing secure protocols for accessing resources and data, policies for allocating resources, and software for managing them. In a very general sense, the focus is on enabling “virtual organizations” to bring together groups of individuals and varied resources to solve common problems.

From the perspective of HPC users, three important grid objectives must be addressed:

Grid computing. Facilitating the transparent migration of computational jobs to appropriate and available resources is a key aspect of grid computing. This capability is already available to a limited extent. Existing software can identify appropriate resources and migrate jobs under pre-defined allocation policies that accommodate constraints such as resource ownership and job needs. A second aspect is the possibility of running single computations across geographically separated clusters. However, there are fundamental limitations associated with the latency involved in transmitting data over large distances. (The latency between two centres 150 km apart, allowing for a request to go out and data to come back, is 1 millisecond or 1 million clock cycles on a 1 GHz processor.) Only certain carefully designed coarse-grained algorithms will be able to avoid starving processors of data under these conditions.

In this sense, the grid is a gigantic capacity computer covering the outer rings of the capability diagram.

Data sharing. This includes the automatic movement and access of data across distributed storage locations and the automatic preparation of data for processing by varied computing resources. Data sharing technologies are already in use by the particle physics community to handle accelerator data, and by the astronomical community to create virtual observatories that combine existing distributed data sets.

Remote collaboration. A central attraction of broadband networking is the ability to shrink geographic distances among researchers. The ability to bring researchers together via high quality audio and video links and incorporating real-time interaction via smart boards in a multi-point environment will be of pivotal importance in fostering Canadian research productivity, training new researchers and creating a critical mass from a dispersed research community.

Grid projects

A number of highly parallel applications lend themselves to grid computing. Perhaps the best-known example was SETI@Home, in which spare cycles on idle desktops were used via a screensaver to search for intelligent extra-terrestrial signals in radio data. Another example is climateprediction.net, in which thousands of climate models running on private PCs will be coordinated to produce the world’s largest climate prediction experiment.

Canada has its own grid capacity, the Canadian Internetworked Scientific Supercomputer (CISS), which runs on our powerful mid-range facilities. CISS has been used for computational chemistry applications, accumulating the equivalent of several years computing time in a single weekend.

In Europe and the USA, hundreds of millions of dollars are being invested in grid computing. The “TeraGrid” in the USA links nine of the leading supercomputing sites with an aggregate computing capacity of 20 Tflops, nearly 1 PB of storage, visualization capabilities and the software to enable grid computing. All components are linked by a dedicated 40 Gb/s network.

Supercomputers to Supersimulators: Keeping Canada's Aerospace Industry Flying

The aerospace industry is one of the most competitive, with a constant global push for leading-edge aircraft, jet engines, avionics and flight simulators. The industry's challenge is to design the most efficient and safest technical product, on time, on budget, and within regulatory requirements. To achieve this, large-scale simulation, analysis and design are used for improving performance, increasing quality and lowering production costs. This cannot be more vividly demonstrated than by the unexpected new challenge and threat from China, which is entering the Regional Jet market with the ARJ21, at rock bottom prices but with very modern technology. AVIC1's new research and design facilities in Shanghai, Xi'an and Beijing are awe inspiring, as is the Supercomputer Center facility that has been put at their disposal in Shanghai. This demonstrates that even in an economy based on inexpensive labor, wisdom dictates that the only way to achieve quality and time-to-market in aerospace is to employ advanced HPC to simulate as many of the processes as possible.

Canadian aerospace, ranking fourth in the world, must keep pace with the rapid acceleration of new developments and expand its computational resources to maintain its position. A simulation that twenty years ago would have taken 365 full days to complete can be solved today in only 10 seconds by HPC. This acceleration factor of 2 million is also accompanied by a substantial price drop that has not been even remotely matched by lower material or manpower cost, making simulation the primary long-term way to save money and time. This phenomenal speedup has changed the virtual design paradigm: instead of assessing several virtual prototypes by simulation, akin to running a numerical wind tunnel, major companies now directly predict, via multidisciplinary optimal design, the shape of wing, compressor or turbine blade that gives the best performance in terms of fuel, strength, reliability, weight, noise, in-flight icing, etc., simultaneously. For flight simulators, while Computational Fluid Dynamics (CFD) has taken a long time to gain acceptance in building more comprehensive training databases, the current trend is accelerating, with ambitious plans to use of CFD in real-time. The impact of HPC is that not only are experimental test cells and wind tunnel usage reduced, but also the manpower of engineering departments is significantly shrinking due to the advent of direct multi-disciplinary design methods.

If one were to ask leading Canadian aerospace manufacturers what major edge they would like to maintain over their competitors, part of the answer would be how much faster or more thorough they are in implementing HPC simulation and design.

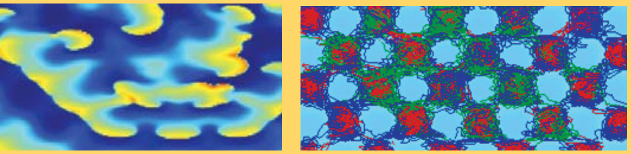
Bombardier is leading the way in the release of the new Global 5000 business jet: it sets a new standard in transcontinental business jet travel, flying 4816-nautical-miles (8920 km) non-stop with an 8-passenger payload at Mach 0.85 (562 mph or 904 km/h). Designed specifically to fly transcontinental missions at Mach 0.89 (590 mph or 950 km/h), the Bombardier Global 5000 sets the standard for speed and range in its class. Designing this aircraft would not have been possible without Bombardier's HPC simulation facilities. [REF 10]

"Today, being second best in an aerospace niche will not keep a Canadian company alive. An aircraft or a jet engine that is ever so slightly more expensive, heavier, noisier, less fuel efficient, prone to frequent overhauls or accidents, or with a shorter lifespan, will lead to the demise of a company, perhaps entraining with it the Canadian Aerospace industry. Hence, it is no longer sufficient for our products to be "good," they have to be "optimal," in other word the best. Only through simulation can we aspire to achieve this. Computational power must continually be invested in, to sustain and grow Canada's aerospace industry!"

Fred Habashi
NSERC Bombardier Industrial Research Chair
of Multidisciplinary CFD
McGill University

> *"Those virtual models require very extensive computations. Present modelling involves 720,000 cells, with the activity of each cell determined by 24 equations that have to be solved 80 to 200 times a second, depending on the conditions. We are working with highly complex models and making them more complex at each stage to reflect the human heart better and better. With 12 processors, it now takes four days to model 20 seconds of cardiac activity. High performance computation is an absolute must, and the use of even more powerful computers will definitely be appreciated.*

Dr Stanley Nattel
Paul David Chair and Professor of Medicine
University of Montreal and Montreal Heart Institute



Left photo: A snapshot representing a brief period of activity (1/1000 second) during atrial fibrillation in a simulated atrium containing 720,000 cells modeled to represent exactly cardiac electrical activity. The red zones are the front of the electrical wave, and yellow, green and pale blue represent progressively less activated tissues, with the dark blue indicating resting tissue. The white arrows indicate the directions of propagation of apparently random atrial electrical waves. Right photo: A quantitative analysis of the organization underlying the apparently totally random activity. Green areas indicate sites of long-lasting activation sources, which maintain activity elsewhere in the system.

To the Heart of the Matter: Modelling Human Arrhythmias

Cardiovascular disease has been on the decline in recent years, but is still the leading cause of death in North America. Atrial fibrillation affects about 10 per cent of people over 70 years of age and can lead to life-threatening complications, such as stroke or heart failure. With an aging Canadian population this percentage will increase over the coming years unless improved ways to monitor and treat atrial fibrillation are found.

The heart consists of four cavities – two atria and two ventricles – and is controlled by a highly sophisticated network of electrical pathways. It works like a pump, with the cavities responding to electrical signals and requiring tight synchronization. The right atrium receives incoming blood from the veins; the right ventricle sends that blood to the lungs. The blood returns from the lungs through the left atrium and is pumped to the arteries. Proper coordination of these activities is essential to keep the heart running smoothly. At times, however, the electrical activity falls out of sync, creating arrhythmias of several possible origins.

The arrhythmia that is most frequent and hardest to treat is atrial fibrillation (AF), which completely disrupts the atrial rhythm. A great deal is known about the different types of heart arrhythmia and their origin as far as the heart is concerned. Atrial fibrillation causes the upper chambers of the heart to beat irregularly, which interferes with blood flow.

What stumps researchers is the cellular and molecular mechanisms involved, and that makes it hard to design original therapies. To learn more about the mechanisms behind this condition, researchers have built computer models of the electrical activity in the heart's upper chambers and conduct simulations to see if new or existing drugs might work.

These simulations allow researchers to investigate roles played by various aspects of heart geometry as regards the physiopathology of atrial fibrillation. The simulations would have taken many weeks to perform on personal computers but by using the high performance computing power the tests took hours and the researchers were able to perform hundreds of simulations. The knowledge gained through these studies improves the ability to treat atrial fibrillation, leading to the development of new drugs or surgical procedures and "ablation procedures" meant to modify the functional geometry and suppress arrhythmia.

Predicting Weather Elements: Thriving in Canada

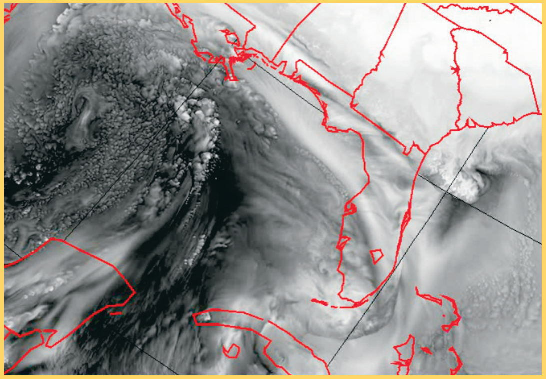
The number one reason people watch or listen to a news program is to get the weather forecast. Accurate weather prediction affects decisions we make daily, and much more. Canada is a world leader in weather forecasting, environmental modeling and prediction. Given the weather we endure, perhaps it is not surprising that Canada's meteorological research is a major international success story. Weather and environmental prediction in Canada is part of a cooperative international program that monitors, studies and predicts changes in the global atmosphere (weather, climate, and atmospheric chemistry), hydrosphere (rivers, lakes, oceans) and the cryosphere (snow and ice).

Central to all of these is high performance computing. Numerical weather prediction was one of the most important motivations behind the first computer applications fifty years ago and is still a major user of HPC. Computational power is needed to improve the accuracy of weather and climate models by improving space-time resolution, the representation of physical processes of models and better quantifying modeling and forecast uncertainties. Our five-day forecast today is as accurate as a day forecast from 40 years ago. Each order of magnitude increase in computing power pushes this accuracy out another day, with huge economic impact for cities clearing snow or peach growers packing fruit.

The goal here, however, is not only to have predictable weather for risk management and emergency response or to decide what to wear, it is about a whole systematic approach to predicting ongoing environmental changes. Future climate change as a result of human activities is an issue of great social and political importance. Global climate modeling is key to both public policy and business strategy responses to global warming, the polar ice melt and long-term pollution trends. The same modeling methods are also being applied to better understand such high-profile environmental issues as water scarcity, water basin management, forest fires, ocean current and temperature changes that impact on local climate and fisheries, and long term trend in ozone.

The policy debates around some of these issues are intense – driving intense demand for better scientific models and simulations. The value of those models depends, of course, on their quality and accuracy, hence on increasing HPC performance. On this point, the 20th century track record of this discipline provides a solid base for confidence.

> *"More accurate weather prediction generates significant economic benefits – in transport, agriculture, tourism and water management. In the public sector, more accurate, timely predictions of high impact weather events – winter storms, summer tornadoes, squalls, flash flood, hurricane and storm surge – are critical to effective emergency response. "*
>
> ***Gilbert Brunet***
> ***Director of Research***
> ***Environment Canada***



This image shows the simulated specific humidity (altitude 325 m) on September 1, 1998, for a 1000 x 1000 km sub-domain over the Gulf of Mexico and Florida. It shows 1% of the total surface of the computing domain! The pixels are at the model 1 km horizontal resolution. The simulation shows realistic small weather disturbances evolving around the spinning hurricane.

Chapter 3: Human Infrastructure

Since 1999, the investments in high performance computing (HPC) made by the Canada Foundation for Innovation (CFI) and its provincial and industrial partners have increased the number of researchers working on HPC-related projects from a few hundred in 2000 to almost two thousand in 2004. If this momentum can be maintained – and even accelerated – these investments will have played a critical role in increasing Canada's commercial, research and development competitiveness. They will also have led to a critical base of human resources capable of leveraging HPC capabilities to Canada's economic and social advantage: another key requirement in a global economy. [REF 20, 21]

All across Canada, world-class research teams are using HPC resources to make new discoveries. The breadth of research applications is staggering, encompassing virtually all areas of the sciences, engineering, and medicine, with growing applications in the social sciences and humanities. High-performance computing is changing the way people think about problems, allowing them to solve problems whose scope was absolutely impossible only a few years ago. These are exciting times, and the research results are often foundational as well as commercially relevant. This is inspiring a new generation of researchers.

Although it is the lead researchers that assume the high-profile roles in the research process, for HPC-related activities there has to be a large supporting team working behind the scenes. An effective HPC facility is much more than hardware; the smooth operation of the facility requires a second tier of highly qualified personnel (HQP) to manage, operate and maintain the facility. It is equally important for these highly trained technical support staff to train and assist researchers in making the best use of this expensive infrastructure.

An investment in people for today and for the future is a critical component of this proposal. In many respects, computing infrastructure can be more readily acquired than human infrastructure. Given adequate funding, the upgrading of the capital equipment is straightforward: one can simply buy whatever is needed. However, human infrastructure is much more challenging to obtain. It can take years to train people with the necessary skill sets, and then they can be easily enticed away from Canada by the lure of better opportunities coupled with higher salaries. If Canada is to invest in people and skills, then it must also invest in creating the environment to attract and retain them.

Research Teams

Over the past decade, there has been a pronounced shift towards the hiring of computationally based researchers at Canadian universities, particularly in the sciences and engineering. The new generation of Ph.D. graduates is computer literate, and many have a strong computational research focus. For example, bioinformatics, computational finance, nanotechnology and photonics represent some of the hottest research areas today and for the foreseeable future. The HQP in these areas are scarce and go to the highest bidder. To complicate matters, the recruitment of these professionals by research institutions is also being met with stiff competition from industry. Universities nonetheless recognize that computing skill sets are fundamental to future successes in research and development, and are establishing the strong computationally based research programs essential to Canada's competitiveness in the 21st century.

Leading this research and training renaissance in Canada is a growing pool of talented professors, ranging from senior Canada Research Chairs with outstanding international reputations to junior assistant professors with tremendous research potential. They have confidently accepted positions at Canadian universities in part because of Canada's excellent track record in advanced infrastructure investments (through CFI and supporting provincial programmes) and because of the favourable research environments that these investments provide.[6] These professors are essential to training the next generation of skilled professionals and researchers, and to creating acceptance by Canadian industry of computationally based applications.

They are clearly pivotal to advanced HPC based research in Canada and to the enhancement of Canada's research and development capacity and reputation.

Computationally based research, which is conducted by graduate students and postdoctoral fellows (PDFs) working under the guidance of their professors, is an important vehicle for the training of highly qualified personnel. Following completion of their studies, these individuals are in high demand by industry, government and academia. The development of HQP and the subsequent movement of these people among organizations and sectors constitute the most effective form of technology transfer. An insufficient supply of HQP, in many cases based on inadequate resources for training, has been identified as a significant bottleneck in developing local, regional and national HPC based innovation systems. To get the necessary training, these ambitious and energetic individuals need access to world-class HPC facilities, help from highly trained experts in computational techniques, and experience in applying these techniques to concrete problems in support of both research and development activities. Expertise is not an incremental benefit. These highly qualified women and men will be the crucial resource needed to enhance research and manufacturing productivity and to accelerate discovery and innovation. Canada's future opportunities are in the hands of these experts.

> *"Mankind is experiencing a data explosion driven by growing data resolution, networking providing data access at a world scale, and an exponential growth in the amount of data created. Companies such as YottaYotta are building technologies that must scale at rates that match this hyper-exponential growth. Hiring development staff from the HPC community provides essential background in parallel computing, efficient data transmission, and multi-node data management autonomics. Partnering with the HPC community allows companies such as YottaYotta to test technology today that will be mainstream tomorrow."*
>
> ***Wayne Karpoff***
> ***Co-founder and Chief Technology Officer***
> ***YottaYotta, Edmonton***

Our country has many world-class researchers who now rely on HPC to make innovative discoveries at the international level. It often takes a decade or more to build strong research teams and have their technology evolve from ideas into commercial products. It takes at least as long to train skilled workers with advanced research and development capabilities. The creative sparks ignited in today's graduate students using the current CFI-funded infrastructure will not realize full potential for Canada unless these students have access to appropriate resources throughout their studies and into their academic, public sector and industrial careers. Long-term support for HPC infrastructure and human resources will allow Canada to derive the maximum benefit from its past and present funding initiatives. The universities play a crucial role in providing computationally based research and training that is an essential component of their long-range research plans.

Research Programs

HPC-related research may be categorized in two areas:

(i) *Applications development*. Projects in this area are domain-specific applications that use the performance gains achievable with more powerful computing resources to create new insights. This is usually characterized as *research using HPC*.

(ii) *Tool development*. In this area, new tools and technologies are created to facilitate the use of the HPC resources. This is usually characterized as *research into HPC*.

Canada is already an acknowledged leader in computationally based applications development – research that uses HPC – with strong research programs

---

[6] Although it is difficult to quantify the effect, there is strong evidence to support the conclusion that in HPC-related areas, a combination of initiatives - including the Canada Research Chairs program and the Canada Foundation for Innovation infrastructure grants - has been successful in slowing and in many cases reversing the so-called "brain drain."

across many disciplines. These strengths, however, are not balanced by correspondingly strong HPC tools development research (essential technologies such as data-mining tools, meta-computing, performance evaluation tools, advanced visualization, operating systems and automatic parallelization). Expertise in these foundational skill sets and technologies must be developed and expanded to fully capitalize on the physical and human investments being made in the computing infrastructure. We must continue to build on our capacity both in research using HPC and research into HPC, strengthening existing areas and facilitating new ones, if we are to be competitive in the global research arena.

The Mandate of the Long Range Plan for Canadian HPC is to support research infrastructure, and not to support research programs directly. While we acknowledge that funding research teams is critical to realizing the full benefits of the infrastructure, no funding is proposed to meet this need. This is best left to the granting councils (NSERC, SSHRC and CIHR) to address through their well-established and respected peer-review systems. This leads us to the following recommendation:

3.1 We strongly recommend that the granting councils, in consultation with the universities, develop funding opportunities which strengthen research programs both in the foundational technologies of HPC and in HPC application development.

Support Personnel

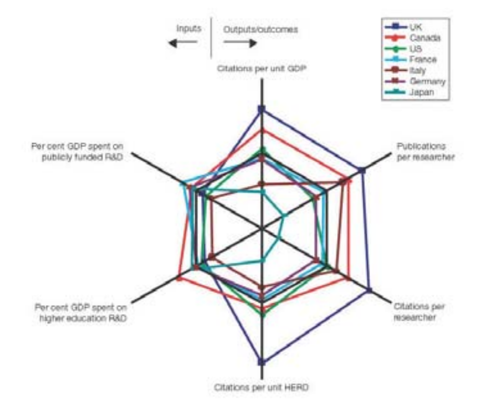
Acquiring a computer - regardless of the size – is only the beginning of a user's commitment to the resource. Users are painfully aware that computers must also be maintained: there are hardware faults to be corrected, new versions of the operating system and software packages to be installed, data backups to perform, networking, security and so on. Computer and systems maintenance represents substantial time and resource commitments, and these are magnified as these systems become larger and more complex. A variety of skilled personnel and support roles are therefore essential to the effective operation and maximum exploitation of any HPC facility. The skills and experience needed are extensive, including (i) managing, operating and maintaining a facility, (ii) training and assisting researchers to make the best use of its resources and capabilities, (iii) ensuring maximal productivity of the HPC sites by, for example, checking that software is run on the most suitable computing platform and reworking code to achieve significant performance gains, and (iv) helping to create new applications in support of innovative research initiatives.

System Administration and Operations

Systems administration and operations are primarily concerned with the day-to-day care of the HPC hardware and software infrastructure. The

Scientific productivity

There have been numerous studies consistently demonstrating that Canadian researchers are among the best in the world. A recent study in the journal Nature assessed the research productivity (scientific outputs and outcomes) of the G8 countries (except Russia) as a function of the amount of money invested in research (financial inputs). For the six metrics measured, the data has been normalized to the average of the seven countries. HERD stands for Higher Education funding of Research and Development. The data shows that Canadian researchers are second in individual productivity (citations per researcher and publications per researcher). Additional data from the study shows that the average United States researcher is only 75% as productive as a Canadian researcher. From the financial point of view, Canada is a making significant investment in research, most notably as a percent of GDP spent on HERD. [REF 22]



Technical analyst support Program (TASP)

TASP people operate many of the Access Grid rooms in Canada, allowing researchers across the country to communicate visually and to "share" applications in real time.

TASP provides Canadian researchers with vital help in operating advanced high performance computing systems, in HPC application development, code analysis and optimization, parallel code design and visualization expertise. This program is run by C3.ca, and is funded by an NSERC Major Facilities Access grant. The funding provides for 19 technical support people and one manager. The support people are distributed across the country at various universities, ensuring physical proximity to the user community. The TASP personnel run the largest virtual laboratory in Canada and one of the largest in the world!

TASP has successfully overcome the many daunting challenges of supporting a large, diverse user community geographically separated by thousands of kilometers. The program has achieved considerable success by

- making limited HPC support available to many researchers in Canada;
- providing advanced technical support, coast to coast, in a heterogeneous HPC environment with different hardware vendors, operating systems and system applications;
- organizing numerous HPC-related workshops each year;
- conducting outreach programs for the academic and industrial research communities; and
- working together as a team to realize the CISS initiatives (see the CISS sidebar on page 65).

Despite its success and its impact on the Canadian research community, the TASP program is not able to meet existing demands and is continually losing skilled and experienced personnel. This situation arises primarily from a lack of sustained long-term funding. As funds are dependent on annual applications for renewal, all contracts for TASP personnel run for only 12 months and provide no long-term job security. In addition, the funding for the program has remained fixed for almost a decade despite the huge growth of HPC activities in Canada since 1999, resulting in a stressful increase in workload for the TASP team.

supporting personnel ensure the proper functioning of HPC facilities, providing systems management and operations support. Specific tasks include installing and maintaining operating system(s), performing updates and patches, managing file systems and backups, and ensuring the integrity and security of the user data. These activities are crucial to ensuring that the system is fully functional and available to the community.

Programmer/Analysts

The role of programmer/analysts is to provide specialized technical assistance to researchers, to conduct workshops and training, and to evaluate and implement software tools to make effective use of available resources. HPC hardware typically operates at a sustained rate well below the theoretical peak performance of the system; this is usually due to a lack of parallelism in parts of an application. A creative team of programmer/analysts can double that rate through code optimizations, algorithm re-design, enhanced cache utilization and improved data locality. The added value from such activities can be huge, and can correspond to *twice* the science delivered for the same hardware. These skills can thus dramatically increase the scientific productivity of the research community. A recent review of the TASP program of C3.ca (see the sidebar on this page) cited one example of the benefits thus achieved: ". . . the efforts of the TASP member had the same result as purchasing ten times more computing resources." By allowing researchers to run their applications faster, analysts support researchers and their students to do better science.

In extreme cases, the support team may be responsible for enabling research that would not have happened without their work. It is therefore extremely important that researchers and their science are matched to the appropriate HPC facilities and personnel resources within a national support model. Examples of this include the highly visible Canadian Internetworked Scientific Supercomputer (CISS) project, where computing resources across the country are combined to give selected high-quality research proposals access to massive computing resources (see the sidebar on this page).

Applications Programmers

Frontier science requires world-class software applications. While much of the development of new scientific functionality is traditionally carried out in a researcher's own laboratory, HPC applications programmers often make valuable contributions to this work by virtue of their own scientific, numerical or visualization experience. The additional skills of the support staff, when matched with researcher needs, often play an integral role in enabling ideas, concepts and advice to flow with greater ease in the subject domain of the scientist. This support has the additional benefit of greatly reducing what is normally a challenging startup period for researchers

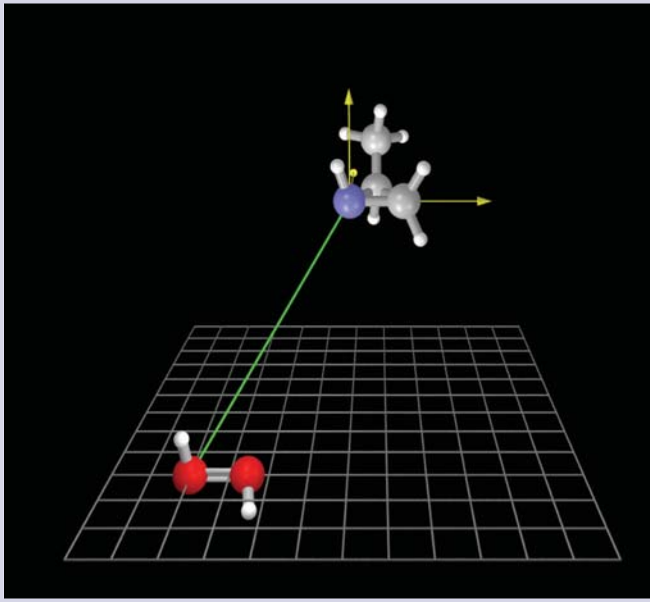
Canadian Internetworked Scientific Supercomputer (CISS)

Investments by the Canada Foundation for Innovation have resulted in over $250 million in high performance computing equipment being installed in Canada. This equipment is distributed across the country at over 20 sites. However, some research problems require access to computational resources that exceed those of any single site in Canada. The Canadian Internetworked Scientific Supercomputer (CISS) project [REF 23] is attempting to address this shortcoming by connecting much of this computing capacity together into one giant virtual supercomputer.

The CISS infrastructure is built using software developed by the Trellis project [REF 24] at the University of Alberta. Trellis provides a simple interface that allows an application to run on collections of different types of computers, possibly running different operating systems, at any HPC site in Canada.

There have been three CISS experiments run to date in computational biology, chemistry and physics. CISS allows these applications to accomplish several years of computing in a single day, greatly accelerating the discovery process. For example, CISS-3, held in September 2004, harnessed over 4000 processors across Canada, resulting in 22 years of computing being done over two days!

The CISS enables the research community to tackle research problems that are considerably larger than anything that could be tackled at a single HPC site in Canada.



On the human scale, a handshake is possible between two left or two right hands, but not between the left and right hand. On the molecular level, there is an energy difference between the interactions of the right-handed and left-handed forms of a molecule. CISS-1 did an analysis of this phenomenon.

TASP Case studies

Here are two examples of how TASP analysts can dramatically improve research productivity and increase the value of computing hardware investments.

Example 1: The SHARCNET HPC consortium in Southern Ontario has a heterogeneous computing environment to support the varied needs of its diverse user base. A TASP analyst was involved in helping to migrate users from *idra*, the principal system for large-scale parallel applications, to *cat*, a Xeon/GigE cluster with faster CPUs. A user complained that to their surprise their software ran much slower on cat (with 2.4 GHz Xeon processors) than on idra (with 0.67 GHz Alpha processors). A TASP analyst was called in and discovered that in addition to CPU speed, there was another difference that could affect performance: the size of the cache. The analyst was able to identify where the cache size influenced the program's performance, and then rewrote the offending code. When the analyst was done, the program was running 10 times faster!

Example 2: A University of Manitoba physicist had serious application performance problems and turned to the TASP program for help. The physicist's sequential program was taking too long to run, even on small tests. This meant that the researcher could not run parameter sets that were large enough to get publishable results. The TASP analyst parallelized and tuned the program. A few days later, the physicist sent the following email:

I ran the program for one of my input data and compared the computational time as follows: Before: elapsed time is about 7 days Now: elapsed time is about 9 minutes This is a huge improvement. Now I am running the program for new input data and waiting for the results. Thanks a lot for your time and kind help.

The improvement was 1000 fold! The researcher then submitted his big job, which ran for six days. Before the TASP analyst became involved, this program would have taken roughly 15 years to complete.

These two examples exemplify the important work done by TASP analysts. Researcher cannot wait for 15 years for a result. A consortium may get higher value from their hardware investment by a sustained investment in the skilled people who understand the hardware and software issues needed to improve application performance The contribution of TASP to the scientific community can be measured by substantial increases in research productivity, by a greater capacity for competitiveness and by dollars saved.

learning to work with HPC. This skill set is imparted to students and postdoctoral fellows as well, giving them both the scientific knowledge and the programming experience necessary to create new computational methods and applications in their various fields, eventually leading to dramatic new insights.

Data Management and Visualization Personnel

The importance of versatile analysis and visualization techniques for simulation work is now self-evident, and both computation and visualization activities are increasingly being driven by "science pull" rather than "technology push."[7] The most challenging aspect of data management and visualization is coping with the massive datasets that are being produced. Simulations in climatology, bioinformatics and astrophysics, for example, regularly produce data sets that are hundreds of terabytes or even petabytes in size. Entirely new techniques and computing resources will be necessary to cope with them: in most cases, interactive visualization is the only practical way to glean insights into these datasets. In addition, the effective exploitation of such volumes of data will require a major development effort in distributed computing across high-speed networks (grid computing). This requires the training and the retention of personnel able to

---

[7] Science is increasingly driving the rate of technology development, rather than merely tracking it.

manage the data resources and to develop the new tools and techniques required to visualize them.

Management and Administration

Administrators and staff are needed to make resource allocation decisions, manage human resources, develop and implement training and outreach programs, handle media relations, support nationally coordinated projects like CISS, and perform everyday business processes. A technical manager such as a chief technology officer (CTO) would take responsibility for the management of the programmer/analysts. The management team would be responsible for prioritizing tasks, dealing with the far-flung nature of the facilities and fostering a sense of teamwork and common purpose. Its mandate would be to forge a geographically scattered set of individuals and skills into a coordinated national resource, one that is available and effective for researchers regardless of their location or discipline. The team would also be responsible for planning and implementing the training and professional development of the analysts themselves—crucial to the overall goal of attracting and retaining HQP.

Training of Researchers

The overall goal of HPC support is to proactively support the needs of a wide variety of researchers by engaging them through their entire HPC lifecycle. In this context, support staff must (i) create awareness of the resources available and the potential for these resources to accelerate research productivity; (ii) provide usage instructions and courses (on topics such as parallelism, programming tools and performance analysis); (iii) help users find the right match between their application and the available technologies; (iv) maintain the HPC facility; and (v) develop new tools or tune existing ones (hardware and software) to optimize the use of HPC resources and the applications running on them. This type of support will enable the researcher to obtain more results in the same unit of time or the same results in less time, thus freeing up facility time for others (see the sidebar on page 66).

HPC support staff is essential for training Canada's next generation of scientists and engineers in the use and improvement of HPC resources. Interactions of HPC support staff with graduate students and postdoctoral fellows will provide a fertile training ground to develop the next generation of researchers, giving the new scientists a grounding in HPC as part of their disciplinary training. Much like today's researchers use personal computers to support their work, the next generation of researchers will rely on HPC resources to improve their research outputs. To do so effectively, researchers will need appropriate training. This leads us to the following recommendation:

3.2 We strongly recommend that universities develop training opportunities at the fundamental and advanced levels to enhance HPC skills in the research and development environment.

Implementation of this recommendation will increase the versatility of Canada's students and researchers, and thus our academic and industrial competitiveness at international levels.

Funding for HPC Support Staff

The current funding of HPC support staff in Canada is very low. In the first two rounds of competition, CFI provided no funds to operate the HPC facilities it approved. In the next two rounds (2002 and 2004), an amount equal to approximately 30% of the CFI's capital award was made available to help operate the facilities. Most provinces have programs that will match CFI capital awards, but few have programs to provide operating funds. Most of the new HPC installations have had to operate on shoestring budgets. It is unfortunate that Canada invests, for example, $10 million in computing hardware, and then has to look hard to find $100,000 to operate it. Capital and operating funds must be linked more closely and realistically.

In contrast, major HPC installations around the world have chosen to make substantial investments in HPC support personnel (see the sidebar on page 68). Their human resource investments typically range from 20% to 50% of the capital investments.

Support staff funding: Canada versus the world

The table below details the level of support staff funding for five major HPC facilities. The National Energy Research Scientific Computing Center (NERSC, USA), the Pittsburgh Supercomputing Center (PSC, USA) and the HPCx (Great Britain) show the level of support that is typical for World Top-50 installations. All three are national facilities available to the general research community. Environment Canada (Dorval) provides a Canadian perspective on a production HPC facility with a targeted application base. The WestGrid facility (Vancouver) shows the typical level of support at a CFI-funded facility, where operating funds are in scarce supply. Compared to similar leading-edge HPC installations around the world, the Canadian facilities have a severe shortage of supporting personnel.

The significant change in the center ranks between June 2004 and November 2004 demonstrates how competitive and rapidly growing the HPC world is. The WestGrid HPC research facility in Canada is limited in its capability to follow this trend by the small number of its support staff. The staff typically operates in a reactive mode, responding to day-to-day needs. This represents a lost opportunity: an investment in a multi-million HPC infrastructure requires a significant investment in support people to maximize the research benefits obtained. To reap the full return on the venture in a fast evolving computing environment, the support staff must be numerous enough to operate in proactive mode, engaging the research community to assist them in making the applications run effectively as soon new computing hardware is available. This fundamental principle is recognized by all the organizations represented in the table.

Category	**NERSC (USA)**	**PSC (USA)**	**HPCx (UK)**	**Environment Canada**	**WestGrid* (Canada)**
Rank in the Top 500 (June 2004)	14	25	18	54	38
Rank in the Top 500 (November 2004)	21	34	27	74	54
Availability (hours/days)	24/7	24/7	24/7	24/7	10/5
Operations Staff	9	6	4	12	2
Systems Support	11	10	2	7	1
Networking and Security	6	3	1	5	0
User Support	15	11	13	15	0
Total Staff	**41**	**30**	**20**	**39**	**3**

* Data is for the UBC site.

Staff members are usually all at a single location, working on a small number of computing systems. These organizations see investing in people as a way of maximizing the return on their investment in the hardware.

A failure to provide the necessary level of support funding will have a broad negative impact on the research being done on HPC facilities in Canada both today and in the future. Not only will the growth in awareness (and thus users) of HPC slow down, but the high start-up costs associated with utilizing HPC facilities that have insufficient support staff (such as training and software conversion) would preclude many researchers from establishing research activities in this area.

In Canada, over $250 million has been invested in HPC infrastructure. The research community using these resources numbers over 2000 professors, graduate students and research associates. In 2004, the staff supporting HPC facilities totalled 43 people—spread out over 30 institutions and dozens of computing installations. A failure to provide the necessary level of support funding will have a broad negative impact on the research being done on HPC facilities in Canada both today and in the future. Not only will the growth in awareness (and thus users) of HPC slow down, but the high start-up costs associated with utilizing HPC facilities that have insufficient support staff (such as training and software conversion) would preclude many researchers from establishing research activities in this area. The Technical Analyst Support Program (TASP) shows what can be accomplished by adding support staff (see the sidebar on page 66). Statistics showed that TASP catalyzed significant growth in the number of researchers involved in HPC: one consortium reported an increase from 40 users to over 200 users as a result of TASP initiatives.

In addition, issues that are currently handled by the existing support staff (such as hardware and software malfunctions in the vendor equipment) will fall by the wayside as other activities take priority in response to reductions in personnel or increase in workload. This will dramatically affect how science is done in Canada, and lessen the potential positive impact of existing and new HPC facilities.

As a cautionary note, there is currently a critical shortage of people with skills to solve problems in HPC infrastructure. Attracting and retaining highly qualified personnel who know how to use and maintain HPC facilities, and hence advance HPC as a discipline in its own right, will require stable financial support. These people will be necessary to develop world-class software applications and to assist researchers in the implementation, analysis and visualization of their computational experiments.

While there is a sizeable need for these people to work in university-based HPC facilities, there is also a substantial and growing need for them in industry. As HPC becomes an ever greater part of research and technology transfer, industry will be continually required to adopt and support new HPC tools and applications if it is to remain internationally competitive. It will require a pool of talented and skilled individuals that are capable of traversing the difficult path from research to commercialization.

The TASP model shows that the development of a nationally organized group of HPC specialists spanning the entire country provides a much broader and efficient pool of expertise than would normally be available if all the different technicians were working separately. The cost effectiveness of this support model is also attractive. This model supports a much higher researcher / support analyst ratio than exists, for example, in the United States. However, there is a need for much more support than is currently available. This leads us to the following recommendation:

3.3 We strongly recommend that an amount equivalent to at least 25% of the annualized HPC hardware expenditures be allocated annually to fund HPC support personnel (adopted when establishing the funds required for Recommendations 2.1 and 5.1).

While there is a sizeable need for these people to work in university-based HPC facilities, there is also a substantial and growing need for them in industry. As HPC becomes an ever greater part of research and technology transfer, industry will be continually required to adopt and support new HPC tools and applications if it is to remain internationally competitive.

A cutaway section of a dynamic virtual cell simulation representing two million biomolecules.

"At this time, there exist significant challenges involved in providing the synchronous execution of a tightly coupled application for cell life simulations. Unacceptably high performance losses are a result of the limited communication rates and computational resources. Our current research takes a week of computer run time to generate a 2 millisecond simulation of cell life. The model relies heavily on a large number of tightly coupled computational processes, operating in parallel and maintaining a high level of close communication. This is at a reduce scale of 1/100th and using 100% of our current computing power. New computing resources are critical to optimize efficiency and yield significant impacts on the resolution achievable by cell simulation."

Dr. Mike Ellison
Professor of Biochemistry,
University of Alberta
Executive Director, Project Cybercell

Simulating the Living

Imagine being able to have your prescription medicine tested on your own cells without experiencing painful side effects. What if your physician was able to more accurately prescribe medication suitable for your physiology without trial and error testing?

Through recent research development in high performance computing, researchers in Canada are able to simulate a living cell accurately within the virtual environment of a computer. This virtual living cell is one that can be manipulated at different levels of molecular resolution, and it can respond, adapt and evolve to exploit this virtual environment.

The prospect of simulating life has profound implications for life science research. Discoveries could be made with greater predictive insight, more quickly and at less expense. Pharmaceuticals and other biologically active compounds could be screened and tested not only for their targeted effects but also for their physiological side effects. Eventually, complex drug regimens could be tested and optimized for patients using personalized physiological simulations. Organisms could be genetically engineered for optimal performance in biomanufacturing. The potential applications are extensive and far-reaching in their impact.

Through recent research development in high performance computing, researchers in Canada are able to simulate a living cell accurately within the virtual environment of a computer. This virtual living cell is one that can be manipulated at different levels of molecular resolution, and it can respond, adapt and evolve to exploit this virtual environment.

For the effective deployment of a virtual cell model, improved computing resources must be available to research teams across Canada. Currently, virtual cell models are being carried out at 1/100th of the scale that is needed. To begin to model more than a millisecond of simulated cell life, the computing requirements are enormous and require at least 8 terabytes of memory!

More processing power and networked computing nodes will be imperative for the success of this research. Simulations will continue to increase in complexity. Techniques for visualizing complex data are rapidly becoming more than simply a means of communicating results. These tools have become invaluable to the continued evolution of the model design and critical to the rapid diagnosis and evaluation of model behaviour.

Nanotechnology: To See a World in a Grain of Sand

Imagine environmental robots cleaning out our landfills and toxic dumps keeping our landscape clean. Imagine having a missing piece of DNA inserted in your genes to prevent cancer. Imagine miniature robots able to work inside your body cleaning your blood. Imagine never having your car rust again.

And this isn't science fiction. This is the potential of nanotechnology research. Nanotechnology is research and development at the atomic or molecular level in medicine, biotechnology, genomics, manufacturing, computing, information technology, communications and other fields. It is manipulating matter and creating devices at the nanometre level—one billionth of a metre.

High performance computing capacity is a huge limiting factor on the practical advancement of nanotechnology. In the nanoscale world, the complexity of problems increases exponentially with the number of atoms involved. When the number of atoms being simulated increases by a factor of ten, the number of interactions to be considered increases by a factor of 100 or more. The length of time over which a nanoscale interaction can be observed is also severely constrained by current computational limits.

These questions are swirling in the semiconductor industry. In the past, industry operated under the assumption there was very definite limit to the number of components that could be fabricated onto a semiconductor wafer. Nanotechnology is dismantling these absolutes. Decreases in size are a part of the equation; increases in speed are the other part. The potential will only be realized with corresponding HPC infrastructure.

Nanotechnology demands unprecedented enhancements in HPC capabilities to support this new scale and scope of research. There is a major demand for enormous computer power: teraflops or petaflops computing capability. This will allow our researchers to explore simulations that involve a realistic number of atoms interacting for longer than a second or two. This is absolutely central to building nanodevices and ultimately creating new and better manufacturing processes, homeland security systems, biomaterials, disease diagnostic tools and environmental cleaning agents, all of which require computer power exceeding teraflops and petaflops computing capacity. The setup costs are substantial, but the payoff even greater – ultimately cutting costs while providing more power, capacity, flexibility and reliability.

> *"In the long term, nanotech is all about computing. It's the only tool we have to bridge the chasm between theory and experiment for hard problems that resist analytical treatment. Which, of course, means almost all problems . . ."*
>
> ***Dr. Mark Freeman***
>
> ***Canada Research Chair***
>
> ***in Condensed Matter Physics***
>
> ***Professor of Physics, University of Alberta***

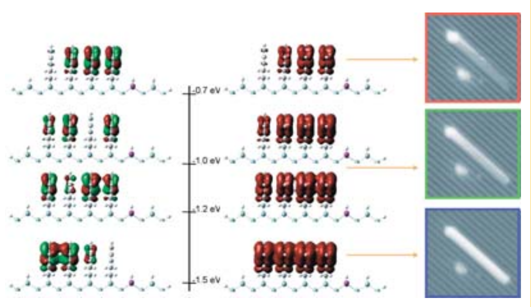
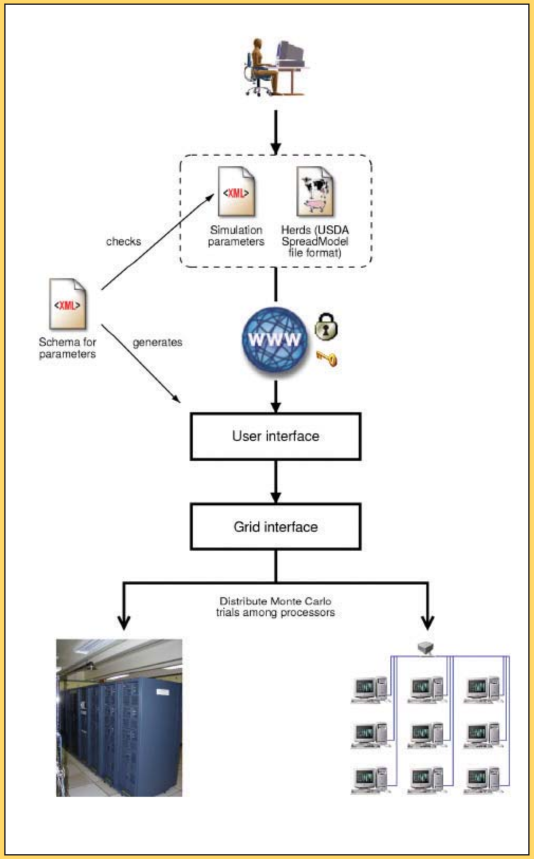


Figure showing correspondence between calculated molecular orbitals (left), calculated total charge density (middle), and "slope" effect in molecular lines observed via scanning tunneling microscopy imaging (right). Quantum mechanically predicted slope disappears with increasing magnitude imaging bias, in agreement with experiment. This phenomenon illustrates how conduction of electrical current through molecules is modulated by a localized negative charge and provides a unique insight into molecular electronics.



The simulation process for animal disease modeling. From the user at the desktop to the modeling algorithms running on HPC computer clusters. Modeling permutations are a complex equation and require large amounts of computing power.

Modelling Infectious Disease in Animals

Foot-and-mouth disease (FMD) is a highly contagious and economically devastating disease that affects cattle, pigs, sheep and other cloven-hoofed ruminants. Many affected animals recover, but the disease can cause severe losses in meat and milk production. FMD spreads widely and rapidly. This combined with the grave economic and physical repercussions makes FMD one of the diseases livestock owners fear most.

Much of FMD disease in susceptible populations is understood by identifying the factors that influence the frequency with which the disease occurs. These factors include: virus, host, host/virus interactions, and the environment. However, each of these factors has so many permutations and combinations that high performance computers are needed to assess how diseases spread and hence allow for the implementation of containment strategies.

Modeling isn't a new tool in epidemiology. However, HPC has raised the bar for animal modeling research and is now an integral part of the science. It has acted as a catalyst, prompting scientists to ask new and more sophisticated questions to help provide industry and government with the tools they need to contain the spread of animal disease.

"With high performance computing, researchers will be able test out increasingly more complex models, with improved speed and larger capacities for better understanding of how disease spread models work and how they relate to biological reality. These results play a crucial role in helping animal health professionals design outbreak containment strategies. The opportunity to work with animal health researchers from Canada and the United States and to expose them to the potential of cluster computing is extremely exciting."

Dr Deborah Stacey
Professor of Computing and Information Science
University of Guelph

Animal modeling research looks at a future where animal disease outbreaks can be better managed and contained where policy makers can best decide to implement vaccination strategies to minimize disease spread. It's not a matter of if an animal outbreak will occur; it's simply a matter of when. Animal disease modeling can give scientists and governments the tools to respond quickly and appropriately, minimizing physical and economic damage. It is also an essential tool for agricultural and public health policy development for analyzing and planning strategies to monitor and combat bioterrorism. HPC is essential for performing these studies and improving our ability to minimize and contain outbreaks that could have devastating impact on the Canadian economy.

Simulation Providing Stimulation for Cardiac Patients

High performance computing is helping biomedical engineers reduce the costs and time involved in developing new life-saving devices, such as implantable defibrillators.

Complex computer models of the heart are being developed to provide an early cost-and-time-saving step in testing new medical ideas and innovations. The medical world has heuristic models about how things work. Computer models are more rigorous. Simulation experiments allow researchers to verify hypothesis in a quantitative way. If the ideas make sense in computer models, then there's a good reason to go ahead with the medical experiments.

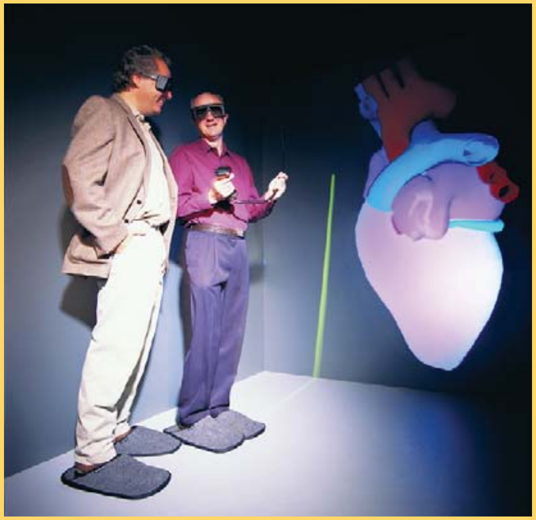
A typical heart simulation takes a day to run on high performance computers. The same test might take a year or more to perform on a desktop computer. However it's doubtful a modern desktop could handle the complex computer programming demanded by the heart models. Heart models are extremely complex. They're built from the sub-cellular level up. This work entails creating millions of 'virtual' cells, linking them together to build tissue and then layering the tissue to form parts of the heart or a whole heart.

This heart simulation capability through HPC will enable powered simulations to improve implantable defibrillators. An implantable defibrillator is a small device implanted under the skin containing a sophisticated microprocessor and a battery. Thin wire electrodes are threaded from the defibrillator into the heart. The device constantly monitors the hearts rhythm, so when a cardiac patient has an episode of sudden cardiac death, the defibrillator delivers a strong electrical shock to start the heart pumping again.

This is a huge step forward in cardiac patient care and monitoring. These devices will literally save patient's lives. Thus high performance computing is playing an essential role in cardiac research directed at saving the lives of Canadians.

"We have a basic hypothesis, based on the dynamics of fibrillation, about improving how the device delivers energy. This will be a major improvement in the safety of the devices and overall heart health of cardiac patients. The simulations that allowed us to arrive at this hypothesis wouldn't have happened without high performance computing resources."

Dr Josh Leon
Department Head of Electrical
and Computer Engineering
University of Calgary



A virtual heart model containing millions of "virtual cells". Researchers can investigate the virtual heart through the use of a 3D cave shown here. This allows for heart investigations to go from viewing the surface cells of the heart to viewing all cellular and tissue levels when monitoring heart activity.

Chapter 4: IMPACT Canada: A National Body for HPC

The effective coordination of people and resources has been a cornerstone of the success of the Canadian HPC community. The current Academic HPC consortia provide a highly effective model for regional coordination and sharing. The C3.ca Association has provided a crucial forum for developing a national strategy for HPC organization/development, but is not configured to provide all the leadership that is needed. As HPC becomes more prevalent, and as systems become increasingly available on our high-speed national and provincial networks, the importance of effective national coordination assumes central significance[8]. This level of coordination and sharing of personnel and computing infrastructure is presently perhaps unique to Canada – thanks in large part to CFI. Canada will gain far greater value from further coordinating its national HPC investments than if these are managed in isolation of each other. Canada will also become much more competitive internationally under this model/paradigm which is being adopted, for example by the European Union[9], and could find itself in a leadership role with respect to the innovative development and coordination of highly geographically dispersed national HPC research infrastructure. To take on this essential leadership and national coordination role we propose the immediate establishment of a new agency, IMPACT Canada.

4.1 We strongly recommend the establishment of a body, IMPACT Canada, to provide national leadership, coordination, and oversight for HPC in Canada.

The primary mandate of this body will be to provide national leadership and coordination by activities such as: (a) advocacy with respect to funding, particularly federal, (b) representation of the Canadian HPC communities nationally and internationally, (c) defining required HPC services, functionality, capabilities, and capacities, (d) coordination and interoperability issues among the major resource providers, (e) coordinating and facilitating Canadian training and education relevant to HPC research and HPC research users, (f) leading public awareness, publicity and outreach communications on behalf of the HPC communities and (e) the annual HPCS conference. In particular reference to the other parts of this report, this would result in an emphasis on:

- creating strategies for enhancing HPC in education and research in the health, environmental, applied, physical, life, and human sciences as identified in recommendation 3.2 by developing an awareness of HPC resources throughout the public and private sectors;
- assisting in the development of expertise identified in recommendation 1.2, including working with post secondary institutions to develop a critical mass of skilled personnel, and encourage computationally based research in Canada;
- advising on the allocation and distribution of resources needed to maximize the competitiveness of Canadian HPC-dependent research;
- assessing future development needs to ensure that Canada's HPC investment remains competitive for the future;
- facilitating the selection and operation of Canada's tera-scale facility;
- enhancing the interface between industry and HPC resources, and connect industry to researchers and research opportunities; and
- providing broad accountability for the expenditure on HPC

---

[8] Already the European Commission Support for Research Infrastructure's Programme has launched HPC-Europa, an initiative that supports European scientists by integrating networking, joint research, and sponsored access to six major HPC infrastructures (CEPBA, CINECA, EPCC, HLRS, IDRIS and SARA). [REF 25]

[9] The € 6.27 billion European Union's 6th Framework Programme (FP6) (2002-2006), designed to compete with US and Japanese research efforts, has a budget of € 3.6 billion to support its "Information Society Technologies (IST)" theme. One of this theme's strategic priorities is to provide Grid computing capacity and middleware for complex problem solving, and support for collaborative research and development opportunities, in Europe. The framework also provides for another € 655 million for the development of a pan-European grid-based fabric of high-speed communications and computing networks. [REF 26]

In this context, IMPACT Canada will foster **I**nnovation, facilitate the **M**anagement of advanced computing resources, enable **P**eople to reach their research potential, assist in the development of the computer **A**pplications that lead to scientific breakthroughs, foster the movement of ideas and interaction through broad **C**ommunication strategies, and oversee and coordinate the acquisition and operation of the enabling **T**echnologies. IMPACT Canada will thus capture the essential elements of the well-connected, highly competitive, digital world that will dominate 21st century Canada. The following sections outline specific strategies and functions that IMPACT Canada should undertake to complete each major aspect of its mandate and makes subsidiary recommendations related to these activities.

Create strategies for enhancing HPC in education and research in the health, environmental, applied, physical, life, and human sciences as identified in recommendation 3.2 by developing an awareness of HPC resources throughout the public and private sectors.

IMPACT Canada will support coordinated research efforts amongst disciplines. In this role the agency will work to develop and strengthen the connections between, and the experiments of, national science initiatives that are beginning to emerge amongst Canada's HPC Consortia. For example, the *Canadian Internetworked Scientific Super Computer* (CISS) project [REF 23] links together computers nationally for targeted needs. *IMPACT Canada* will also provide information to researchers as to how HPC resources are being used and how to access them through its website.

IMPACT Canada will also provide technology outreach, training and education to researchers and staff, assist schools with program development. IMPACT Canada will work with other organizations and all levels of government to foster professional development programs for teachers, from pre-school to high school, to help them better teach how computational science models and solves real-world problems. These programs will introduce teachers to computer modelling and simulation, and improve their knowledge of the use of information technology in the classroom. Teachers in turn will use their new skills to inspire their students to become excited about mathematics and problem solving within the core subjects of the curriculum. Furthermore, teachers will be assisted to arouse the interest of their students through problems of local or regional interest. This objective will complement the CRYSTAL (Centres for Research in Youth, Science Teaching and Learning) initiative of the Natural Sciences and Engineering Research Council of Canada. Success in this key objective is crucial for the long-term economic health of Canada.

Assist in the development of expertise identified in recommendation 1.2 including working with post secondary institutions to develop a critical mass of skilled personnel, and encourage computationally based research in Canada.

IMPACT Canada will also work to develop a network of industry and public partners interested in both participating in HPC initiatives and leveraging Canada's HPC resources. IMPACT Canada through its own initiatives, and in cooperation with other organizations, will develop outreach programs to complement the formal education and training programs within Canadian schools and universities. For example, the potential benefits of visualization tools and collaboration tools, such as the Access Grid (a set of protocols for advanced collaboration), to educators and to Canada's *small and medium* sized enterprises (SMEs) could be promoted through cooperation with the National Research Council of Canada. Furthermore, IMPACT Canada will work with the media to ensure that the Canadian public is aware and well informed of emerging technologies.

Without a critical mass of qualified personnel, Canada will be unable to effectively utilize, develop, and support HPC activities of international stature. We therefore strongly endorse the following actions regarding Education, Outreach, and Training: (i) Universities, consortia and IMPACT

Canada should work together to create and deliver education, outreach, and training programs that will assist in the effective exploitation of all HPC facilities; that (ii) IMPACT Canada encourage and assist the development of apprenticeship programs to give talented students the opportunity to obtain hands-on experience with HPC applications and HPC operations, with a view towards training the next generation of HPC researchers and support personnel; that (iii) Access Grid, or similar technologies, be used as a means of actively fostering the collaborative and educational aims of IMPACT Canada; and that (iv) IMPACT Canada undertake outreach programs with universities that will promote understanding of the role of computationally–based research in all levels of education.

IMPACT Canada will play a key role in education, outreach, and training. It will work closely with universities, hospitals, colleges, and consortia to deliver to young Canadians, via the World Wide Web, an early introduction to computational science. This is widely recognized to be a vital component of science education in general, and such content provision is very well funded in the United States. IMPACT Canada will develop innovative outreach programs to build awareness in the minds of our next generation of young researchers of the central role that computing can play in both learning and research.

IMPACT will be Canada's principal instrument for facilitating the development of human resources in all facets of HPC through education, training and outreach. As noted earlier, a strong cohort of talented HPC researchers exists within Canadian universities, many of whom hold Canada Research Chairs and are international leaders in a broad range of fields. Our leading researchers attract excellent students and produce graduates who in turn are poised to be the leaders of tomorrow. Despite the tangible benefits that have accrued from the infusion of funding and infrastructure through the granting councils and new initiatives such as the CFI and the CRC program, the status quo is not sufficient to maintain Canada's position among the leading nations of the world. Indeed, in order to rank among the world's most innovative countries, much more must be done to prepare Canadians for living and computing in a well-connected digital world.

Within Canada's universities there is an urgent need to create programs in computational science and engineering (CSE). Whereas the well-established universities have had strong academic programs in the core disciplines of science and engineering for many decades, the development of programs in computer science is relatively new. Despite Canada's growing strength in computer science, there is an urgent need to develop programs in CSE that fill the void between conventional science programs. These tend to be discipline specific, while computer science programs are not normally oriented to the application of computational methods for problem solving.

IMPACT Canada will work with Canadian universities and other organizations to foster the development of CSE programs at the undergraduate and graduate levels. These programs are inherently interdisciplinary and require graduates to develop expertise in advanced computing technology and in one or more disciplines. At the undergraduate level, these programs include a core curriculum in selected aspects of computing technology, numerical computation and the practical use of advanced computer architectures. In addition to meeting the requirements of the core curriculum, students may choose a stream leading to a BSc in computational science with specialization in a wide range of disciplines such as applied mathematics, atmospheric science, biology, biophysics, chemistry, geophysics, materials science, etc. BSc graduates in CSE provide a valuable talent pool for MSc and PhD programs in all modern scientific and technological fields.

IMPACT Canada will also work with Canadian universities and industries to facilitate student participation in co-operative (CO-OP) education and other internship programs that provide good training for careers in advanced computational fields. Furthermore, to foster internal competitiveness and to make Canadians aware of the best international practices, IMPACT Canada will work with Canadian universities to make use of foreign exchange programs.

There is a corresponding requirement for teams of scientists and computational mathematicians able to design, to modify and to optimize applications for future systems. These teams should ideally involve computer scientists in the task of analysing and abstracting the requirements of scientific applications, developing appropriate software environments that allow scientists to extract the maximum performance and capability from that hardware. With IMPACT Canada co-coordinating such an approach, we are optimistic of increasing both the scalability of existing codes to larger number of processors and, just as importantly, the percentage of peak performance actually delivered.

An equal need is to develop expertise in the design of HPC software tools. Canada has acknowledged strength in computationally-based research and in exploiting HPC. These strengths are not balanced by correspondingly strong HPC technologies (e.g., data-mining tools, meta-computing, performance evaluation tools, advanced visualization, and automatic parallelization). It is important to address this imbalance by developing this expertise, since these skill sets and technologies are vital if we are to fully capitalize on the physical and human investments being made in the computing infrastructure.

Coordinate and standardize the implementation and use of HPC infrastructure and personnel resources across the country

As the availability of, and the demand for, HPC increases in all fields, and in industry[10], IMPACT Canada will lead and coordinate the adoption of common standards for hardware, software, data and communication – an essential aspect of ensuring the long-term effective use of HPC resources. IMPACT Canada will also coordinate initiatives that enable seamless access to geographically separated HPC resources–processors, storage, visualization capacity, and collaboration facilities. This will require the careful coordination of resource providers, and the undertaking of strategic planning activities to develop technology and training standards. IMPACT Canada will be able to provide advice and specialized assistance to HPC centers across Canada, and will actively accept the challenges of efficient technology transfer.

Successful grid computing must not only overcome the inevitable but surmountable technological hurdles, but, more crucially, foster a highly coordinated and effective HPC organizational culture across Canada. The existing NSERC-funded Technical Analyst Support Programme (TASP) provides an innovative and key framework that allows various resource providers to work together on complex initiatives. It is proposed that IMPACT Canada replace TASP in coordinating the personnel resources that are distributed across the country with various resource providers. The far-flung nature of HPC activity will require that IMPACT Canada foster a strong sense of teamwork and common purpose. IMPACT Canada would also be responsible for interacting with network providers at both the national and regional levels and the training and professional development of support personnel. This latter task is crucial to developing, attracting and retaining HQP in technology based industries.

To effectively fulfill IMPACT Canada's mandate will require a dedicated full-time staff, including a director, a technical manager, outreach coordinators and technical experts. This first-rate support team will serve as a national resource, supporting national initiatives and broad-based collaborations. The emergence of advanced collaboration tools, such as the Access Grid, will allow researchers and the support team to interact and share applications using "voice and image over IP", a capability that makes this pan-Canadian initiative both exciting and realistic.

Enable and ensure that Canada plays a lead role in international computing initiatives

CANARIE has done an excellent job of positioning Canada internationally with respect to advanced networking. Indeed, we are perceived as a world leader in

---

[10] A new IDC report suggests that GRID computing (HPC) is on the verge of a major expansion, with users looking to optimize their resource utilization capacity. The market is expected to "exceed $12 billion in revenue by 2007 across high performance computing (HPC) technical markets and commercial enterprises" [REF 27]

such activities. We suggest that IMPACT Canada should play a similar role in advanced major computing initiatives such as PRAGMA (*the Pacific Rim Applications and Grid Middleware Assembly*). This will help access intellectual and technical resources of other nations while significantly facilitating international research.

Advise on the allocation and distribution of resources to maximize the competitiveness of Canadian HPC-dependent research

A key premise of this plan is that effective use of HPC resources maximizes scientific return. A strategy must be developed which both identifies and couples scientific applications with the appropriate computer architecture, thereby opening a sustainable path to the highest levels of computer performance. As part of this role, the agency will advise Canadian research funding agencies on HPC issues, trends, technical requirements and priorities.

IMPACT Canada will be accountable to funding agencies and the HPC consortia for its decisions. In addition, there is a clear requirement of funding agencies to appropriately measure the scientific results and value elicited from the proposed HPC investments. Whereas individual grant holders and consortia provide this type of information, it is performed piecemeal and does little to construct a national picture of evolving HPC requirements. There is also currently no coordinated effort to assess where Canada is in terms of international HPC investment and production. Consequently IMPACT Canada will review and report on HPC activities and investments from a national perspective. Such a picture would be of immense value to both national and provincial agencies as well as the research community with respect to future decision making and anticipating coming needs and trends.

Assess future development needs to ensure that Canada's HPC investment remains competitive for the future

IMPACT Canada will initiate a "technology watch". The rapid change of HPC technology requires the ongoing assessment of technology as it evolves and is adopted internationally. IMPACT Canada will be ideally placed to play a lead role in this effort, ensuring awareness in the HPC community, and avoiding the duplication of effort and investment. Technology watch activities include gathering and disseminating technical information, coordinating vendor presentations, undertaking benchmarking, and providing advice on appropriate software architectures for user applications. These activities will also support outreach efforts and promote public awareness.

Assist in the selection and operation of Canada's tera-scale facility

IMPACT Canada will establish a scientific committee that will recommend the site, composition and use of the tera-scale facility. In consultation with stakeholders such as the consortia and funding agencies, IMPACT Canada will oversee all facets of this process, as well as provide the leadership, management and training support needed to ensure that researchers choose and exploit the most appropriate resources for their research. One of IMPACT Canada's most important tasks will, therefore, be to oversee access to this unique national capability-computing facility. IMPACT Canada will thus require a dedicated full-time staff for this facility – perhaps modeled on the NRC-IRAP (Industrial Research Assistance Program) model. These personnel will help users migrate their applications from mid-range facilities and enable them to extract the full scientific potential from the central ring of HPC. In terms of the overall HPC strategy, tera-scale computing provides a component viewed as essential by our major competitor nations; one that is conspicuously missing in Canada.

Facilitate an interface between industry and HPC resources, and connect Industry to researchers and research opportunities

Small and medium sized enterprises (SME's) or businesses are recognized to be the economic engine of our economy. Unlike many large corporations, these businesses do not typically have access to HPC resources and cannot, therefore, take advantage of

HPC research innovation. We thus strongly urge that the National Research Council (NRC)/Industrial Research Assistance Program (IRAP) and IMPACT Canada establish mechanisms to facilitate access to HPC, highly qualified personnel, and advanced research outcomes and equipment by the SME base in Canada. This includes linking SMEs with researchers and research opportunities.

Canada could stand at the leading edge of the 21st century economy: we offer an HPC strategy to make Canada *the* country that most fully realizes the potential of HPC. Initial investments within our universities/hospitals have sown the seeds for new and exciting economic gains for Canada. An excellent example of this is the Montreal-based biotechnology company Neurochem Inc., which grew out of research conducted at Queen's University. The company is publicly listed (TSX: NRM) and now has a market capitalization of over $1 billion.

Both start-up and more established businesses, whether they are involved in R&D supporting manufacturing, biotechnology, the life sciences, communications, finances, e-commerce etc... will require access to HPC to remain competitive and grow. In fact, they must "plug into" Canada's HPC capacity if R&D productivity gain estimates of 47% through the use of grid technology are realized (as suggested in "The Global Grid Computing Report 2002 – Technology and Market Assessment" [REF 28]) By closing this access gap, IMPACT Canada will catalyze the ability of SME's to convert research opportunities and HPC capacity into social and economic benefits for Canadians.

Accountability for the expenditures on HPC

Given the substantial mandate of the agency, and investment needed to establish and operate IMPACT Canada, we recommend a regular review of its operations and role as part of a full review of Canadian HPC investment. This will provide the opportunity to assess its success as well as future funding needs and priorities. It will also ensure the development and implementation of a consistent roadmap that evolves in response to the rapidly changing technological environment within which HPC exists. We recommend that IMPACT Canada be re-evaluated every five years as part of a national review of Canada's international HPC related research competitiveness. This review would, as appropriate, incorporate the following performance indicators: **academic excellence** (refereed publications, national and international awards, retention and recruitment of faculty); **highly qualified personnel produced** (graduate students, post doctoral fellows, technicians and research associates); **scientific impact; and societal and economic impact** (enhanced health outcomes, improved public infrastructure, patents, spin offs, industrial collaborations and partnerships, technology transfer).

Summary

IMPACT Canada will coordinate HPC facilities and offer technological outreach, and education. It will create strategies for enhancing and applying HPC in education and research in the health, environmental, applied, physical, life, and human sciences, for coordinating existing resources across the country, and for allocating new resources so as to maximize competitiveness of Canadian research. It will have sufficient staff to assist with program development, training, co-ordination of symposia, and to provide advice and specialized assistance to HPC Centres across the country. The agency will also always be ready to advise Canadian research funding agencies on HPC issues, and provide accountability for national HPC activities.

Mining Memories and Mending Minds

Treatment of the brain has always been a fragile and challenging medical practice. Diagnosing brain abnormalities and disorders has been a painful procedure for both patient and doctor. The risks associated with any cranial investigation are huge.

Until now, seeing inside the skull without risky and non-invasive imaging technologies was virtually impossible. Now, we can get inside using new tools and techniques. It is none too soon. With an aging population, increasingly afflicted by degenerative brain disorders and psychiatric illness, these approaches are desperately needed. Health Canada estimated the economic burden of mental illnesses at more than $14 billion per year, though a more realistic estimate may be over $35 billion.

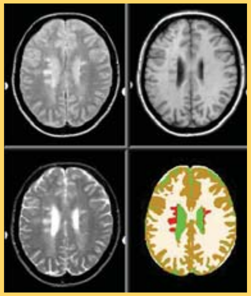
Canada can be the global leader in brain imaging research. We have a distinguished history in brain research, while present advances in neurology, psychology and cognate fields are breathtaking. But high performance computing is crucial to securing Canada's place. The power of computational tools, which was essential to the triumph of the human genome project, is being brought into brain research and the time for Canada to act is now. And it is worth remembering that every step of a brain scan relies on fast and effective computing, which produces three-dimensional images that may then be mined.

The implications for Canada's health programs are immense. One brain scan is useful, especially if it is your own brain acting up, but thousands of brain scans may be needed to determine the appropriate treatment. When well archived, brain scans become a wealth of information that doctors and researchers may mine to analyze and understand brain function and disease. Scientists must develop highly sophisticated search engines, expert systems and image analysis tools to allow researchers to fully utilize these raw data. Patients will be better diagnosed without risky and invasive procedures, minimizing the chance of neural complications and possible cranial infections. Brain imagery tools will also enable new analyses with exciting repercussions for patients who, for example, are about to have surgery to replace brain tissue and recover lost capability.

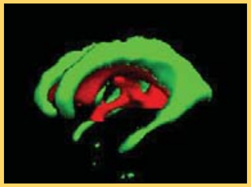
Brain research in Canada is increasingly making use of large brain image databases, often constructed from research sites around the world, to understand brain disease and treatment. These strategies are also used in the assessment of new pharmaceuticals for treatment of brain disorders, requiring massive computational and network bandwidth capabilities. Continued support of high performance computing infrastructure to enable these outstanding research breakthroughs is therefore imperative. The societal implications alone warrant providing the requisite computing for such exceptional medical efforts.

> *"With competitive HPC resources, Canada will be able to compete, and possibly lead, in applying information technology to brain research. The benefits of such research strength in terms of brain disease treatment are incalculable. With an aging population, increasingly afflicted by degenerative brain disorders and psychiatric illness, it is essential that we have access to state-of-the-art neuroscience research to develop effective therapeutics."*

Dr Alan Evans
Coordinator, McConnell Brain Imaging Centre
Montreal Neurological Institute and Professor
of Neurology and Neurosurgery
McGill University



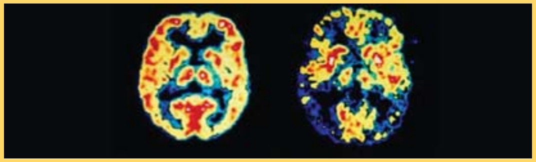
Three different kinds of MRI image of the same brain (grey) showing multiple sclerosis (MS) lesions. These are identified (red in 4th panel) by using automatic image processing to identify different normal tissue and MS lesions (red).



Combining data from hundreds of MS cases, we build a 3D 'map' of MS lesion distribution. The lesion map (now in green) surrounds the central fluid space in the brain (red). New pharmaceutical treatments can be evaluated by measuring the change in the green lesion distribution with and without the treatment. This summarizes a 500-GByte database from 460 separate patients and over 1800 separate MRI studies. [REF 29]

"Historically, Canada has not had home-grown drug companies with a multinational presence to contribute to the wealth of our nation. Countries much smaller than Canada have traditionally enjoyed the economic benefits of vibrant drug companies: Denmark (Lunbeck), Sweden (Astra), Switzerland (Ciba-Geigy, now Novartis). An evaluation of all drugs discovered worldwide (by country of origin) from 1982-2002 reveals a linear relationship correlating average number of drugs discovered with size of the gross domestic product (GDP) for the top ten drug producing countries: number of drugs = -1.7 + 36.9GDP (GDP expressed in $US trillions) Canada should be discovering 16 drugs per decade. We do not achieve even a fraction of this, and thus are not in the top 20 NCE discovering countries. Accordingly, we need to build capacity in NCE discovery and only through HPC can we do this."

Dr Donald Weaver
Professor at Dalhousie University
Canada Research Chair in Clinical NeuroScience



Low metabolism PET scans show lower activity (blue and black regions) in the brain of an Alzheimer's patient (right) than in a healthy brain.

High Performance Economics with Computer Aided Drug Design

Safe and effective. Economically produced. Useful therapeutics. Linked together these terms have a major impact on Canada's pharmaceutical sector. How? Through bioinformatics and computational chemistry research, experts now have the ability to develop and discover New Chemical Entities (NCEs), which are the building blocks to safely designed and tested drugs.

High performance computing will be central in creating this capacity for NCEs discovery. HPC is now fundamental to the drug discovery process, is crucial to the massive data handling issues associated with modern drug design and is critical in visualizing the process whereby a drug binds to its receptor site. HPC is absolutely essential to enable the various calculations which permit the shape and geometry of a drug molecule to be computed (molecular mechanics, molecular dynamics, molecular graphics, semi-empirical molecular orbital calculations, ab initio quantum mechanics calculations. How might this impact the Canadian economy?

The pharmaceutical sector is an important component of the economies of developed countries. NCEs as therapeutics for human disease will be the "oil and gas" of the 21st century. As the world's population increases and confronts the ever-expanding health problems of the modern world, the need to discover NCEs in the first half of the 21st century will be as great as the need to discover petroleum was in the early years of the 20th century. Moreover, the "pay-offs" will be just as great; an effective drug for Alzheimer's disease or a single new antibiotic drug with widespread usefulness will be a "billion dollar molecule".

The time is now to support bioinformatics and chemoinformatics as the most rapidly growing area in biotechnology in Canada.

The time is now to support bioinformatics and chemoinformatics as the most rapidly growing area in biotechnology in Canada. These research areas are crucial in the development of useful therapeutics (i.e. useful drugs), molecules that are not only efficacious and safe, but also can be easily and economically produced, patented, successfully complete clinical trials and ultimately marketed. These computer-driven informatics processes are central to the health and well being of our pharmaceutical sector and our Canadian society.

HPC Helps Define the Structure of Proteins

It is impossible to overstate the importance of proteins to plant and animal life. Much of the tissue in the human body is made of protein, as are all of the enzymes that catalyse reactions in the body, the globins that transport and store oxygen, the antibodies responsible for the body's immune response and hormones such as insulin. By understanding biological processes at the protein level, it will be possible to improve drug therapy, obtain better biomarkers for health and diseases, and improve the application of proteins in plants, animals, food and nutrition.

Today, we know that humans do not have many more genes than lower organisms, and that the secret lies in the proteins that each gene codes for. To put the challenge in perspective, there are about 100 000 proteins in the human genome and the structural information about these proteins is critical in understanding how proteins work for drug development, cell differentiation, protein production and regulation.

In order to investigate the 100,000 proteins and their different crystal structures researchers must be able to solve the structures of these proteins using a radiation source, which produces massive terabytes of data.

Facilities like the Canadian Light Source make it easier to obtain x-ray diffraction pictures of protein crystals, which in normal x-ray labs would be impossible. Currently it takes less than a day of beam-time at the Canadian Light Source to produce enough synchrotron radiation to produce the diffraction data of a single protein structure.

In order to determine all of the protein structure of the human genome, high performance computing support of the Canadian Light Source is crucial. Once beamlines are used to determine a protein's structure they can be used to determine how well hundreds of drug candidates synthesized by pharmaceutical research can fit into the active sites of target proteins. These developments have huge implications on the pace of drug synthesis and may answer questions such as why some people can defend themselves against the AIDS virus while others cannot. It will provide a better understanding of protein-drug interactions, helping create better and more effective drugs.

There is unlimited potential in the field and Canada can position itself as a global leader if it can provide better infrastructure and build better teams of researchers than its competitors.

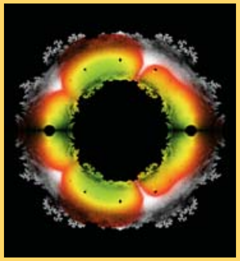
> *"Protein crystallography would be impossible to perform without the availability of massive computing power. The ability to generate a picture of protein's structure and perform comparative analysis continues to plunge us deeper into the depths of understanding why proteins evolve, mutate and function abnormally in certain environments. This not only aids Canadian researchers in understanding genetic disorders and diseases, it allows for a better understanding of protein-drug interactions. Continued growth, however, relies upon the assurance of continued access to high performance computing systems. Without this, Canada will be unable to continue to compete at a global level."*

Mark De Jong
Project Leader
Canadian Light Source

Ribbon plot of the trimeric arrangement of KDPG aldolase. Each monomer is shown in ribbon representation with a different color for each. KDPG aldolase is an enzyme that provides a mechanism for carbon-carbon bond formation in living organisms [REF 30]

"HPC is an essential research infrastructure for these solutions. Without it, our knowledge society has weak foundations."

Dr Ian Lancashire
Professor of English
University of Toronto



"Numbers are not the only thing that computers are good at processing. Indeed, only a cursory familiarity with fractal geometry is needed to see that computers are good at creating and manipulating visual representations of data. There is a story told of the mathematician Claude Chevalley, who, as a true Bourbaki, was extremely opposed to the use of images in geometric reasoning. He is said to have been giving a very abstract and algebraic lecture when he got stuck. After a moment of pondering, he turned to the blackboard, and, trying to hide what he was doing, drew a little diagram, looked at it for a moment, then quickly erased it, and turned back to the audience and proceeded with the lecture. It is perhaps an apocryphal story, but it illustrates the necessary role of images and diagrams in mathematical reasoning – even for the most diehard anti-imagers. The computer offers those less expert, and less stubborn than Chevalley, access to the kinds of images that could only be imagined in the heads of the most gifted mathematicians, images that can be coloured, moved and otherwise manipulated in all sorts of ways." (Nathalie Sinclair, 2004) [REF 31]

Real or Fake? Dynamically Validating the Reliability and Authenticity of Digital Records

A paper-only letter today is relatively secure. Its letterhead, its author's signature, and perhaps its watermark breed faith in its reliability as a product of its stated author. Proper pagination and a list of any attachments tell us whether the letter has changed during transmission. Without obviously missing pages or physical damage, we will accept that the document is authentic.

In contrast, few people trust digital documents. Email or web documents often have no digital signature or secure evidence of provenance. A sender's address can be very easily forged, and routinely any attempt to reply is returned for lack of a true address. Digital content can also easily be plagiarized so that the identity of the true creator is hidden. Digital records also are vulnerable to authenticity problems as a result of transmission or use as computer displays may fail to reproduce what was sent. Once digital records are stored, other potential problems can occur. Records can be damaged during routine processing, and no one is the wiser.

So how might one create a level of authenticity and security to instill a level of confidence in digital records creation, management and retrieval? Only by working with grid-based high performance computing technologies can the problems in determining, in real-time, the reliability and the authenticity of digital records as they are created and transmitted be dynamically solved each and every day. By using high performance computing resources each record receives a global identity, persisting over time, within a so-called logical name space. That name space receives information about the states that a record assumes during copying. It also stores descriptive metadata, some of which would characterize the digital record's reliability and authenticity.

The significance of this level of authenticity and security are huge. Government agencies increasingly rely on the Internet to investigate terrorism and crime. Multinational and pan-Canada industries and financial institutions use digital records every day. Accidental loss or corruption of these records can cause significant economic, security or privacy problems. Education at secondary and post-secondary levels looks to web-accessed databases as a medium for "distance" education – increasingly on a scale with tens of thousands of concordant users. The humanities and the social sciences are stakeholders in storing all Canadian historical, artistic, and literary records of the past thirty years and well into indeterminate future.

With a high performance computing architecture we can combat these threats and create excellent digital record systems, which are reliable, safe and can be preserved for our children and for their children.

Chapter 5: Funding

"The tools for scientific discovery have changed. Previously, science had been limited to experiment and theory as the two pillars for investigation of the laws of nature. With the advent of what many refer to as "Ultra-Scale" computation, a third pillar, simulation, has been added to the foundation of scientific discovery. Modern computational methods are developing at such a rapid rate that computational simulationis possible on a scale that is comparable in importance with experiment and theory. The remarkable power of these facilities is opening new vistas for science and technology. Previously, we used computers to solve sets of equations representing physical laws too complicated to solve analytically. Now we can simulate systems to discover physical laws for which there are no known predictive equations."

Testimony of Raymond L. Orbach, Director, Office of Science, U.S. Department of Energy, before the U.S. House of Representatives Committee on Science, July 16, 2003

Introduction

Modern computing is transforming research in Canadian universities, research hospitals, and high-tech industry. Traditional "wet science" performed in the field or laboratory is often supplemented by "dry science" using computer simulations and models to study the same phenomena. Dry science often takes the research into realms previously inaccessible, and has provided scientific insights that were inconceivable a few years ago. In turn, this has dramatically increased the demand for high performance computing (HPC), not only by the traditional science and engineering based researchers, but increasingly also by researchers in the social sciences and humanities. Numerous examples of the importance of HPC in these areas are provided in the previous chapters.

As we enter the 21st century, our shift to the use of HPC resources will continue to grow, and with this comes the pressing need for a coordinated, efficient and effective approach to making them available throughout Canada. The current approach to HPC funding is not sustainable: it is piecemeal, does not allow for planning beyond the immediate short-term, and is not competitive with the coordinated efforts of other nations investing in HPC. To become competitive Canada must also invest in the manpower, operational and computing infrastructure resources needed to support Canadian researchers. The leverage provided by this investment will have direct impacts on our research capacity and international research profile. It also provides significant downstream economic and quality of life benefits, through new knowledge, technological innovations, and a critical mass of highly qualified personnel.

Investment in competitive HPC is not inexpensive, but it is clearly crucial. It requires a continuing investment in mid-range computing systems, the addition of a high-end computing resource, the personnel needed to operate and leverage the value of these resources through coordination, management, development and outreach support (IMPACT Canada), and the expansion of existing facilities. HPC investments must also recognize Canada's geographic reality, which requires the dispersion, but not dilution, of HPC capabilities across the country. These factors lead us to propose that Canada grow its annual public investment from over $40 million today (for the 6 major, and 2 single University (Victoria, Toronto) based HPC Resource Providers) to $76 million in 2006 and to $97 million in 2012 (2004 dollars).

This chapter explains current HPC investments in Canada, presents a number of key funding dimensions that support an effective HPC strategy, and outlines our funding proposal for the future. Canada's status in terms of its HPC investments as compared to those made by other nations provides a context for our funding recommendations.

It should be noted that this chapter has been written so that it may be read in isolation of the other chapters. There is, therefore, some repetition of information from previous chapters.

HPC Investment in Canada to Date

Between 1980 and 1998 there was no Canadian strategy for funding HPC infrastructure. Alberta and Ontario attempted to create HPC computing centres, but these efforts were short-lived principally due to the lack of sustained funding. However, these initial forays into creating a shared HPC resource provided the foundation for the present HPC community, and resulted in, for the first time, research communities formally defining their HPC resource and funding needs, and establishing their visions for the future. In 1997 this led to the establishment of C3.ca, the national voice for high performance computing in Canada, which today represents over 50 academic and industrial organizations.

It was not until the Canadian government provided a vehicle to pursue this vision, the Canada Foundation for Innovation (CFI), that Canada was able to become internationally competitive in HPC-related research. As a result of the 1998, 1999, 2001, and 2003 competitions, CFI and its provincial and industry partners have invested about $250 million to establish, operate and maintain mid-range advanced HPC facilities to serve Canadian researchers. This investment has produced a vibrant and productive research community, catalyzed the attraction and retention of high-quality researchers, created the technology needed to maximize the benefits of HPC resources, and enabled the training of the highly qualified people that are essential to maintaining Canada's competitiveness in the 21st century. Six regional, and two university specific, HPC Resource Providers are now in place for Canada to build on. However, funding investments are still sporadic and uncoordinated, and future investments, although essential, are uncertain.

Thinking Strategically

A long-term stable funding strategy is needed to create and sustain a vision for HPC that guarantees the best value for the Canadian public, and ensures the long-term benefits of the resource for Canada's innovation and research environments. To be effective this strategy must exemplify the following funding investment dimensions:

1. **Continuity:** Provide continuous and stable funding to ensure that Canada can reap the maximum benefit from longer term strategic research and development initiatives, and exploit any short term opportunities that arise. It is essential that there are no funding gaps as this leads to the loss of: opportunity, invaluable expertise and trained human resources, and competitive advantages. This is particularly important as a new system can take up to three years to implement.
2. **Growth:** Recognize evolving HPC needs as more researchers, from diverse disciplines, integrate the use of HPC into their research plans. There must be commensurate growth in resources to support continued reasonable access.
3. **Competitive Research Infrastructure:** Be aware that Canada's international competitiveness in many research areas and industrial sectors is dependent upon local, regional and national research, development, and training strengths. This requires HPC users to have access to world-class capital infrastructure.
4. **Capital and Operational Investments:** Support a complete solution (human resources, operational support, infrastructure and facilities) in order to successfully exploit the capacity represented by HPC. The lack of any of these resources will seriously erode the value obtained from the others.
5. **Planning for the Future:** Recognize that effective funding and resource coordination and management is essential to strategically building on and exploiting existing and future HPC investments. IMPACT Canada is

intended to provide long term planning and effective HPC exploitation.

Canada cannot depend on others for its HPC needs. We cannot become second-class HPC users, we cannot afford to lose our expertise and trainees to other countries, or erode our research capacity. In addition, our industry partners cannot lose their capacity to develop and export "made in Canada innovations", or lose access to the well-trained workforce that makes this possible. Canadians must have direct access to HPC tools. Canada must grow and retain its own well-trained workforce. Our nation must be able to compete internationally. We must invest now.

Choosing Wisely: HPC Investment Requirements

The three areas requiring investment are i) the support of mid-range and advanced HPC infrastructure, ii) investments in people and operations, and iii) the establishment of IMPACT Canada[11]. These investments must be made within the context of the five funding dimensions discussed above if they are to be effective. In addition, it is essential to understand that although ramping up a hardware investment can be achieved relatively quickly, building a competitive and skilled human resource base can take much longer, especially if these personnel turn over or are lost due to inadequate or inconsistent funding.

The following funding proposal uses a six year timeline (2006 – 2012) as a basis for HPC planning. Funding beyond 2012 is subject to review and future planning, as the numbers are difficult to predict beyond this time considering changing technology costs, increases in human resource costs, and other factors outside of our control or not yet known. It is, however, essential that a stable funding base be in place at that time to sustain the value of Canada's HPC investment. Nationally coordinated strategic planning, proposed to be undertaken by IMPACT Canada, will assess specific needs on a yearly basis. This work will ensure an optimal and coordinated approach to HPC use and development, and will assist funding agencies to plan for anticipated investments.

Funding Requirements for Mid-range and Advanced HPC Infrastructure

HPC resources need to be upgraded more frequently than most other types of infrastructure. Moore's Law dictates the need for upgrading HPC equipment about every three years if research facilities are to be sustained at an internationally competitive level. The frequency and pervasiveness of this *technology refresh* is without precedent in the history of scientific research.

Mid-Range Computing Infrastructure funding

Mid-range computing facilities, as noted in Chapter 2, are the bedrock of our HPC model. HPC facilities in this range enable a vast range of leading research initiatives, and currently support, through the eight HPC Resource Providers, almost two thousand faculty, graduate students and research associates. Mid-range computing facilities currently include systems that, in a multi-user environment, can provide an average of about 32-128 dedicated processors per research application (the average desktop provides 1), with some applications requiring substantially more processors for short periods of time. A mid-range facility will rank between 30 and 500 in the Top 500 listing of HPC worldwide. In Canada all HPC Resource Providers can only offer the existing body of users resources at this level.

The most recent CFI competition saw four major HPC sites (SHARCNET, HPCVL, ACEnet, University of Victoria) funded to the level of $88 million of public funds over an average of four years (roughly $22 million of capital infrastructure per year for the four years that represent a typical CFI application request). The next competition will see the other major sites (WestGrid, CLUMEQ, RQCHP and the University of Toronto) pursue similar or even higher funding levels (although it is not clear yet whether CFI and the provinces will have enough money to

---

[11] It should be noted that this funding does not include the continued support of foundational research computing infrastructure (desktops, workstations and dedicated machinery required for certain project specific applications) by the granting councils.

cover these and the multitude of other funding requests). Current public infrastructure support for HPC is over $40 million per year for these facilities (see also the table 5.1).[12] If in the next round of the CFI the HPC applications are renewed at the same rate as in the 2003 competition, the average annual infrastructure expenditures are anticipated to grow to at least $44 million in 2006. However, CFI has limited remaining funds. In fact, CFI itself has no long-term viability unless it receives a substantial infusion of funds from government to deal with the nation's research infrastructure needs for 2006-2010 and beyond.

In the following section we outline our proposal for funding mid-range computing in Canada. The key potential impacts (benefits), and most substantial risks associated with not investing in HPC, are listed to reiterate the importance of the funding to Canadian research and researchers. The proposal takes into account the following basic assumptions:

- Technology costs are assumed to remain similar over time. This assumption is made on the basis that as technology costs drop, greater performance can be purchased for the same dollar.

- The expected lifespan of a primary HPC system is three to four years[13]. This lifespan is consistent with federal computer technology amortization rates (45% per year starting in 2004), which recognize the rapid depreciation and limited lifespan of these assets[14]. At these rates, a computing investment will be worth less than 10% of its original value in four years.

- All of the major HPC facilities are currently reporting that they are either close to or at capacity. Current infrastructure refreshes, as supported by the 2003 CFI competition, are providing new capacity, but there is no guarantee of future funding or that this will be sufficient. If operational funding is not tied to the capital investment, it will become even more difficult to sustain Canadian HPC.

- As users become more sophisticated, and as their computational models become more complex, there will be a steady growth in single application processing demand. This will reflect the desire to gain greater insight into the research problems supported by HPC, and the ability of users to leverage more of the resource.

- The requested funds do not include the capital infrastructure contributions from industry, only the portion required from public sources. Industrial contributions are anticipated to range from $13 million in 2006 (about 30% of the public expenditures) to $16 million in 2012. These figures are less than the industrial matching contributions achieved thus far in the CFI program, and reflect the move to a greater use of commodity components for which vendor margins are much lower. The capacity of vendors to contribute is, therefore, substantially reduced.

- That user demand in excess of the increased capacity represented by processor refreshes will be 3.5% per annum. Consequently, the time taken to run a simulation remains constant.

Our proposal for funding assumes an increase of 3.5% per year in HPC costs based on the above noted assumptions. Using the current investment of $41 million per year as a base, it is anticipated that mid-range computing in Canada will require the following capital investment shown in Table 5.1:

---

[12] Public funds are accessed from multiple sources, which include federal, provincial, research council and foundation funds. Most applications require partnerships between such sources.

[13] Experience shows that many of these systems will be repurposed after this time, space and cost issues permitting, but they are no longer adequate to support competitive HPC based research.

[14] According to the Federal 'Budget 2004' speech, the *Capital Cost Allowance* for computing equipment increased to 45% from 30% per year, and for Internet, broadband and other data networks from 20% to 30% per year, in order to better reflect the depreciation of these assets. "These changes will bring the rate at which businesses can write off their investments in line with the useful lives of these assets." [REF 32]

Table 5.1: Proposed Capital Investment for Mid-range HPC Facilities

Type of Cost	2004	2006	2009	2012
Capital Costs	$41 million	$ 44 million	$ 49 million	$54 million

Notes: 1. All figures are based on 2004 dollars and do not make allowance for inflation.
2. Capital costs only; additional funds are for needed for human resources and operations as discussed later.

Anticipated impacts (benefits) of funding include that:

- It will put Canada in a position to support its existing investments in HPC to the benefit of Canadian research and researchers.
- It will provide a coordinated growth in resources. This coordination will help to avoid the fragmentation of funding into piecemeal HPC assets, as small groups will not need to duplicate the resources provided by HPC Resource Provider
- Our capacity to support and collaborate with industry, as well as our ability to participate in international research initiatives, will be much enhanced, providing increased opportunities for technology transfer and research partnerships.

The risks associated with not proceeding are high. Of most concern, is the risk that Canadian research will not progress in comparison to that of other Nations. Difficult problems in sectors such as the life sciences, manufacturing, information technology, and energy will, for example, not be solved here. Other risks of particular concern include:

- An inability to support, and thus recruit and retain, exceptional researchers such as Canada Research Chairs who require HPC resources to accomplish their research.
- To access HPC resources our graduates will seek employment outside of Canada.
- Canadian innovation capacity is reduced as Canadian research falls behind, and Canada loses the expertise to innovate from using HPC as a technology platform.
- The likelihood of piecemeal acquisition will increase as individual researchers seek to compete for needed resources. This will decrease the value per dollar achieved by funding agencies, and further decrease access.
- That the existing user community will be constrained for HPC resources to what processor refreshes can support. This dramatically reduces their capacity to invent, and limits the scope of the research problems tackled in Canada.
- A loss of industry partnerships, and the research opportunities, support funding and matching capacity that these partnerships represent.
- The eventual erosion of local opportunities, as Canadian industries purchase related goods and services from outside of Canada. This weakens the capacity and competitiveness of our own industries, and, furthermore, decreases the competitiveness of Canada's supporting value added product and service businesses.

Advanced Computing Infrastructure Funding – A High-end Facility

As per Chapter 2 of this document, a high-end HPC facility has thousands of processors, and is capable of sustaining teraflop (Tflop) performance on individual user applications in a multi-user environment. These are the systems that find a place in the top 30 of the World's Top 500 list, and support the most computationally intensive research – the grand challenge problems. High-end computing facilities are available only to key peer-reviewed projects of exceptional scientific quality and significance. In this sense, these facilities are similar in nature to other large scientific instruments such as particle accelerators and astronomical observatories.

Canada currently has no high-end computing facility, and is, therefore, not competitive as compared to other nations in terms of the research supported at these levels. To fill this significant gap it will require a capital investment as given in Table 5.2:

Table 5.2: Proposed Capital Investment for a Pan Canadian A High-end HPC Facility

Type of Cost	2004	2006	2009	2012
Capital Costs	$0 million	$ 10 million	$ 12 million	$14 million

Notes: 1. All figures are based on 2004 dollars and do not make allowance for inflation.
2. Capital costs only; additional funds are for needed for human resources and operations as discussed later.

Investing in a high-end facility provides Canada with strategic advantages in terms of its capacity for research innovation. See for example the sidebars in Chapter 2 on climate and meteorological (weather) modelling, data management, analysis and visualization needs, and grid computing capacity and applications. See also the case studies on the importance of HPC to Canada's aerospace industry (Chapter 2), and to virtual cell modelling and nanotechnology (Chapter 3). It is also an important attractant to exceptional researchers, for example those who would take up Canada Research Chair positions. Many of Canada's industries (e.g.: biotechnology, manufacturing, information security) would benefit from this resource, through their research collaborations with academia. Specific funding impacts (benefits) include:

- It will enable Canadian researchers to access internationally competitive high-end computing, advancing their research in key competitive areas.
- It provides Canadian researchers with the tools needed to solve complex research problems and allows their partners to use this job-critical knowledge to innovate.

The down-side risks associated with not funding a high-end facility include:

- Losing our international competitiveness in being able to address the most computationally intensive research areas, and competing on grand challenge problems.
- Lacking a path for the advancement of research problems from the mid-range to the highest level. This will result in the loss of our researchers to nations with better facilities, so they can perform their research and/or maintain their personal competitiveness.
- Foregoing collaborations and partnerships in the many high technology research areas that require capability computing facilities.

The significance of these risks can be seen in the context of research problems discussed in previous chapters.

> *"To allow Europe to compete scientifically in future with the US and Japan and in light of recent developments also with China, it was essential that the supply of computing resources be expanded continuously in both quantitative and qualitative terms."*
>
> ***Regarding the German Science Council's recommendation to create three super-computers of the highest performance in Europe.***

How do we compare today? Why we MUST invest in Mid-Range and High-End HPC.

The following are examples of investments that other nations are making in HPC resources. While each represents a distinct funding and implementation strategy, these examples are intended to show the importance and significance placed on HPC investments by those nations serious about their research competitiveness.

- Canada currently represents 1.4% of the Top 500 Supercomputer Sites in the world (November 2004), representing 7 of our HPC systems. Our highest ranking is 54th on this Top 500 list – the benchmark for supercomputing around the world.
- We are behind all G8 nations excepting Russia, and our top system ranks behind systems in countries such as Saudi Arabia, Brazil, and Israel.
- Nations such as India and Korea have more HPC systems on the Top 500 list – 17 and 11 respectively. Neither country comes within 40% of Canada's per capita GDP. It is worth noting that Giga Flops (GFlops) per GDP (per capita) is the default metric for assessing each nation's supercomputing investment. Figure 5.1 on page 92 shows the relative investment in HPC of the 14 countries at the top of this list. Canada, in 14th place, lags significantly behind thise nations with which it typically compares itself.
- By 2006, the United Kingdom will have invested £50m (~C$110m) over four years to support the creation of a national supercomputer (22 teraflops)
- The German Science Council estimates that the German governments (federal and state) currently must expect to face costs of 30 to 50 million euros to situate one peak performance machine in Germany [REF 33]
- Spain hosts the most powerful supercomputer in Europe. This system will be enhanced from 31 teraflops to 40 teraflops by 2006, through an investment of € 70 million. Spain is ranked 8th in Gflops per GDP (per capita) – 6 spots ahead of Canada. Based on todays expenditures, adoption of the recommendations in this LRP would move us to 6th place (assuming others do not invest any more).

"Discovery through simulation requires sustained speeds starting at 50 to 100 teraFLOPS to examine problems in accelerator science and technology, astrophysics, biology, chemistry and catalysis, climate prediction, combustion, computational fluid dynamics, computational structural and systems biology, environmental molecular science, fusion energy science, geosciences, groundwater protection, high energy physics, materials science and nanoscience, nuclear physics, soot formation and growth, and more"

Testimony of Raymond L. Orbach, Director, Office of Science, U.S. Department of Energy, before the U.S. House of Representatives Committee on Science, July 16, 2003

Human and Operational (Facility) Resources

In common with other major scientific infrastructure, the effective and productive use of the physical infrastructure requires substantial continuing investments in technical management, operational and scientific support, administration, outreach and training: the *human infrastructure resources*. Additionally, there are substantial costs associated with providing power, air conditioning, security, data backup, software acquisition and maintenance, hardware maintenance, uninterrupted electrical power and a host of other support services: the operational or *facility costs*.

Current Support

Prior to the 2002 CFI competition, there was no CFI funding for HPC operations – researchers had to cobble together operational funds from a variety of sources (some provinces, universities and industry partners provided funds). The 2001 and 2003 CFI competitions supplemented successful grants by providing an additional 30% of the CFI award to be used towards operating costs. The provinces did not follow suit, which meant that the operating funds applied to only 40% of the total project budget – or about 10% of the total budget.

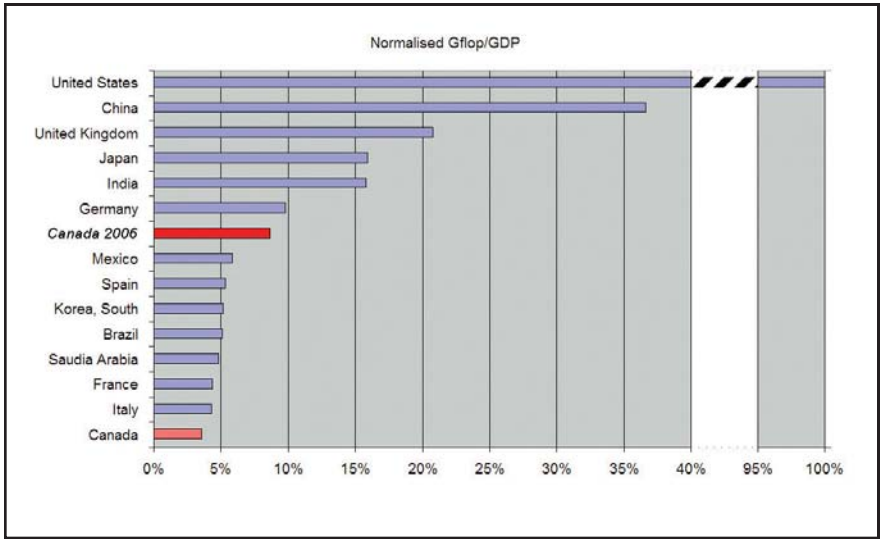


Figure 5.1: Normalized graph showing GFlops per GDP (per Capita) and where we would be positioned in 2006 if the recommendations in this LRP are adopted [15,16]

Experience with the infrastructure that was put into place as part of the 2002 CFI awards indicates that the CFI funds are sufficient to provide a minimal operational level of human infrastructure support, but no more. This assumes that the facility overhead costs (such as electricity and air conditioning) are paid for by the universities. This is becoming an increasingly problematic situation, as this type of facility does not attract federal indirect costs of research support. There are, therefore, no funds left over to provide proactive support services – analysts to help develop, enhance, port, and optimize HPC applications – and this is a serious deficiency. In the future, this matter will be compounded if CFI is not able to provide operational funding at the current 30% of their portion of the capital infrastructure cost.

The Department of Energy High-End Computing Revitalization Act of 2004 authorizes DOE to establish supercomputer user facilities that will provide U.S. researchers access to some of the world's most advanced computers on a competitive peer-reviewed basis.

Proposed Funding

> *"Government agencies responsible for supercomputing should increase their levels of stable, robust, sustained multiagency investment in basic research. More research is needed in all the key technologies required for the design and use of supercomputers (architecture, software, algorithms, and applications)".*
>
> ***Recommendation 6 of the US Committee on the Future of Supercomputing (National Research Council) in their publication: Getting Up to Speed: The Future of Supercomputing***

Human Resources

Based on experience at Environment Canada (Dorval) and international HPC centres, a typical *human infrastructure budget* for basic operations and HPC researcher support services is roughly 25% of the TOTAL capital infrastructure cost. This translates into an investment of $13 million in 2006 growing to $15 million in 2009 and $17 million in 2012 to reflect increased HPC use[17]. The funding needed is significant, and must take into account that all researchers and Resource Providers must have direct access to a variety of dedicated skills in-house or locally. Naturally, a supporting facility investment is required to house these personnel. This investment is significant given the broad geographic context, and the dispersion of skills (e.g.: technical, managerial, systems admin, programming) required. Table 5.3 outlines the human resources needed to support HPC Resource Providers. The reader is referred to Chapter 3 of this report for an analysis of the types of technical personnel needed.

The human infrastructure budget of Environment Canada (and other major HPC sites) is in direct contrast to the very limited resources (~$1,000,000 per year) currently available through the Technical Analyst Support Program (TASP), and other Major Facility Access (MFA) grants made available by NSERC to support HPC. These, and the operational funding supported by CFI (about $6,000,000 per year on average), are the only formal support mechanisms for human resource and operating costs for HPC in Canada. This underfunding has resulted in a small base of expertise being available, and is a serious problem given the length of time it takes to access needed skills and train personnel. The lack of stable funding has also resulted in the loss of personnel to other nations able to provide sustained and competitive salary support.

In addition to the invaluable services provided by the technical personnel (programmers, analysts, etc), it will be key to provide researchers and their students with: 1) training, 2) information as to the services available, 3) support for research and industry partnerships by connecting research needs with resources, and 4) assisting the implementation, resource allocation, personnel and maintenance activities associated with each provider. The proposed resources will provide the HPC Resource Providers with the critical expertise pool needed to effectively use Canadian HPC investments.

---

[15] Computing statistics are taken from Top500.org, and GDP figures for 2003 from the CIA World Factbook 2004.

[16] The Figure shows Canada could expect to stand assuming the proposed Canadian investment for the year 2006. It does not adjust for the fact that these nations will also continue to move ahead with HPC investments.

[17] It is important to note that the personnel that this proposal would fund will not be working for individual researchers. Canadian researchers will still be responsible for funding their own research programs. However, the HPC personnel will be able to guarantee that these researchers have access to, and can get the most benefit from, the Canadian HPC research environment.

Table 5.3: Average Staffing Required to Effectively Support one HPC Facility along with Total Cost

Type of position	**2006**	**2009**	**2012**
Manager	1	1	1
Administration	1.5	2	2
Technical HPC Personnel	14.5	16	17.5
Communications and Outreach	1	1	1
Training	2	2	2
Total Staff positions	20	22	23.5
Number of Resource Providers*	8	8	8
Average salary and benefits (2% increase per year beyond basic cost of living)	$80,000	$84,900	$90,000
Total Human Resource $'s**	**$13,000,000**	**$15,000,000**	**$17,000,000**

* We aggregate the single-University Resource Providers as one Facility, since they are small in comparison to the six others which each support multiple universities. The High End facility is likewise included as one Resource Provider.

** All numbers are in 2004 dollars

Table 5.4: Summary of facility costs

Year	2006	2009	2012
Facility Costs*	$8,000,000	$9,000,000	$10,000,000

* 2004 dollars

Operational (Facility) Resources

The facility costs of running an HPC centre have risen dramatically in the past few years, in large part due to the increasing price of electricity. These systems are power hungry and require extensive cooling equipment. In addition, there are the other sizable costs mentioned above, such as networking, data archiving, security, uninterruptible power supplies, and so on. Depending on the hardware configuration purchased, the *total facility expenses* can vary; however, a survey of the Canadian HPC Resource Providers indicated these annual operating costs were roughly 15% of the annualized hardware investment. This number should be viewed as volatile, due to the unstable energy market. Based on current trends and the increase in infrastructure assets, we predict facility expenditures to grow from $7 million today (2004) to $10 million per year by 2012 as shown in Table 5.4. (These figures assume stable energy prices).

IMPACT Canada: Coordinating HPC needs across the country

We anticipate that IMPACT Canada will have many responsibilities, including national coordination of HPC, government and industry liaison, HPC technical leadership, promotion of HPC, the establishment of a national and international presence, and the development and delivery of outreach and training material. The agency will have a three-pronged mandate of doing reporting (past), developing business plans (near term), and creating strategic plans (future). We initially envision a staff of ten people led by a CEO who will answer to a Board of Directors. Within three years, we estimate the staff to grow to 19 people, mostly on the technical side. This requires a dedicated investment of $1 million per year by 2006, growing to $2 million per year by 2009. (Table 5.5)

Recommendation for Funding

The HPC funding strategy proposed here will allow Canadian Universities and their partners and collaborators to gain access to competitive HPC resources. These resources will address immediate HPC needs, and allow Canada to maintain its research capacity.

5.1 We strongly recommend that a long-term funding envelope be developed to support HPC in Canada. We anticipate that this funding could be phased in over a period of six years as follows (Table 5.6):

The analysis does not accommodate all of the smaller computation-based proposals that are currently being funded by the CFI. If the Long Range Plan is to include these costs also, the resources committed will have to be expanded by 10-15%.

Table 5.5: IMPACT Canada – Number of staff positions

Type of position	**2006**	**2009**	**2012**
Chief Executive Officer	1	1	1
Administration	1	2	2
Technical HPC Personnel	6	14	14
Communications and Outreach, Training staff	2	2	2
Total Staff positions	10	19	19
Average salary and benefits*	$80,000	$85,000	$90,000
Operating costs **	$200,000	$300,000	$300,000
Total Personnel Costs	**$1,000,000**	**$1,900,000**	**$2,000,000**

* All numbers are in 2004 dollars. Increases include, for example, merit increments (2% per year beyond basic cost of living)

** Operational costs associated with positions. These include travel, training materials, outreach activities, minor upgrades, etc. For example, a week-long training workshop for 25 to 40 people can cost on the order of $40,000. All of C3.ca's current activities are subsumed.

Proposed Mechanisms to Disseminate Funding

The allocation of HPC funds should be based on periodic proposals that clearly demonstrate excellence (internationally competitive research), importance (intellectual, economic or societal) and compatibility and coordination with computing capacity across Canada. There are various ways that this new program could be implemented using existing mechanisms, for example: 1) designating new funds within the CFI program or within a special tri-council initiative such as the Canada Research Chairs directorate; or 2) establishing a separate funding envelope to *IMPACT* Canada. Irrespective of the mechanism, it is critical that there be a single funding envelope to cover both capital infrastructure and operating funds (both human infrastructure and facility costs) as these are intrinsically connected.

We believe that the long-term funding initiative should be re-evaluated every five years as part of a national review of Canada's international research competitiveness. Reviews of HPC expenditures, and

Table 5.6: HPC Funding breakdown for 2006 to 2012

Type of position	2006	2009	2012
	(Funding in $ millions[1])		
HPC capital infrastructure: Consortia	44+13*	49+14.5*	54+16*
High-end (tera-scale) facility	10+3*	12+3.5*	14+4*
HPC operations: Human infrastructure	13+4*	15+4.5*	17+5*
Facilities	8	9	10
IMPACT Canada	1	2	2
Total public contribution:	**76**	**87**	**97**
Total industrial contribution*	20	22.5	25

1 In 2004 dollars (no allowance for inflation)

* Industrial contribution

> *"Without adequate HPC facilities in a country or region, scientists and engineers will struggle to compete with leading edge researchers in other countries, and some of them will actually migrate to other countries that provide the facilities to support their work."*
>
> ***– Australian High-Performance Computing Project, Feasibility and Capability Study Report for the Illawarra Regional Node.***

a technology watch, would separately be undertaken by IMPACT Canada on a more regular basis. The following performance evaluation criteria should be considered for this review: *academic excellence* (refereed publications, national and international awards, and retention and recruitment of faculty); numbers of *qualified personnel produced* (graduate students, postdoctoral fellows, technicians and research associates); and measurable *societal and economic impacts* (patents, spin-offs, industrial collaborations and partnerships, technology transfers, improved public infrastructure, positive health outcomes and educational outreach activities). The review should also consider the effects of the investments in HPC on Canada's international competitiveness.

It is anticipated that the facilities provided by adopting the recommendations in this Long Range Plan will address the majority of HPC needs (large and small), and reduce some of the pressure on CFI and the granting councils. We recognize that any requests for HPC must be evaluated in the context of whether or not the existing HPC facilities can already meet the need. However, we also recognize that there will be researcher requests for specialized high performance computing equipment to CFI or the tri-council funding agencies that do not properly fall into the normal production mode of the proposed HPC facilities. In such cases, these research programs will need access to their own equipment. This leads us to the following recommendation:

5.2 **We recommend that funding requests for HPC computer resources be evaluated in the context of what can be provided by existing HPC facilities, either general-purpose consortia or domain-specific facilities, so that the most cost-effective use can be made of investments in HPC. Requests that genuinely**

cannot be reasonably satisfied using this existing infrastructure should compete on their merits for existing sources of funds such as conventional CFI envelopes or equipment funds from the granting councils.

"To get the maximum leverage from the national effort ….

The Congress should provide adequate and sustained funding".

Recommendation 1 of the US Committee on the Future of Supercomputing (National Research Council) in their publication: Getting Up to Speed: The Future of Supercomputing

We recommend that funding requests for HPC computer resources be evaluated in the context of what can be provided by existing HPC facilities, either general-purpose consortia or domain-specific facilities, so that the most cost-effective use can be made of investments in HPC. Requests that genuinely cannot be reasonably satisfied using this existing infrastructure should compete on their merits for existing sources of funds such as conventional CFI envelopes or equipment funds from the granting councils.

Summary

HPC funding in Canada must accommodate the need for mid-range and advanced HPC infrastructure, investments in people and operations, and the establishment of IMPACT Canada. Funding must take into consideration the need for stable sustained funding to allow for long term planning, evolving needs, demands from industry and a coordinated approach.

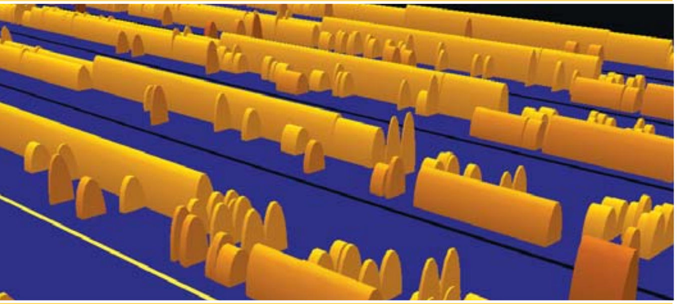
In this chapter, we have outlined Canada's needs for HPC funding over the next 8 years, within the context of key funding dimensions: continuity, growth, competitive research infrastructure, capital and operational investments, and planning for the future. We anticipate that Canada must invest $76 million annually in 2006, increasing annually to $97 million by 2012, in infrastructure, human and operational resources. These resources will provide Canadian researchers across the country with the tools they need to excel, and will ensure that there is a development and implementation continuity that is crucial to their success.

HPC is not a passing fad. As shown in previous chapters, the technology has become pervasive, growing from a handful of HPC sites worldwide in 1984 to thousands today. Furthermore, the past decade has seen the technology embraced by industry. The number of companies on the Top 500 list has grown from 3% in 1994 to over 50% today. These companies run the gamut of products and services, including the traditional HPC product areas such as automotive, aeronautics and petroleum as well as the non-traditional areas of entertainment, financial services and retailing. High performance computing is a universal tool, a cost-effective shared virtual research laboratory for all areas of research and development that is supporting a continually growing diversity of applications.

Canada itself will not only benefit from the research outcomes enabled through this investment, but from the jobs and expertise that this investment produces. Canada must, therefore, continue to invest strategically and effectively in order to maintain our HPC based research strengths, and maximize the benefits of the investments already made. Our failure to invest now will seriously damage Canada's future research and economic competitiveness.

"The rapid dissemination of the SARS genomic sequence through the Internet did allow groups all around the world to participate in the computational analysis of the genome and the 3D modeling of the protein sequences. Ultimately, resulting in an unprecedented number of scientific publications and vaccine initiatives just over one year after the discovery of this emerging disease".

Dr Steve Jones
Head of Bioinformatics
Genome Sciences Centre
Director of Bioinformatics at Genome BC
University of British Columbia



This diagram displays the genetic makeup of the human chromosome illustrating genes and other critical information. This graphs allows scientists to review genetic information from human chromosomes in hopes to better understand and develop vaccines for infectious diseases, like SARS, which has drastically affected the human population in the last two years.

HPC Helps to Fight Epidemics in Human Populations

As summer descends on North America, West Nile Virus is anticipated to cause widespread disease and chronic neurological problems. The summer months will also be accompanied by increasing problems with malaria in North America and other developed countries as global warming heralds fundamental changes to our planet's ecosystems. In Africa, bacterial meningitis is reaching epidemic proportions. It is of great concern that antibiotic drugs (the miracle agents of the 20th century) are becoming increasingly ineffective against microorganisms. The spectre of SARS (Severe Acute Respiratory Syndrome), with accompanying hospital quarantines emphasizes the looming importance of infectious disease to the health of our nation and our planet.

The World Health Organization has identified infectious disease as the greatest challenge to global health. A recent NIH report identified resistance to conventional antibiotics and the need for new antibiotics as crucial. The "To Market, To Market" reports of the American Chemical Society (which document all new drugs introduced for human use on a yearly basis) verify this and reveal that the majority of "new" antibiotics being developed by drug companies primarily continue to be penicillin and cephalosporin derivatives. A Kalorama Information study (The Worldwide Market for Antibacterial Drugs) concludes that there has been only one new class of antibiotics in the past 30 years (oxazolidinones) and that there is an immense need for "new" antibiotics to service the $30Billion market.

HPC enables Canadian scientists to tackle these crucial world health issues. Through the use computer modelling and simulations researchers can better understand human genetic sequencing and how these sequences react to dangerous pathogens. Through analysis of these models and sequences scientists hope to develop improved control strategies for infectious diseases.

HPC must be supported in order to stimulate research methods to contain or eradicate infectious disease like SARS. Millions of dollars are spent preventing and treating measles and other similar infectious diseases but support is needed to ensure research facilities have the tools they need for developing vaccines to prevent outbreaks. HPC plays a fundamental role in permitting these dreaded diseases to be attacked. This research can become a global health reality by supporting HPC resources in Canada to allow research to continue to discover preventions and treatment for the world's most troubling diseases.

Helping Canadians Breathe Easier with HPC

Breathing is critical to human survival. In Canada about 11% of children suffer from asthma and are heavily dependent on the delivery of some form of drug to the lungs to enable them to complete their day-to-day activities. One problem with current respiratory drug delivery is drug waste through deposition in the mouth and throat. If the effectiveness of delivery to the lungs could be improved this would be a major health-care benefit to all Canadians suffering from respiratory ailments.

How can drug delivery be improved to increase the effectiveness of respiratory drug treatment? Through Computational Fluid Dynamics (CFD) researchers have greater insight into why drugs are poorly delivered to the lungs through simulations. These simulations provide data, which can effectively help physicians and pharmaceutical develop improved respiratory treatments. Additionally, CFD minimizes the need for human test subjects.

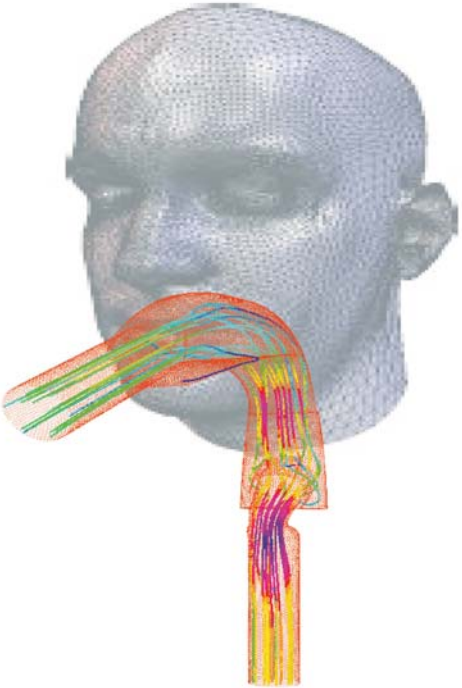
Enhanced drug delivery to the lungs can only be realized with the evaluation of good data generated by high performance computing infrastructure. These simulations need to be produced in a timely manner to be effectively appreciated by physicians and the pharmaceutical industry. That can only be achieved by access to the absolute best computing environments.

CFD is a major driver for manufacturers of HPC and related devices. Canada must invest in world-class computing facilities if it is to remain a leader in CFD technology and its increased use in industry. In fact, CFD is used everyday in almost every industry in Canada: aircraft, aero-engines, power generation including fuel cells, automotive, petrochemical, and the steel and aluminum industries.

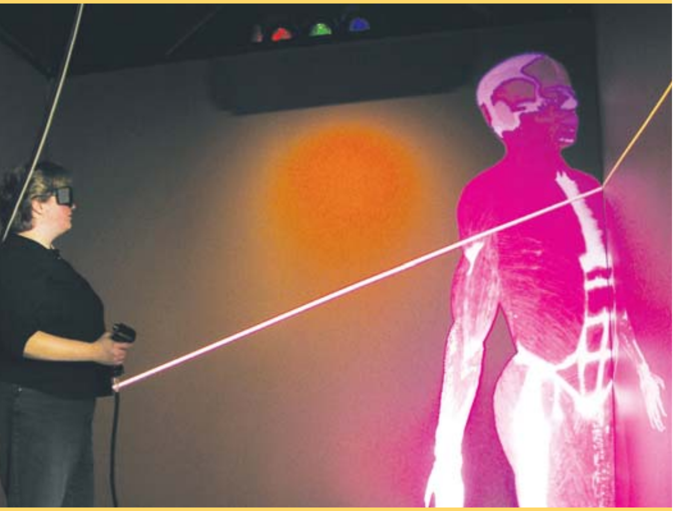
HPC enables many of these industries to compete internationally. More importantly, universities continue to be the incubators for the applications in common use today; and they continue to be inventors of new applications for use in new emerging areas, such as nano-scale technologies. Growth in demand for people with CFD expertise continues.

"As a CFD vendor and particularly for those customers we serve around the world, an over-arching need in Canada is an increased pool of highly qualified personnel and highly advanced computing resources"

Dr Jeff Van Doormaal
Team Leader, Product Development
ANSYS Canada Ltd
and founder of Advanced Scientific Computing Ltd.



CFD model of inhalation through a simple metered dose inhaler. [REF 34]



Through the visualization of genomics, researchers are able to see how the genetic code is related to the functions of the human body. By visualizing these genomic functions researchers have the ability to pinpoint potential problems in genetics and how they impact the human physiology.

Unlocking the Language of Life

Can you imagine receiving medical treatments customized precisely to your body's genetic makeup? Not having to worry about an adverse reaction to a prescription? Your cholesterol put back in balance?

The determination of the human genome sequence is considered the single largest scientific achievement of the 20th century. Today scientists are taking the necessary experimental, computational and theoretical steps to exploit genome sequence information to better understand the causes of disease, viruses, and the human genetic code. This involves taking genetic information, finding fragments, organizing and rearranging them – to provide medical interventions and preventions to promote wellness in Canada.

Such audacious research will be one of the greatest achievements of our century. It could not be undertaken without massive computing power. Manually trying to decode and understand the genetic code is virtually impossible. High performance computing resources are critical for Canadian researchers to expand a valuable genetic database, shaping the physiological equivalent of an alphabet, text segments and sentences. The ongoing challenge is to assemble these letters, sentences and text segments into the language of life. So why is this breakthrough significant? Because unlocking the genetic code – and the genomic research that follows – allows for the timely development of vaccinations to combat viruses like SARS, offers development of cures for cancers, cystic fibrosis and HIV/AIDS and even prevents strokes and introduces gene therapy. The benefits of this program are truly staggering for the health, wellbeing and productivity of Canadians and our global society. Supporting our Canadian genomics researchers supports our own future wellbeing and the economic viability of Canada's healthcare system.

Canada is aging and with the power of HPC resources and the ability to harness genetic information we can discover, create and develop some of the best techniques, medicines and preventative strategies to support our aging population. Health care is Canada's pride so marrying the research and development in genomics with High Performance Computing resources we can provide the best health care service possible to all Canadians.

> *"The human body is undoubtedly the most complex machine ever created. Genome researchers are undertaking the challenging task of unraveling how it is organized and how it functions. High performance computing plays a dominant role in this research. Without extremely sophisticated computational infrastructure, Genomics research would not be possible."*

Dr Christoph Sensen
Professor of Biochemistry and Molecular Biology
University of Calgary

Supercooling Liquids with Supercomputing

Solidification is a familiar and important process. Water freezes to ice. Molten silicates are mixed in a glassblower's furnace and then cooled to create glassware. Ice and glass are both familiar solids, but quite different in their molecular structure. Ice is a crystal, a regular array of molecules; glass is an amorphous solid, a random, disordered array of molecules (see figure). Why would one liquid create an ordered material, while the other doesn't?

Surprisingly, the answer to this basic question continues to be debated by physicists, chemists and materials engineers. A general way to predict the structure and properties of the material created by solidifying a given liquid does not yet exist. Virtually all liquids are at least momentarily "supercooled" before solidifying, that is, they have a temperature less than the melting temperature of the crystalline form. So understanding solidification begins with understanding supercooled liquids at the molecular level, with particular emphais on how its structure and properties are affected by molecules colliding and interacting under different conditions leading to solidification.

High performance computing resources provide an ideal virtual laboratory for obtaining a detailed microscopic picture of these processes. Model systems of thousands of molecules – studied over thousands of different conditions – allow theories of solidification to be tested, and allow new experimental results to be clarified. Crucially, HPC simulations also allow exploration of totally new conditions that have not yet been considered in theories, or explored by experiments.

The aim is to predict, and ultimately control, the process of crystal and amorphous solid solidification. This is critical for enabling research and development that today is experimental, but that five years from now could enhance production quality and the creation of new industrial materials.

"Computer simulations using HPC facilities are a central tool for Canadians researching the creation of novel materials. For the solidification problem, HPC based research is really the only way to study in detail the molecular motions occurring during solidification, and to explore temperature and pressure ranges presently inaccessible to experiments. The results are critical for testing competing theories, and to allow us to better predict and control the solidification process".

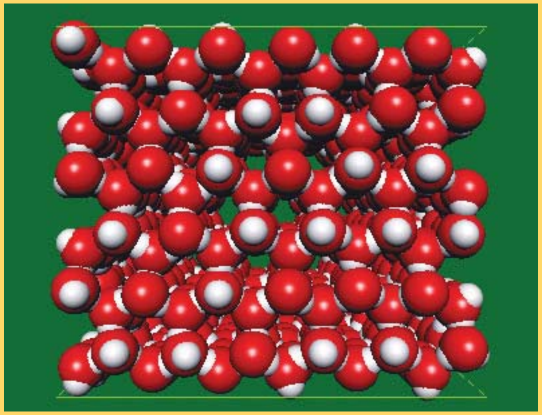
Dr Peter Poole

Canada Research Chair in Modeling

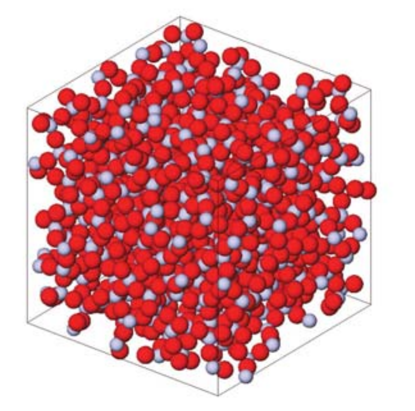
and Computer Simulation

Professor of Physics

St. Francis Xavier University



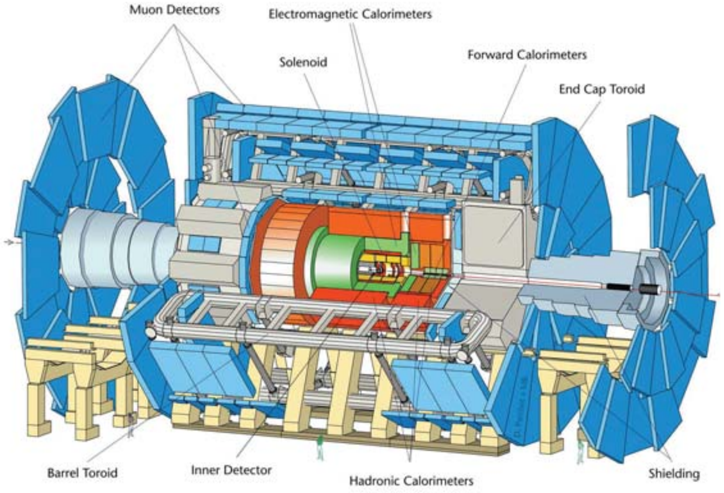
Ordered, crystalline structure of ice, as modeled in computer simulations. Red spheres are oxygen atoms, white spheres are hydrogen atoms.



Disordered, amorphous solid structure of silica glass, as modeled in computer simulations. Red spheres are oxygen atoms, grey spheres are silicon atoms.

"High performance computing is an essential tool used in the analysis of high-energy physics data. The size of the data sets from upcoming experiments will be unprecedented, requiring an international grid of HPC facilities. HPC is one of the enabling tools to unlock the secrets of the fundamental structure of our Universe."

Dr Mike Vetterli
Professor of Physics
Simon Fraser University and TRIUMF



Schematic of the ATLAS detector. The dimensions are: diameter= 25 meters, length= 46 m, weight= 7000 tons. One can get a feeling for the size of ATLAS by considering the tiny green figures at the base of the detector, which are people.

Pushing HPC to the Edge to Understand the Universe

Particle physicists have been researching the fundamental structure of matter for decades. For example, physicists search for the mechanism that is responsible for generating the mass of elementary particles and by extension of all matter, and seek to understand why there is more matter than anti-matter in the universe. This research addresses fundamental questions related to our universe and will shed light on its origin and its eventual fate.

Particle physics experiments use very sophisticated detectors to essentially take a picture of the results of subatomic collisions. The response of these detectors is recorded for each collision so that the experiment can be "re-played" over and over with ever improving analysis techniques.

Canada is heavily involved in this arena through the LHC/ATLAS project, a greater than one billion dollar investment by the international research community to construct the most powerful proton-proton collider in the world. The high performance computing resources required to analyze the collisions in ATLAS will be immense. It is anticipated that ATLAS will produce several petabytes of data per year. No single HPC site in the world can provide for the storage, management and analysis of these data. Instead, the project has adopted the novel solution of creating a worldwide computational collective, with many countries contributing massive computing resources.

This research gives the world insight into the fundamental nature of our universe. As such, it is one of the grand intellectual challenges of our time and Canada can be at the foreground of research in supporting these discoveries.

As a bonus, this research can lead to economic developments in technology products. Particle physics experiments push technology to its limits, leading to new products and techniques, which find various uses in society. Some examples are diagnostic tools for medicine, such as Positron Emission Tomography (PET), treatments for cancer using radiation, the use of radio-frequency techniques to quickly dry lumber, and maybe the most famous spin-off from high-energy physics, the World-Wide Web.

HPC in the Digital Content Creation Industry

Canada is a world leader in computer graphics, well represented by academic research and industry application software companies. Alias, Side Effects, and Softimage are the major Canadian international success stories in the 3D software tools marketplace, now complemented by artistic, technology and production expertise for feature filmmaking.

Canada is a "near-shore" destination for live-action productions already, and in 2003 the Computer-Generated-Imagery (CGI) feature animation industry got a toehold in Toronto with C.O.R.E. and IDT announcing CGI feature productions. In the same way that skilled crews and artistic talent draws producers to shoot live-action pictures in Canada, the availability of suitable production studios will draw CGI pictures here for near-shore production.

A typical compute cluster or "render farm" for CGI features consists of more than 500 dual-processor machines. Each frame of film-resolution animation may take between 2-4 hours to compute. If a typical CGI feature runs 80 minutes, then you must calculate 115,200 frames X 3 hours/frame = 14, 400 days of processing time (on a single CPU) for only one iteration. Each iteration incrementally adds to the quality of the final CGI image. Although there are many other industries with compute-intensive needs, the data sizes are typically small compared to that of high-end photorealistic CGI data. A single shot with multiple characters and effects may require 100 to 1,000 gigabytes of disk space (comprising both the input data and the output rendered frames). With an average of 1,500 high-end photorealistic shots in an 80-minute feature, the demands on the infrastructure are enormous, and the payback period is limited to the duration of the picture (typically 2-3 years, after which the infrastructure will be refreshed completely).

Greater throughput means the artists and technicians can run more iterations of a given shot, resulting in a better-looking final product. The more iterations a farm is capable of running, the better the director's creative vision can be realized. As audiences become more sophisticated, the director is constantly asking for more and more realistic animation and rendering, as well as visual effects. This is accomplished due to the ready availability of the artistic talent (especially from Sheridan College's world-renowned computer animation program) and computer graphics scientists who are required to create custom code for visual effects and other special-purpose elements that can increase the realism (e.g., ocean waves and spray in the movie *The Perfect Storm*) or novelty (e.g.,"Bullet Time" in *The Matrix*) of a given sequence. To stay at the top of this market, we must provide flexible, powerful computing solutions that support creation of the most realistic imagery in the shortest period of time.

> *"Access to HPC is a pre-requisite for computer-animated features. Since the market for CGI features is growing rapidly, more than the average 35% yearly growth in the digital content creation market overall, companies like C.O.R.E. face a very competitive market chasing somewhat scarce resources."*
>
> ***Tom Burns***
>
> ***Core Feature Animations, Toronto***

Frequently Asked Questions (FAQs)

The issues raised by the following questions are dealt with in detail through the body of this report. They are listed here to illustrate the essential rationale behind the report. References to sections dealing with these issues are provided (e.g., E2 – Section 2 of Executive Summary; C2 – Chapter 2).

If HPC is so beneficial, why can't industry support the infrastructure?

Industries develop mature applications derived from basic, curiosity-driven academic research. Few corporations can afford to invest in this critical early discovery phase. Also, many HPC applications are in the public domain: infrastructure (transport), big science (meteorology, astronomy, physics), government jurisdictions (fisheries, health, education). Increased support for HPC in the public domain would be primarily a government responsibility.

(Relevant Sections: Neurochem Case Study, "One Day" Story)

Why can't the researchers apply for HPC funding through existing mechanisms?

Some funding mechanisms exist for mid-range HPC infrastructure (e.g. CFI) but there is no long-term coordination for a national HPC strategy that oversees equipment, personnel and infrastructure, and provides the linkages between government, academe and industry. There is no mechanism for securing the large investment needed for a top-end HPC facility in Canada. Finally, the rapid timescale required for refresh of the technology is not well suited to current funding mechanisms.

(Relevant Sections: E2 ; E3 ; C3-Funding for HPC Support Staff ; C4 All ; C5 Introduction)

Why can't we just link up lots of increasingly-powerful, cheap desktop computers to handle the big problems?

This is possible if the problem can be broken down into many self-contained components, each of which can be handled by one desktop computer. However, there are many important research applications which simply cannot be scaled in this manner, since they have CPU, memory, data, or communications needs that far exceed that of a desktop computer. Such problems can only be solved by HPC equipment.

(Relevant Sections: E1 ; C2 – Capability and Parallel Applications sidebars ; C2 – Central Rings section)

Isn't Canada too small a population to compete in this field?

A number of smaller countries have invested in HPC strategies, recognizing this is a key engine of future economic development. Moreover, Canada has unique characteristics that make it ideally suited for HPC leadership. Canada is not so small that it lacks the resources (human, financial, knowledge) to make a significant impact on the world stage. Neither is it so large that national efforts are so daunting that they tend to break up into incompatible, regional initiatives. This principle has already been demonstrated with CANARIE's CaNet*4 network. Canada's vast extent but limited number of major cities makes it ideal for a coordinated computational grid which combines the high-speed data communication of CANARIE with the HPC resources described in this document. Much like the railways of the last century, such a national grid would bind together Canadian science and society in a manner that few other countries could match.

Moreover, external metrics demonstrate that Canada's research community ranks second among G8 countries (behind the UK) in research productivity. Canada is well-placed to be a leader not a follower.

(Relevant Sections: C1-CANARIE Sidebar ; C2 Networks and Grids section and Sidebars ; C3 Scientific Productivity Sidebar)

Why invest in expensive HPC ? Why not simply wait for other countries to do the basic research while we concentrate on development and production phases ?

The principal benefits of information technology accrue to those who develop it first. If Canada waits for other countries to make the breakthroughs, then we will inevitably fall behind in the race to bring new knowledge to the market-place. This has never been more true than in this field. Imagine if Silicon Valley had grown up in a Canadian city…

(Relevant Sections: C1 Introduction)

Why can't we just buy time on computers in other countries?

This is ultimately more expensive and counter-productive to the development of Canada as a leader in high-end research and development. As discussed in Chapter 2, the true cost of running one application on a tera-scale facility may run from $100K’s to $M’s. This is all money lost to Canada. An offshore HPC strategy, if scaled up to a national level, would cost more than the in-house strategy proposed here. The use of the Japanese Earth Simulator to model a hurricane weather system is an illustration of the potential benefits and pitfalls. The true cost of this work was $5-10M; the Canadian team was able to provide technical resources instead of payment but still suffered from low priority access and onerous international bureaucracy.

After 9 months of use the work is still only about 40% complete due to low priority. Buying CPU time, if it could be secured, would be a signal that Canada is not a significant player in the G8. We would lose many young computational scientists to other countries since they could not compete from Canada due to lower priority access than their peers in the country providing the facility.

(Relevant Sections : C2 The Central Rings section and footnote ; C2 Earth Simulator Sidebar, C4)

What about provincial support ?

It would be desirable for provincial governments to participate in this national strategy. This would allow them to augment the national HPC infrastructure in ways that address specific provincial HPC priorities. Addressing these individual perspectives was beyond the scope of this plan and would require discussions between the two levels of government.

References

[1] Diagram provided courtesy of Service météorologique du Canada, Recherche en prévision numérique

[2] For more information, see www.googleguide.com

[3] For more information, see www.aerospace-technology.com

[4] Neuromancer page 51, William Gibson, 1984, ISBN 04-441-00068-1

[5] http://www.dgp.toronto.edu/~patrick/ryanTech/index

[6] For more information, see www.brainrepair.ca

[7] For more information, see www.bcgsc.ca/bioinfo/SARS/index_html/view

[8] For more information, see www.projectcybercell.ca

[9] For more information, see http://www.oecd.org/home/

[10] Images provided courtesy of Bombardier Aerospace

[11] Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003, www.cise.nsf.gov/news/special_reports/cyber

[12] "Unraveling of SARS VIrus Relied onTech Giants", Bio-IT Bulletin, April 15, 2003 www.bio-itworld.com

[13] Copyright owner: Intel ; www.intel.com/research/silicon/mooreslaw.htm

[14] Source: TechDecisions, September 1999

[15] For more information, see www.top500.org

[16] © 2002.Time Inc. Best Innovations of 2002. Reprinted by permission.

[17] Earl Joseph, Addison Snell and Christopher Willard "Council on Competitiveness Study of U.S. Industrial HPC Users", www.compete.org

[18] Images provided courtesy of Virgo consortium

[19] Sources: HEC and Scientific Opportunity, Martyn Guest, 2003; 21st Century High End Computing, D. H. Bailey, NERSC, 2004

[20] "Achieving Excellence: Investing in People, Knowledge and Opportunity: Canada's Innovation Strategy." Government of Canada, 2002, available at www.innovationstrategy.gc.ca p. 19.

[21] "Knowledge matters: Skills and Learning for Canadians: Canada's Innovation Strategy." Government of Canada, 2002, available at www.innovationstrategy.gc.ca.

[22] David A. King. "The Scientific Impact of Nations", Nature, vol. 430, July 15, 2004, pp. 311-316. Reprinted with permission

[23] CISS: http://www.cs.ualberta.ca/~ciss/

[24] Trellis Project: http://www.cs.ualberta.ca/~paullu/Trellis/

[25] From: "HPC-EUROPA (2004-2007), a project funded by the European Commission Support for Research Infrastructures Programme." By C. di Reno (28 Jan 2004). Accessed Sept 10, 2004. Source: http://www.hoise.com/primeur/04/articles/monthly/AE-PR-03-04-27.html

[26] From: Academic Supercomputing in Europe. Accessed Sep 10, 2004. Source: http://www.arcade-eu.org/academicsupercomputing/europeanprogrammes.html

[27] IDC Sees Big Opportunity for Grid Computing, By Paul Shread, April 27, 2004. Found at: www.itmanagement.earthweb.com/erp/article.php/3345931 – Sep 9, 2004

[28] "Grid Technology Partners issues report on Global Grid Computing market." June 2002, Source: http://www.hoise.com/primeur/02/articles/live/AE-PL-06-02-4.html. Accessed Sep 10, 2004.

[29] Images provided courtesy of the McConnell Brain Imaging Centre

[30] J. Allard, P. Grochulski & J. Sygusch, PNAS, 98, (2001) 3679-3684

[31] From Making the Connection: Research and Practice in Undergraduate Mathematics, M. Carlson and C. Rasmussen (Eds), MAA Notes, in press.

[32] http://www.fin.gc.ca/budget04/pamph/paecoe.htm

[33] Source: http://www.supercomputingonline.com/article.php?sid=7505 (Accessed October 2004)

[34] Pollard et. al, Mechanical Engineering Queen's University at Kingston, © 2003