
# *High Performance Computing Development for the Next Decade, and its Implications for Molecular Modelling Applications*

Sectoral report from the HPC Technology Roadmap Study within the ENACTS project

Jan Fagerström, Torgny Faxén, Peter Münger,
and Anders Ynnerman,
NSC

J-C Desplat
EPCC

Filippo De Angelis, Francesco Mercuri, Marzio Rosi,
Antonio Sgamellotti, Francesco Tarantelli,
and Giuseppe Vitillaro,
CSCISM

April 25, 2002

**Abstract**

Technological and economic trends determining the development of High Performance Computing (HPC) hardware architectures and systems in the time range 2006–2011 is investigated, as well as its implications for scientific applications in the area of molecular modeling.

In part I of the report, information regarding the HPC development has been obtained by means of interviews with representatives of 7 vendors of HPC systems (Compaq, Sgi, Sun, IBM, Cray, NEC and Fujitsu). The interview material is compiled in a concise form, and conclusions are made regarding the likely development of hardware, architectures and systems in the HPC area. In the report, a summary of each interview is presented followed by an overview of all interviews and some general conclusions. Vendors are often portrayed as having quite different views on supercomputing, especially in areas such as processors, vectorisation, memory architecture and programming models. Interestingly enough, the interviews showed that there is general agreement on many of the major trends that will be important for future computer architecture in general and HPC computing in particular. Most future computer systems can very generally be described as scalable parallel architectures based on the notion of clustered SMPs. When looking "under the hood" however, there is actually a wide array of different solutions presented. New advances in all relevant computer technologies, made possible by an increasing consumer market as well as the Open source movement, seems to make the number of available solutions even larger than before. Actually, if anyone believed that future HPC systems would all move towards a homogenous computer architecture, it seems likely that they will be disappointed.

Part II of the report provides an assessment of the implications of future computer hardware and architecture for the key molecular science community. The assessment is made by several experts on molecular applications at CSCISM. Using the results from the vendor interviews in part one, two typical future HPC solutions are configured. This is used when investigating the impact of future computer HPC development on three main approaches to the definition of interatomic potential: 1) Model potential based on Force-Field (FF) parameterizations, 2) DFT (Density Functional Theory) derived potentials, and 3) Ab initio potentials.

# Contents

# Chapter 1

# Introduction and purpose

This report presents the results obtained within the HPC Technology Roadmap study within the ENACTS [7] project. ENACTS (European Network for Advanced Computing Technology for Science) is a European project funded by the European Union. It is a collaboration between 14 centres for high performance computing in 13 countries across Europe.

The aim of ENACTS is to evaluate future trends in the way that computational science will be performed and the pan-European implications. One specific goal is to enable the formation of a pan-European HPC meta-centre. To this end several studies will be undertaken within ENACTS: six studies in a first phase focusing on key enabling technologies, and four studies in a second phase studying practical implications of the results from phase one, dissemination of the project results and preparation of the case for capital investment in improved research infrastructure. The project started in February 2001, and the present report presents the results from the second study within ENACTS: The HPC Technology Roadmap.

The HPC Technology Roadmap study is undertaken by the National Supercomputer Centre (NSC) in Linköping, Sweden in collaboration with Center for High Performance Computing in Molecular Sciences (CSCISM) in Perugia, Italy. The objective of the study is to determine the likely technology and economic trends, which will prescribe the hardware architectures of HPC systems over the next 5 to 10 years, and to evaluate the effects that this will have on applications software. The work within the study has been shared between NSC and CSCISM in the following way. NSC has provided a survey of the technology roadmap for processors, memory, networking (closely coupled and LAN), data storage, custom-built solutions, and software paradigms and standards. This survey was accomplished by interviews with several major HPC vendors, and is reported in section I. NSC has also coordinated the study. Based on the results of the technology roadmap survey CSCISM has provided a case study focusing on the usefulness and implications of the technologies discussed in the technology roadmap, for the key molecular science community. The case study is presented in section II.

Up-to-date information about the ENACTS project and the studies performed in the project is continuously published at the ENACTS web site.[7]

# Part I

# Technology watch report: Trends in HPC development

by Jan Fagerström, Torgny Faxén, Peter Münger, Anders Ynnerman, and J-C Desplat

# Chapter 2

# Methods

The purpose of this technology watch study is to determine the likely technological and economic trends which will prescribe the development of HPC hardware architectures over the next 5–10 years. There are several possible methods which can be used for such a study. Perhaps the first method which comes to mind would be to perform literature searches of technical reports, articles etc. Information regarding HPC technical and economic trends would be distilled from the search results and compiled in a readable form, from which conclusions could be made. However, it was decided that a more attractive way to perform the present study would be by means of interviews with major HPC vendors. There are several reasons for this choice:

1. The HPC area develops rapidly. It is therefore necessary that the information is up-to-date.
2. Different ideals and visions in the area of HPC architectures and development, as represented by different HPC vendors, can be compared on an equal footing, at an equal point in time.
3. The interview format allows for follow-up questions, and elaborations on especially important points, as well as in areas where the interviewee has particular expertise knowledge.
4. The interview format makes it possible to get a response from the interviewees, on the conclusions made from the interview material, thus increasing its relevance.
5. The material is exclusive, and obtained for the particular purpose of this study, and therefore more interesting and relevant.
6. The interview format is personal which opens up the possibility to also present visions and speculations about the future, apart from pure facts and extrapolations of figures.
7. From the point of view of the authors, it is more interesting to work with interviews than with literature searches, since it includes the possibility to make new contacts.[1]

[1] It is likely that it has also resulted in a lot more work.

Initially eight HPC vendors (Compaq, Sgi, Sun, IBM, Cray, NEC, Fujitsu and Hewlett-Packard) were contacted and invited to participate in the interviews with one or several representatives of their own choice. Representatives from the first 7 vendors in the list participated in the study. The interview programme is shown in Table 2.1. In retrospect,

Table 2.1: The programme of the vendor interviews. All interviews where performed during the summer of 2001.

| **Date** | **Vendor** | **Interviewee** | **Place** |
|---|---|---|---|
| June 8 | Compaq | Dr. Martin Walker | NSC |
| June 11 | Sgi | Dr. Wolfgang Mertz | NSC |
| June 18 | Sun | Mr. Benoit Marchand | NSC |
| June 19 | IBM | Dr. Jamshed Mirza | NSC |
| June 20 | Cray | Dr. Burton J. Smith | NSC |
| June 21 | NEC | Mr. Tadashi Watanabe | International Supercomputer Conference 2001, Heidelberg |
| August 22 | Fujitsu | Dr. David Snelling | NSC |

it is clear that much of the information expressed in the interviews is more coloured by the visions and opinions of the individuals being interviewed, rather than by the official vendor programs. This is very satisfying, considering the fact that all persons being interviewed are notable general experts in the field of high performance computing.

The interviews were performed in the following way. In order to present a starting point for the discussions during the interviews it was decided to use the results from the ENACTS user questionnaire produced by the Grid Service Requirements study.[25] Since the report from this study was not available at the time of the interviews, the results were compiled in simple tables and diagrams for use at the interviews.[2] In addition to the questionnaire material a number of specific questions were prepared, see Appendix A. The results from the ENACTS questionnaire and the specific questions were sent out to the interviewees a few weeks in advance of the interviews.

The character of the interviews varied depending on, e. g., the kind of preparation of the interviewee, and the available time. Some interviews followed the specific questions in Appendix A rather strictly, while other took the starting point at view graph presentations prepared by the interviewees. However, all interviews were characterised by an open discussion.

NSC was represented by at least 2 persons taking notes at each interview. All notes were compiled in a first readable version after the interviews. This first version was sent out to the vendors for revision. Comments and corrections from the vendors were added to the final versions of the interviews, which are the versions presented in Sec-

[2]Support from Dr. J-C Desplat in providing us with the raw data from the ENACTS user questionnaire is gratefully acknowledged.

tion 3. A personal reply to the vendors views on some of the Grid issues discussed in the interviews are given by Dr. J-C Desplat in Section 4. Conclusions made after all of the interviews are presented in Section 5.

# Chapter 3

# Interviews

## 3.1 Martin Walker (Compaq Computer)

Compaq Computer was represented by Dr. Martin Walker at the interview. The interview was held at NSC on June 8, 2001.

Dr. Walker has a background in mathematics and physics. He worked as a researcher for 15 years at the Max Planck Institute for Physics and Astrophysics. In the early 1980's he joined Myrias to work on the building of supercomputers from microprocessors. In 1990 Dr. Walker moved to Cray Inc. to work in the design team for the Cray T3D computer. Since 1996 he has been working at Digital Equipment, which was later acquired by Compaq. He is now responsible for scientific computing at Compaq in Europe.

### 3.1.1 Summary of key ideas of Martin Walker

The most important key ideas conveyed at the interview with Dr. Walker, Compaq, can be summarised as follows:

1. Moore's law for transistor density will continue indefinitely.
2. The pressure from commercial markets on the development of hardware is very high. It is unavoidable that HPC systems will be based on COTS in 5–10 years time from now.
3. The discrepancy between CPU instruction execution time and memory access latency is increasing rapidly. The discrepancy can be resolved in part by deeper memory hierarchies. It is clear that application programmers will have to pay attention to this increasing memory hierarchy depth.
4. Both shared and distributed memory architectures will be used. We can expect an increase of applications using a mix of shared memory and distributed memory.
5. The bandwidth of internal and external networks increases at such a rate that bandwidth limitations will have disappeared as a bottle-neck in a few years time from now.

6. There is no definite upper limit for the size of computer systems. However, large machines will always be partitioned.

7. A majority of all HPC vendors will use a common open source operating system (Linux?) 5–10 years from now.

8. Fortran, C, C++, Java, MPI and OpenMP will continue as the dominating programming languages and parallelization environments 5–10 years from now.

9. It can be expected that open source will dominate in applications and research software. There is also a slow trend, in industry as well as in academia, towards an increased use of $3^{\text{rd}}$ party codes.

10. HPC computer architectures will be influenced much more by commercial than by scientific applications, since the market for commercial applications is at least ten times larger than for scientific applications.

11. We do not yet know what the Grid will become. The Grid can be several different things, and we should try them all! What people actually do will crystallise out in time. Meta-computing might not even be a key part of the Grid.

12. A large European centre is lacking. Europe need to create a large centre in order to influence the vendors.

### 3.1.2 Interview

#### General remarks

The views expressed by Dr. Walker in the sections to follow are his personal views, and not necessarily the official views held by Compaq.

#### Moore's law and hardware development

Moore's law for transistor density will continue indefinitely. There are no technological limitations like power consumption, storage etc. which will prevent this development. According to Kurzweil[32] the doubling time is even getting shorter, such that the development is described by an exponential of an exponential function. This is empirically verified by Kurzweil by looking at technological advances in history from 1900–2000. [The origin of this double exponential function could be the simultaneous increase in a) technological capacity and b) economical investments in research.] It also shows that new technologies emerge when old technologies have reached their limit. What will follow on todays lithographic technologies? Perhaps molecular transistors based on carbon nanotubes.

Thus, transistor "real estate" growth will continue. However, we do not know how all new transistors will be used. Design teams work on this problem. The number of transistors does also cause practical problems: There will be an issue regarding signalling within the chips.

All transistors can not respond to the same clock. Perhaps the clock will be distributed in different clock regions on the chip.

**Computer building blocks**

The commercial market for scientific computing will not be able to support the development of special purpose hardware. Within 10 years from now almost only mass-produced general purpose processors will be used, simply because these are cheaper than low volume special purpose processors. In this way, scientific computing will live off the tail of development in the area of commercial high volume markets. For instance, the development of database engines, an application area which is growing faster than scientific computing, will be of increasing importance for HPC architectures and performance. Another interesting example is the 3-D games market which will provide graphics power needed in scientific visualisation. The movie industry use large computers for rendering applications, but this is no big market. One can conclude that there is no market for special equipment for scientific applications.

Thus, future computer systems will most certainly be built from COTS (Commercial Off-The-Shelf) components rather than proprietary building blocks. Proprietary components have no hope, except perhaps for high performance interconnects. However, vendors will not become system integrators. System integration is not, and will not be, a big market. Rather, vendors will focus on computer architectures and design. The development of Beowulf clusters is desirable and complementary to the development at vendors.

**Computer architectures**

Dr. Walker continues by emphasising that the discrepancy between CPU instruction execution time and memory access latency (i. e., the ratio of memory access time and CPU cycle time) is increasing rapidly, and might even exceed 400 within 10 years according to some authors.[26] The discrepancy can be resolved in part by deeper memory hierarchies. This is further discussed in Reference [26]. It is clear that application programmers will have to pay attention to this increasing memory hierarchy depth.

At the same time both shared and distributed memory architectures will be used. We can expect an increase of applications using a mix of shared memory and distributed memory.

Dr. Walker strongly emphasise the rapid increase of bandwidth in communication networks. “Bandwidth will have disappeared as a problem within a few years time from today.” The doubling time in Moore’s-law-for-bandwidth is much shorter than in Moore’s-law-for-transistor-density. This implies that bandwidth limitations will effectively disappear, at least for long distance networks, but very likely also for internal networks in a few years time. Network latency, on the other hand, is an increasing problem.

The number of different architectures and vendors existing side-by-side on the market will not change much over the coming years. Monopoly is undesirable for everybody.

### Computer systems

Dr. Walker comments on the ENACTS user questionnaire that highly parallel systems are not considered very important by the users. This can most likely be explained by bad scaling properties of most applications. It is also hard to write efficient parallel codes! The result in the user questionnaire probably reflects the fact that only a small fraction of all users (the ones that have parallel codes) use most of the computer resources. A few "capability users" will always exist with applications scaling to several thousands of processors, for which only Amdahl's law sets a limit on application speed-up by parallelization.

Dr. Walker sees no definite upper limit for the size of computer systems. However, large machines will always be partitioned. It will be desirable to administrate them as a unit even if they are not actually used as one unit. It is possible and desirable to have several different systems within a single administrative cluster. There is work going on in this direction. Compaq, HP, Sgi and 3$^{rd}$ party companies work together on scheduling solutions for the Grid within the New Productivity Initiative[11], and this work does also apply to highly parallel systems. In this way does larger and more powerful systems not imply less reliable systems.

A larger fraction of peak performance will be delivered to applications in the future. This has been the evolutionary trend at Compaq the last 10 years.

A general observation concerning HPC procurements is that several recent initiatives at European institutes underestimate cost of future HPC resources. There is an important issue concerning how much one can get for money. The cost estimate for HPC systems can not be based on PC cost! This issue is a problem in Europe due to very rigid rules for requests for proposals (RFP) — the offers from vendors can not be changed after the RFP. This means that all discussions have to be done before the RFP.

### Linux and open source

Concerning the development of operating systems Dr. Walker says that Linux will increase in importance, while proprietary Unix will decrease in importance. All vendors will move to the same, open source OS environment — perhaps Linux— in a few years time. This is attractive for the vendors, which will be relieved of the burden of developing their own proprietary OS. Vendors will not be big in open source development and activities. Rather, they will be relieved to work on the core work (hardware architecture and design) instead of software development.

### Storage

Dr. Walker says he is not an expert on storage. In general, tapes are being replaced by spinning media, and it is very unlikely that storage will be a bottle-neck in the future.

One should note that the commercial market is 10 times larger than the scientific market. Commercial applications are dominated by databases from which a large number of small reads are done. In contrast, large writes dominate in scientific applications. There will be memory hierarchies, and it might be important for the scientific community to make sure the development at vendors will not only be focused on performance to do many small reads.

### Programming

Fortran, C, C++ and Java will continue as the dominating programming languages 5–10 years from now.

Concerning parallelization Dr. Walker says he is afraid that we will still have mainly MPI and OpenMP even 10 years from now. Shmem will be replaced by MPI. HPF will not be used — it is already a failure. Neither will co-array Fortran ("F−−") be dominant. UPC (Unified Parallel C) on the other hand, is an interesting alternative. It can be expected that most programming work at the parallel HPC level will be done by fewer and more professional programmers, developing codes that will be widely used. Performance will always require hardware awareness. However, new hardware does not imply new programming models.

OpenMP is not used much according to the ENACTS user questionnaire, probably due to its low scalability and portability. OpenMP can only run on shared memory computers while MPI can run on any computer.

Programming paradigms will not have changed much 10 years from now. Some tools will appear to simplify some parts of programming. For example, high level problem solving environments, like Cactus[17, 2] and other similar solutions, take the burden of disciplined programming from the scientist and user.

The pain/gain ratio of programming, i. e., programming "pain" versus the obtained performance gain, will unfortunately not improve in the future.

### Applications and benchmarking

There was an especially clever response in the "Future needs" section of the ENACTS user questionnaire: "we have defined the complexity of the problems we have studied to be those which live just at the borderline of being doable, and we will most likely continue to do so." There are several application areas that will become doable and benefit from the power of HPC systems 5–10 years from now. Some notable examples are the usage of computations in life sciences such as biochemistry and genomics. "Proteonics" and protein folding are other examples, which are also potential users of the Grid.

It can be expected that open source will dominate in applications and research software. There is also a slow trend, in industry as well as in academia, towards an increased use of $3^{rd}$ party codes. In this context it should be noted that architectures will be influenced much more by commercial applications than by scientific applications, since scientific applications constitute only approximately 10% of all applications.

It will be easier to disseminate knowledge, not only data, in the future. However, concerning security, the tendency in industry is towards less security. The request from customers for security in communication, file transfer etc. is not large enough to sustain a company (as Cray learned). Thus, security will not be strongly developed.

Visualisation and graphics development will be driven by video games. Most visualisation systems will be based on commodity devices, COTS. An important exception, however, is immersive visualisation. This development will not be driven by the games industry, since games probably will use heads up glasses instead of large rooms.

Concerning bench-marks it would be desirable to make bench-marking out of well-known, portable, well-structured, standard, professional codes. Peak performance can be misleading: all vendors will have fantastic peak performance! One should also note that vendors do not want to debug user bench-marks. However, conscious users and thoughtful analyses from users will influence vendors. For instance, prior to the ASCI[1] investment the discrepancy in performance of the memory hierarchies between what was needed by the users and what was available from the vendors was studied. ASCI then balanced the investment to compensate for large discrepancies. Within Compaq, application specialists work close together with architecture specialists.

**Grid**

We do not yet know what the Grid will become. The Grid can be several different things, and we should try them all! What people actually do will crystallise out in time. It is clear that the Grid will enable virtual communities to collaborate on projects. Meta-computing might not even be a key part of the Grid.

Dr. Walkers first reaction to the “power grid” vision about the Grid (a computer grid as easy to use as the ordinary electric power grid) was that “this is a solution looking for a problem”. This observation was also made by Ed Seidel: there is a need for problems and applications to be solved by the Grid. One purpose of the Cactus code[2] developed by Ed Seidel *et al.* is to simplify development of applications to be run on the Grid.

However, the Grid is still immature! One should remember that the first global Grid forum meeting was held in March 2001 and the first Grid forum in 1999. There are no existing examples showing its possibilities. This is why results from the ENACTS user questionnaire shows that the experience of Grid enabling technologies is very low.

The Grid will have no impact on computer architectures — the computers are nodes on the Grid.

**Future computer centres**

The number of HPC centres in the future is a political question. What kind of user service will these centres provide? Basic support? Application specific support? There will be application specific centres! Centres will have to differentiate to compete.

A trivial common denominator of several recent HPC initiatives (Los Alamos, Earth simulator, ASCI) is that they are building new huge buildings with huge computers.

PFlop performance will be reached within ten years. This leads to a large power consumption at the centres, on the order of, say, $5 M per year in cost for electricity.

A large European centre is lacking. Europe need to create a large centre, comparable to Earth simulator in Japan or the ASCI initiative in the US, in order to influence the vendors. A possible application area for such a centre could be knowledge management in a multilingual environment — this would fit well into the European situation.

## 3.2 Wolfgang Mertz (Sgi)

Sgi was represented by Dr. Wolfgang Mertz at the interview. The interview was held at NSC on June 11, 2001.

Dr. Mertz received a PhD at the University of Innsbruck. He worked two years as assistant professor at the University of Innsbruck. After that he spent a year in Brisbane, Australia to work for CSIRO (Commonwealth Scientific & Industrial Research Organization) in a mining research project. From 1989 till 1994 he was HPC Presales Analyst for Convex Computer, Germany. Since December 1994 he works for Sgi, since 1999 as an HPC Consultant in Sgi's EMEA (Europe, Middle East & Africa) organisation.

### 3.2.1 Summary of key ideas of Wolfgang Mertz

The most important key ideas conveyed at the interview with Dr. Mertz, Sgi, can be summarised as follows:

1. Moore's law for transistor density will hold for at least five more years. The development will move towards system-on-chip.

2. More and more COTS products will be used as building blocks in future HPC systems. Interconnects might be an exception to this.

3. Sgi predicts an evolutionary development of computer architectures from todays technologies over the next coming years, rather than revolutionary changes.

4. There will be a decreased number of architectures in the future, as compared to today. Parallel, shared memory architectures will be most common, and vector architectures will not be used.

5. Some general trends are that the memory hierarchy depth will increase, the bandwidth will grow for both internal and external networks, and the latency will decrease.
6. The size of HPC computer systems 5–10 years from now will be an extrapolation from today. The largest systems will probably contain less than, or even much less than 100 000 processors.
7. The importance of Linux is growing.
8. Tapes will continue to be used for long term storage. SAN will be the dominant kind of storage system technology.
9. Fortran, C, C++, perhaps Java, together with MPI and OpenMP will be the programming languages and parallelization environments for HPC computing in 5–10 years.
10. Several mixed programming models will be used, applied to both shared and distributed memory architectures.
11. The use of visualisation will become simpler in the future. Powerful desktop computers, with graphics cards, in combination with high capacity remote resources and faster networks, will bring high performance graphics to every desk-top within 3–4 years.
12. There will still be computer centres 5–10 years from now. However, the Grid development will cause more international collaboration between centres, and also emergence of application specialised centres which are accessed through the Grid.

### 3.2.2 Interview

#### Moore's law and hardware development

As far as one can see right now, Moore's law for transistor density will hold for at least five more years for processors and memory. As the transistor density increase, more and more functionality will be put on a single chip. The transistors will be used for on-chip caches, multiple CPU's, etc. The development will move towards system-on-chip.

The clock frequency of processors will, of course, also increase. However, Sgi does not consider increased clock frequency as a goal in itself. There will not be a "floating point operations per second world record" by Sgi. Instead the focus is put on achieving systems with a performance balance between CPU, memory and storage.

Resolution of displays will not change much.

#### Computer building blocks

It is clear that more and more COTS products will be used as building blocks in future HPC systems. For instance, mass-produced general purpose processors will be used rather than low volume special purpose processors. The reason for this is the cost of the components. Proprietary hardware leads to higher performance, but is also much

more expensive. It is very possible that the price of COTS is so much lower that several components can be bought to the same price as a single proprietary component. Price/performance is the bottom line, not only performance!

Sgi will build computers with building blocks based on NUMA ("Non Uniform Memory Access") and NUMA clusters. The architecture will remain proprietary, while the individual components like processors, memory, and disks will be "off-the-shelf" components. However, interconnects might be an exception where native components will be used.

**Computer architectures**

First of all, Sgi predicts an evolutionary development of computer architectures from todays technologies over the next coming years, rather than revolutionary changes. A general goal for Sgi is to avoid the imbalance of processor performance versus memory bandwidth and latency. Shared memory distribution is used, and will continue to be used, rather than distributed memory. As mentioned in the "Computer building blocks" section on the facing page, processors will be distributed in an architecture based on NUMA. Very large processor configurations will be made out of clusters of NUMA systems. In the near future, Sgi uses the proprietary "NUMAflex" system based on modular "bricks" to maintain a flexible and scalable architecture, in which the performance of I/O, CPU and storage can be tailored to fit specific applications.

Shared memory architectures will be most common in the future. Mixed shared and distributed memory will also be used to some extent. In general, there will be a decreased number of architectures for HPC. NUMA and clusters of NUMA architectures will dominate. There will also be a decreased number of microprocessor architectures.

Some specific trends in computer architectures are that the memory hierarchy depth will increase, the bandwidth will grow for both internal and external networks, and the latency will decrease.

Regarding Beowulf clusters, these are not general purpose machines. They are only suitable for certain types of applications. The same applies to vector machines. Vector architectures will not be used in the future.

**Computer systems**

The number of processors in the largest HPC computer systems 5–10 years from now will be an extrapolation from today. This means that there will probably be less, or even much less than $100\,000$ processors in the largest systems, and for single image less than, or even much less than $10\,000$ processors.

There is obviously a larger chance for component failures at some point on systems with a large amount of components. This will cause problems with single point failures in very large and powerful computer systems. Thus, there is a risk that more powerful imply less reliable systems. However, there are technical work-arounds to decrease this

risk. One example is the use of "back-up" components. Another way to get around this is to copy the work on several processors and use "voting" for the correct result: if two out of three processors agree, that is the correct result.

Regarding performance, we do not know what the sustained system performance compared to peak, and compared to todays ratio, will look like.

### Linux and open source

The importance of Linux is growing. There will be a cross-over later this year (2001) where there will be more Linux installations than proprietary Unix versions taken together.

Sgi recognise the importance of working with the Open source community. This is important since it is a way to make valuable, formerly Sgi proprietary, technologies the de facto standard within the Open source community. It is better to have Sgi's own technology used in the Open source world than having to conform to worse standards developed elsewhere. For this reason, Sgi already works together with the Open source community. For example, XFS, and other Sgi software, has been released as open source.

### Storage

Tapes will continue to be much cheaper than disks. Therefore, tapes will continue to be used for long term storage. Moreover, there is no need to keep long term storage on-line. SAN will be the dominant kind of storage system technology.

### Programming

Fortran, C, C++, and perhaps Java will be the programming languages for HPC computing in 5–10 years. However, for most engineers and scientists C, C++ are unnecessarily complicated. Fortran is complex enough for these users and application programmers, and there is no need for them to use C, C++. Therefore Fortran will continue as the basic programming language for this important group of HPC users.

Co-array Fortran, "F−−", was a "nice try". However it failed because it is no standard and is too complex. As a consequence there is no vendor who wants to develop compilers for Co-array Fortran.

MPI and OpenMP will probably be the dominating parallelization environments, even 10 years from now. Sgi believes in an evolutionary development of hardware — there will be no hardware revolution within 5–10 years. And without a revolution in hardware development emergence of new programming models is very unlikely. Several mixed programming models will be used, applied to both shared and distributed memory architectures.

It is unclear why the ENACTS user questionnaire shows that the usage of OpenMP is low. Programming paradigms should be divided in

shared and distributed *memory*, not OpenMP and MPI. High scalabilty often requires MPI.

There will be no major improvements regarding the pain/gain ratio of programming. Architectures get more complicated, but compilers will not do the work to simplify programming. Thus, the pain/gain ratio will not decrease in the future.

**Applications and benchmarking**

What areas of research and applications will benefit the most from future HPC systems? This is very difficult to predict. However, large individual investments, such as the Earth simulator in Japan, or large climate investments in Europe, could influence the chart. Influence in the other direction is also possible. At least $3^{rd}$ party codes does influence the system architectures, to some extent.

Vendors have to regard what users want. However, there is also a possibility to "teach" users to use new technologies. An example is the transition from vector to parallel architectures. Thus, the vendors should not be influenced by every "wish" from the users. There are cases where new technologies should be taught to the users.

The use of visualisation will become simpler in the future. For example, there will be more intuitive and simple sharing of data by graphical methods. Powerful desktop computers, with graphics cards, in combination with high capacity remote resources and faster networks, will bring high performance graphics to every desk-top. Within 3–4 years from now the problems with interactive graphics applications will be solved.

A general trend for applications is that the application memory size per processor will grow.

Peak performance is not a good metric for benchmarking. The use of real codes as benchmarks is better, but can also be misleading. More realistic benchmark codes are needed. User specific codes are still the best benchmark metrics, but are also very expensive for the vendors. Thus, in the future new ways and procedures might have to be established to perform those tests. SPEC and other standard benchmarks (including $3^{rd}$ party application benchmarks) are OK as well. Peak performance, however, is the worst benchmark metric for computer performance.

**Grid**

The Grid could be defined as a way for closer cooperation between centres such as application execution across sites, and grand challenge applications distributed across different sites and systems. Using "spare cycles" is another likely future use of the Grid. Moreover, there is and will be more confidence in certain applications at certain sites. This points at the possibility of application specialised centres which are accessed through the Grid. One should note that "Power grid" usage of the computer Grid already exists in a way, on a local scale, e. g., in terms of common batch systems at computer centres.

Some starting points for the Grid exists already. For example, there are many European Grid projects such as EuroGrid[8] and others. Many problems have not yet been solved, however. The Grid is at a very early stage. This is why the ENACTS user questionnaire shows that the experience of Grid enabling technologies is very low.

The Grid will have an impact on software development, but no impact on hardware architectures. The development of networks is fast, but this is not due to the Grid.

#### Future computer centres

What will a future computer centre look like? First of all, there *will* be computer centres. HPC experts will always be needed. There will also be more national and international collaboration. Regarding the development of new hardware architectures, there will be an evolution, no revolution.

From the point of view of Sgi, vector architectures have been given up. Only parallel architectures will be used in the future. Originally MPP-systems had no shared memory. Now, however, does Sgi focus only on shared memory systems since programming is made easier already by the shared memory architecture. Moreover, a common complaint about distributed memory systems, e. g., the Cray T3E, was about too little memory.

It will be increasingly important with good compilers and tools in order to obtain high floating point performance. There will be no major change in programming languages and parallel paradigms. The main reason for this being the great mass of code which is already developed in traditional languages. Fortran will persist — it is easier than C and C++ to use and understand for scientists and engineers.

## 3.3 Benoit Marchand (Sun)

Sun was represented by Mr. Benoit Marchand at the interview. The interview was held at NSC on June 18, 2001.

Benoit Marchand is the High Performance Computing Manager for the Europe, Middle East and Africa (EMEA) region of Sun Microsystems, responsible of all the business development activities of the Sun HPC Servers primarily targeted for commercial and technical marketplaces requiring top performance. Prior to joining Sun, Marchand has been employed by Hewlett-Packard as Systems Engineer, Nelma Data company as Research Engineer, Gould Computer Systems company as Systems Engineer and Silicon Graphics as the European HPC Pre-Sales Support Manager.

He was involved in many HPC and research projects including, but not limited to, the world's first commercial aircraft simulator project running on Unix, the communication protocol for World Cup 98 real-time VRML news feeds and the world's first ever HPC Web Tutorial.

### 3.3.1 Summary of key ideas of Benoit Marchand

The most important key ideas conveyed at the interview with Benoit Marchand, Sun, can be summarised as follows:

1. The increasing manufacturing cost of integrated circuits will cause a break in the development of performance as compared to Moore's law.

2. By 2010 very few general processors will exists. We will see general purpose systems built out of COTS as well as something new: application specific commercial derivative systems developed on the basis of high volume special purpose systems.

3. There will be a proliferation of hardware solutions. One size does *not* fit all applications!

4. Infiniband will cover 80% of the market of interconnects within 4 years if it makes it to the market.

5. There will be an increased number of architectures. Shared as well as distributed memory architectures will be used.

6. Network bandwidth will double every year. However, there will still be problems with latency.

7. The size limit for computer systems is set by logistics: space, housing, power, cooling, transportation, and management. The largest systems will by 2005 contain 50 000 processors.

8. The price/performance ratio will get worse for general purpose systems, and over time approach a linear curve. In contrast, special purpose processors (SPP) will in three years time have a price/performance ratio 10–100 times better than general purpose processors.

9. Linux is a general purpose operating system and will continue to grow. However, it is not suitable for large systems. Proprietary Unix such as Solaris will survive and be important for the stability of very complex systems and environments.

10. No tapes will be used for storage 10 years from now. There will be new techniques based on, e. g., molecular technology. Many different kinds of storage systems such as SAN, NAS and HSM will be used.

11. The dominating programming languages and parallelization environments 5–10 years from now will be C, C++, Fortran, Java, Co-array Fortran, Shmem, OpenMP, MPI, and also Grid based solutions.

12. Efforts to improve performance need to be focused on applications. Applications are the key to better performance, the hardware is not the limiting factor, hardware develops faster than software. A stronger emphasis on in-house developed codes is required. ISV application codes are often too general and not specifically developed for a target architecture, which reduce the performance.

13. Big application areas of the future are life sciences (genetics and biochemistry), risk management and finance, Internet real time transactional applications, and visual computational interactive steering of applications.

14. The Grid term is a fad. Grid computing, or meta computing, has been there for some time. The Grid is already there, also in terms of the possibility to distribute software on the web.

15. It will be an increasingly important task for HPC centres to enable higher productivity for key application codes by in-house development. Future computer centres will focus on software to 90%, and only manage and maintain the computers of its customers. Centres will provide researchers with application code, not only hardware resources. They will be specialised in specific application areas.

### 3.3.2 Interview

#### General remarks

Marchand points out that what is said during the interview are his personal views, and not necessarily the official views held by Sun.

#### Moore's law and hardware development

The increasing manufacturing cost which follows upon the increasing transistor density of integrated circuits will cause a break in the development of performance as compared to Moore's law. Regarding types of components, one should note that a large fraction of the investments in hardware development is put on embedded systems, such as DSP (digital signal processors). DSP's is the area where much of the money is going, and that is also where the fastest development will be.

#### Computer building blocks

Mass-produced special purpose appliance processors will be used in future HPC computer systems, rather than low volume proprietary processors. By 2010 very few general processors will exists. Why? First, because 90% of investments in research and development of processors in the chip industry are in embedded appliance processors, like DSP's. Second, Moore's law imply a 0.074 micron chip line width within 3 years from now, which implies a price greater than 2–10 times more

than today. This means $10 B to build one FAB, which in turn implies that there will not be enough volume for general purpose (GP) processors to cover investment costs in the FAB. However, embedded processors have a much larger volume which will lead to a 10–100 times smaller price/performance ratio. The influence from high volume commercial markets on HPC computer architectures and performance will therefore be very large. Within 2 years appliance computing will be used on a large scale.

There is another, new approach which also will emerge: application specialised commercial derivatives based on large volume commercial components which have been adapted to certain types of applications. COTS will not be enough since they sometimes contain too much — which adds up to a problem in cost for very large systems — and sometimes too little, such that they miss some vital feature. Sun works with hardware vendors to create these commercial derivatives. Certain classes of applications can gain tremendously if they can use consumer product type of hardware. Price performance will start to differ dramatically depending on whether the application can use application specific hardware (low price performance ratio and on an exponential Moore's law curve) or has to stay with general purpose systems (high price performance ratio and moving towards a linear curve). These systems will, however, require a different programming model based on data-flow, see section "Programming" on page 29.

There will be a proliferation of hardware solutions, some based on COTS and some based on special purpose hardware adapted to different applications requirements. One size does *not* fit all applications! In other words, we will see general purpose systems built out of COTS as well as something new: application specific commercial derivative systems developed on the basis of high volume special purpose systems. In this sense the vendor will become a system integrator.

Beowulf clusters are a foe for Sun. Don't do Beowulf! Sun have their own clustering technique.

Regarding interconnects: Infiniband will cover 80% of the market within 4 years if it makes to the market. It will be *the* interconnect for a long time.

**Computer architectures**

Referring to the discussion above, one can conclude that there will be an increased number of architectures since commercial derivatives adapted to specific applications will be used in addition to general purpose systems.

Network bandwidth will double every year. However, there will still be problems with latency. The latency can be reduced by removing excessive steps, but obviously nothing can be done to increase the speed of light.

Some measures which can be taken to decrease the imbalance between processor performance versus memory bandwidth and latency are the following:

1. Latency hiding by multithreading.

2. Place memory close to processor.

3. Use asynchronous memory.

4. Tuning applications for better locality of references.

Regarding item 4, it is probably less than 1% of all programmers who know how to tune their applications for, e. g., several cache levels.

The kind of memory and processor distribution which will be used will depend on the market, which in turn is determined by the types of applications which are used. Memory will be closer to, and tighter coupled to processors. Both shared and distributed memory architectures will be used.

**Computer systems**

Highly parallel systems are not considered very important by the users responding to the ENACTS user questionnaire. This is because the systems are too difficult to use — the tools do not exist. One should also remember that the performance gain is limited, even for highly parallel systems, at least by Amdahl's law.

How large systems can we expect to have in the future? The upper limits for very large systems is set by logistics: space, housing, power, cooling, transportation, and management. To be specific, we can expect to have systems with $10\,000$ processors by the end of 2002, 20 000 by end of 2003, and 50 000 by end of 2005. These systems will be based on commercial derivatives. This fact does also mean that more powerful does not imply less reliable systems. Appliance systems (finite state machines) are more stable.

The price/performance ratio will get worse for general purpose systems, and over time approach a linear curve ($2\times$ in performance will correspond to $2\times$ in cost) since the produced hardware volume otherwise will be too low to amortise the full cost of the investment. In contrast, special purpose processors (SPP) will in three years time have a price/performance ratio 10–100 times better than general purpose processors. From there on SPP solutions will be on a Moore's law curve for price/performance. This development has a price however: the programming model has to be completely changed.

The best performance will be achieved by changing the programming model! Furthermore, and independent of the programming model being used, it is a fact that the sustained system performance compared to peak (and compared to todays ratio) will decrease. The doubling time of Moore's law for applications is 2 years (instead of 1.5 years for hardware) which leads to a smaller slope of the Moore's law curve for application performance. These points are further discussed in the "Programming" section on the facing page.

### Linux and open source

Linux is a general purpose operating system and will continue to grow. However, it is not suitable for large systems. Proprietary Unix such as Solaris will survive and be important for the stability of very complex systems and environments.

Sun works with the Open source community by providing *all* our source code on the net under so-called "Source code software licensing", SCSL. However, there are too many open sources! The large number of open source codes is worrying because it dilutes the development.

### Storage

No tapes will be used 10 years from now. There will be new techniques based on, e. g., molecular technology. 3-D storage will greatly expand storage capacity. Many different kinds of storage systems such as SAN, NAS and HSM will be used.

### Programming

Which programming languages will be used in 5–10 years? The "traditional" languages such as C, C++, Fortran (but preferably not!) will still be much used. Java will also be used, in part because it will enable gate arrays and DSP's to be used in large systems.

Parallelization environments for the future include Co-array Fortran, Shmem, and — in particular for high performance "capacity" applications — OpenMP, MPI, and also Grid based solutions. High Performance Fortran (HPF) will not be used, it is already dead. However, there are other possible environments like JavaWulf which is based on data flow engine programming, which might increase in importance.

The usage of OpenMP is low according to the ENACTS user questionnaire. There are several reasons for this:

- Expertise to tune code is lacking.
- People are lazy. They rather wait for faster chips than learn to program well.
- Amdahl's law.
- Restricted portability.
- Lack of tools to develop OpenMP applications.

Efforts to improve performance need to be focused on applications. This is clear from the discussion in the "Computer systems" section on the preceding page, since the gap between sustained performance and peak performance is growing. There are means to improve the situation, e. g.:

- Independent software vendors (ISV) need to be involved early on in the development of new hardware.

- In the case of in-house codes, continuous software development is necessary to be able to take advantage of new hardware.

Applications are the key to better performance, the hardware is not the limiting factor. There will always be a remaining gap between the performance of the hardware and the performance of the ($3^{\text{rd}}$ party) applications. This is because applications do not, and will not, keep up with the rapid development of general purpose hardware development. Furthermore, new hardware based on SPP's requires new programming models using data flow, as mentioned in the "Computer systems" section on page 28.

### Applications and benchmarking

What areas of research and applications will benefit the most from future systems? Life sciences (genetics and biochemistry) will use 50% of the HPC resources in 5–10 years. However, this is mainly due to the enabling of genome data, not due to the computer development. The hardware is already available. (Note that research in genetics includes studies of the human genome but also genomes of, e. g., fish and vegetables used in research about new varieties, etc.)

Risk management and finance are other areas which will grow and take advantage of future HPC systems. Application areas include, e. g., simulations of the evolution of the economy. A HPC centre offering services in this area will not just host the hardware for a bank, but the expertise in system tuning will be needed. A further application area which will benefit much from future HPC systems is Internet real time transactional applications, i. e., analysis, transformation and interpretation of web content.

A further application area which will develop and benefit from future HPC systems is visual computational interactive steering during calculations. Presently too many (unnecessary) calculations with simple variations of parameters are done. We compute 10 times more than we need. However, the cost for visualisation CAVEs is still prohibiting. We haven't done anything regarding the software. There has not been enough effort too integrate computing and visualisation. The use of computers for rendering and graphics will also grow, but this will not be a large application area.

There are too many $3^{\text{rd}}$ party codes in use today, but this will have to change. In fact, critical, key applications are developed in-house in industry already today. Users must refine their needs into *commodity*, *essential*, and *critical* applications. *Commodity* applications can be provided by ISV's. For *essential* applications it is useful to work with the ISV's. Finally, *critical* applications should be owned by the users. It will be necessary to have control over codes (in-house) to adapt programs to the new, rapidly developing computer architectures and systems based on appliance hardware. From HPC centres point of view it is reasonable to have less than 50% $3^{\text{rd}}$ party codes. Otherwise they become a commodity provider, ASP (Application Service Provider).

$3^{\text{rd}}$ party codes do not influence the system architectures. It is the

other way around: new architectures influence $3^{rd}$ party code developers. They develop codes that should run on *all* architectures (they are "agnostics" to hardware architectures), which leads to worse performance than if the application would be adapted to the most suitable architecture.

The way benchmarking of new systems is done will change in the future. Customers will pay for benchmarking as a service. Sun will in the near future provide free access to hardware, for benchmarking purposes, to customers in the process of purchasing equipment. Today, the market is abusing the benchmarking services of the vendors! However, when new systems are developed the vendors take impression from the users. Sun works with customers in advance to the release of new systems.

**Grid**

The Grid term is a fad. Grid computing, or meta computing, has been there for some time. The Grid is already there, also in terms of the possibility to distribute software on the web. A few groups already do it. A likely, more advanced version of the Grid is the situation where many special purpose machines will be connected to each other. Jobs will be submitted through portal and rerouted to suitable resource. All machines have to be connected to route the application to the right machine. One could note that, in this scenario, the Grid will have an impact on computer architecture through the seamless integration of various architectures into a single environment.

Sharing, visualisation and dissemination of data will be performed through the Grid in the future. A limiting factor of the Grid is the spread of data.

The ENACTS user questionnaire shows that the experience of Grid enabling technologies is very low. There is a lot of talk, but no-one is doing anything. There is too little experience of Grid usage, although the technology has been there already a long time. However, I don't see any fundamental reasons why the Grid experience must be low.

**Future computer centres**

A typical future HPC centre can be characterised as follows:

- Centres will focus on software to 90%, i. e., how to get applications running with high performance. Centres will evolve into software optimisation, tuning and portability experts. The reason for this is that there is an increasing gap between the performance of the hardware and the performance of the ($3^{rd}$ party) applications. Applications do not keep up with the rapid development of general purpose hardware. The doubling time of Moore's law for applications is 2 years (as compared to 1.5 years for hardware) which leads to a smaller slope of the Moore's law curve for performance.
- Centres will not host large general purpose computers, rather manage and maintain the computers of its customers.

- More effort will be put into "top applications" (the most critical applications) to obtain larger fraction of peak.
- Centres will provide researchers with application code, not only hardware resources.
- Centres will have to foresee future needs in collaboration with ISV.
- There will be other types of services than today at computer centres, such that consulting for external partners. Consulting for building large scale computers will develop. This type of expertise is already wanted and needed in industry. The focus now is put on hosting of hardware and applications. In the future the focus will be put on consulting (optimisation, workload management, etc.) for external customers.
- Centres will be specialised in specific application areas and manage specialised hardware for various markets.
- Centres will hold a large capability for applications tuning for its specific area of focus. And they should have a large capability of tuning for in-house applications.

## 3.4 Jamshed Mirza (IBM)

IBM was represented by Dr. Jamshed Mirza at the interview. The interview was held at NSC on June 19, 2001. The interview had the format of a view-graph presentation by Dr. Mirza, with open discussions and questions answered during the presentation.

Dr. Mirza is a Distinguished Engineer in IBM's Server Group. His current responsibilities include systems architecture and technical strategy for Scalable Systems and Clusters. Most recently he has been working on IBM's Linux cluster solutions. Prior to this, he was the lead systems architect for the RS/6000 SP, which is a Unix-based scalable system widely used for Supercomputing and for high-performance Enterprise applications.

### 3.4.1 Summary of key ideas of Jamshed Mirza

The most important key ideas conveyed at the interview with Dr. Jamshed Mirza, IBM, can be summarised as follows:

1. The transistor density will continue to grow according to Moore's law during the foreseeable future.
2. The extra space enabled by the Moore's law development of transistor density might be used to obtain customised chips, i. e., chips for different purposes which are based on a common core.
3. The decline in manufacturing cost per transistor may slow for several reasons. However, system performance will likely continue to improve at an exponential rate through improvements in circuit and system design.

4. IBM will leverage standard technology wherever possible, and judiciously add custom technology only when this is required.

5. With current semiconductor technology trends, DRAM latency as measured in number of clock cycles will likely continue to grow, and the bus speed and bandwidth will not keep pace with increases in microprocessor speed. Several design issues can be utilised to mitigate this problem, e. g., deeper memory hierarchies.

6. Clusters of SMP's will be the dominant future architecture for HPC.

7. Infiniband could become the standard for a common network, if it takes off. Optical interconnects could provide a solution for higher bandwidth and lower latency access to memory within the SMP. Aggregate memory bandwidths an order of magnitude greater than what is available today in the largest SMP's may be possible. But such technology within an SMP is probably several years away.

8. We can expect systems with more than $10\,000$ processors 5–10 years from now. The upper limit in terms of number of processors is set by the price, not by the technology.

9. Linux is strategic and very important for IBM. IBM does not plan to develop their own variant of the Linux kernel. Instead they plan to support standard Linux distributions. IBM is working with the Linux community to improve the enterprise-readiness of Linux. Where relevant, they will infuse IBM technology into the open source Linux kernel through the work at IBM's Linux Technology Center. Proprietary Unix versions will likely "fade away" if open source Linux becomes enterprise-capable over time.

10. Future storage systems will increasingly be networked rather than directly attached. NAS and iSCSI (SCSI over IP) will be used.

11. Disk storage density continues to improve, but in the HPC environment it will often be the bandwidth rather than the capacity which will determine the number of disks required in a particular system.

12. Fortran, C, C++, MPI and OpenMP will (continue) to be the dominant programming languages and parallelization environments 5–10 years from now. Evolutionary enhancements in these languages will continue to react to evolutionary changes in system and processor architectures.

13. The pain/gain ratio of programming will get worse in the future. Lack of adequate tools for software development will likely become a key issue for HPC.

14. The life sciences will benefit a lot from future systems. Visualisation will become more and more important as the amount of data increase. More and more $3^{\text{rd}}$ party codes will be used.

15. A Grid is a virtual computing system formed by aggregating the diverse services provided by distributed resources. The Grid will emerge much as e-Sourcing and e-Utilities will.

16. Computer centres will move toward a utility-like model, i. e., deliver HPC services to their customers on a pay-as-you-use basis. This model will evolve in synergy with the development of Grid computing. It is also likely that HPC centres will become fewer and larger.

### 3.4.2 Interview

#### General remarks

The information contained in the view-graph presentation by IBM has not been submitted to any formal IBM review and is distributed as is.

The presentation was based on a combination of technology and market trends as observed by experts (both external and within IBM) which were used to get a view of how information technology might evolve and how it might be exploited in the next five to ten years. It is not to be construed as IBM's future direction in any way. Rather it is merely an assessment of the general trends in information technology, based on available information at this time. Finally a caution: in a dynamic industry such as this, trends can change in unanticipated and dramatic manner when "disruptive technologies" emerge suddenly.

#### Moore's law and hardware development

Chip performance is expected to continue along the Moore's law trend line for the next 5 to 10 years. This will be enabled by new materials, device structures, continued scaling, and system design. However, significant technical challenges remain. We are approaching atomic dimensions where statistical variations begin to matter. Furthermore, reducing the on-chip wire delays requires dielectric materials improvement.

There are nanometer scale device alternatives to conventional semiconductor techniques. Numerous approaches are under investigation, but there is still no credible competitor to silicon for logic circuits.

The historical decline in manufacturing cost per transistor may slow down due to lithography lag and other reasons. However, system performance will likely continue to improve at an exponential rate through improvements in circuit and system design.

The transistor density will continue to grow according to Moore's law during the foreseeable future. The processor core on the latest IBM processor chip today (the IBM Power 4) occupies approximately $\frac{1}{4}$ of the transistors (and area) of the processor using todays CMOS technology. This implies that in two processor generations from now, i. e., in about three years, the core will only occupy $\frac{1}{16}$ of all available transistors. Development economics and high demands on time-to-market will drive the use of common processor cores across many applications using a

common base instruction set and a common micro-architecture. It is therefore likely that the extra space enabled by the Moore's law development of transistor density will be used to obtain customised chips, i. e., chips for different purposes which are based on a common core. For example, a low end chip could be obtained by adding fast I/O and L2 cache using the extra space for transistors; a game chip by adding high bandwidth memory interface, high bandwidth graphics interface and special arithmetic processors; a network processor chip by adding external cache controller, packet processor, protocol accelerator, and Ethernet interface; or a dense server chip by placing several cores plus external cache controller and memory control on a single chip.

Regarding basic technology, CMOS will likely continue as the technology used for processors. Standard Silicon CMOS technology promises continued performance improvement at least over the next several years. Clock speeds up to 4.5 GHz are possible, without "supercooling." Over the next decade however, we can expect that traditional CMOS scaling will end unless some significant innovation in process technology appears.

As clock speeds and chip sizes keep increasing, number of clock cycles required to traverse across the chip is also increasing. Future processor chip designs may use multiple asynchronized clocks to control different parts of the chip to improve performance. Similarly, researchers are beginning to look at the feasibility of using optical technology to interconnect components of an SMP on a board. However this is unlike to happen anytime soon.

To conclude, the performance trends of processors and architectures in approximately 10 years can be summarised as in Table 3.1.

Table 3.1: Summary of performance in approximately 2011 (10 years from today) of processors and building blocks.

| **Unit** | **Performance** | **Bandwidth (B/F)**[a] | **Latency** |
|---|---|---|---|
| Chips | 10s of Gflops | $\sim 4$ | 1–2 cycles to on chip cache. |
| UMA building block | 100s of Gflops | $< 1$ | 10s of cycles to off chip cache. 100s of cycles to local memory. |
| NUMA node | multi-Tflops | $< 0.25$ | A few 1000 cycles within NUMA node. |
| System | multi-Pflops | $< 0.25$ | Several 1000s of cycles (message passing). |

[a] $\text{B/F} = \frac{\text{Memory bandwidth (GB/s)}}{\text{Compute perf. (Gflops)}}$

Finally one should note that an alternative trend aimed at improving performance/Watt and total cost of operation will be to use less powerful processors that use much less power, and constrained memory and I/O. Such a trend will mean that many more processors may be

required to reach a desired aggregate performance, further increasing the burden on programmers and users of HPC systems.

**Computer building blocks**

Market dynamics will not support substantial unique development for ultra large systems. A guiding principle for building HPC systems at IBM is to leverage standard technology wherever possible, and to judiciously add custom technology only when this is required. Further, any new technology must ultimately be applicable to mainstream commercial applications and systems for it to be viable from a business perspective.

The vendor is becoming more and more of a system integrator. In this context it should be mentioned that Beowulf is not a foe for the vendor, certainly not. The Beowulf development is more of a friend, since it is getting more and more people into the cluster environment. The "build yourself" model will not survive however, especially outside of universities. The commercial market needs and requires support. It could be mentioned that services is a major part of IBM revenue source.

**Computer architectures**

Microprocessor performance will continue to increase much faster than memory performance. This implies that with current semiconductor technology trends, DRAM latency measured in number of processor clock cycles will likely continue to grow, and that the bus speed and bandwidth will not keep pace with increases in microprocessor speed. Furthermore, the circuit density will continue to increase faster than circuit speed. Server designs must evolve to reflect these trends, in order to decrease the imbalance of processor performance versus memory bandwidth and latency. To this end, there are a number of design issues that will be utilised, such as:

- Increased concurrency.
- Modest levels of hardware multithreading.
- Hardware accelerators and offload engines.
- Deeper hierarchies/multiple levels of interconnect.
- Large bulk cache.
- Hardware intelligent prefetch.
- Point-to-point buses and switches.
- Distributed directories.
- Increased integration on the chip.
- Fault detection, isolation, and recovery.
- Optical interconnect within the SMP.

Clusters of SMP's ("symmetric multiprocessor") will be the dominant future architecture for HPC. In the specific case of IBM such SMP's will over the next several years be based on the recently released Power4 family of processors. There will be 100s of large SMP's or 1000s of smaller, totalling 10 000s of processors in the largest systems. Scalable shared memory across subsets of nodes can alleviate the need for explicit management of data (as in message passing) and storage allocation by programmer. This is unlikely to be possible across large systems though. Thus, from a programming point of view, MPI and OpenMP will co-exist over the next several years. This point is further discussed in the "Programming" section on page 39.

Regarding the development of networks, Infiniband has the potential of becoming the standard for a common network, if it takes off. Proprietary networks will still be around but will probably be replaced by commercial alternatives. Optical interconnects are becoming more interesting and could be "in the system box" within 5–10 years. This might enable new system architectures for wide bandwidth and low latency interconnection. Aggregate bandwidths greater than 1 TB/s is conceivable within an SMP with this technology. Low cost WDM ("Wavelength Division Multiplexing") electro-optical switches will likely move into enterprise and local area networks.

**Computer systems**

We can expect systems with more than 10 000 processors 5–10 years from now. The upper limit in terms of number of processors is set by the price. Only few organisations will be able to afford the largest systems that can be built technically.

There is also an issue regarding the reliability of complex systems. Systems are getting more powerful, but also more complex in terms of number of processors and components, the complexity of networks, the increasing number of users, devices, and number of transactions, the appearance of new types of workloads and data. This development does obviously imply more possibilities of system failure. It also leads to a general trend where support takes an increasing fraction of the total system cost. This problem needs to be addressed, and IBM has started the project eLiza[6] to do that. The project includes studies of AI based management of systems based on self-awareness, self-management, and self-repair.

Overall system level performance will follow Moore's law for the next 5–10 years, but of the roughly 60% yearly improvement in HPC system performance, only about 20% is due to improvement of CMOS chip performance. The rest comes from many other sources related to the system, such as:

- Application and middle-ware tuning
- Operating systems: tuning/scalability
- Compiler improvements

- Increased SMP and cluster sizes
- Improved board design
- Improvement of memory subsystem
- Improved power, packaging, and cooling design
- Design tools and designer productivity
- Architecture, microarchitecture, logic and circuit design improvements
- New device structures and new process technologies

The decrease of the price versus performance ratio for servers might slow down. The reason is that the price pressure is mainly put on low end hardware components.

#### Linux and open source

Linux is strategic and very important for IBM. IBM works with the Open source community through its Linux Technology Center (LTC). The LTC has a large number of programmers who work on improving open source Linux and making it enterprise-capable. The centre has people in multiple sites across IBM who work with the Linux community on several different Linux projects. IBM has no plans to develop their own variant of the Linux kernel. IBM's approach is rather to stay mainstream and through the work at the Linux Technology Center, infuse technology into the open source Linux kernel. Their goal is to make Linux enterprise capable, keep Linux available across all IBM platforms, to deliver robust solutions based on Linux and IBM solutions, and to encourage the adoption of Linux. Linux is strategic and very important for IBM.

Proprietary Unix will, however, still stay around, most likely for several more years. The rate at which Linux improves will determine for how long. Perhaps proprietary Unix will "fade away" as Linux grows.

#### Storage

Future storage systems will increasingly be networked rather than directly attached. NAS and iSCSI (SCSI over IP) will be used. The use of SAN is also expected to increase.

Disk storage density continues to improve, but it will often be the bandwidth rather than the capacity which will determine the number of disks required in a particular system, especially in HPC. One should also note that the time that data is kept "live" on disks before archiving is increasing. The trend is that storage takes an increasing fraction of the total server cost, at least for many commercial applications.

Sustained I/O rates and bandwidths of disk drives are not keeping up with the increase in storage capacity. For example, assume we are using disks with 100 GB storage capacity and a bandwidth of

10 MB/s, and that our requirements are 5 TB of storage capacity and a bandwidth of 5 GB/s. This would require 50 disks to meet the need for storage capacity, and 500 disks to reach the required bandwidth.

**Programming**

One can expect an evolutionary development of computer hardware and systems, rather than revolutionary changes. It is therefore unlikely that new hardware will lead to new programming models.

Fortran, C, C++ will be used as programming languages during the foreseeable future. In addition, annotated languages will be used, allowing explicit low level parallel coding mapped to underlying system characteristics.

Co-Array-Fortran/UPC, Shmem, HPF may be interesting for smaller user communities. However, general support of these languages by major vendors is unlikely. So the question is: will an active Open source development community build around any of these language extensions to sustain them?

Instead, standardised, fully general programming models and frameworks will be used, which exposes diverse types of parallelism and provides mechanisms for managing locality, latency and coherence.

MPI and OpenMP will (continue to) coexist over the next several years. This will enable programming the hybrid architecture of clusters of SMP's, also mentioned in section "Computer architectures" on page 36:

- Flat models will be used for ease of use and portability.
- Hybrid models (MPI + OpenMP) will be used for performance.

However, portability is an issue. OpenMP is not supported on all parallel platforms, and does not extend across a cluster. These are some possible reasons why the ENACTS user questionnaire shows that the usage of OpenMP is low.

The pain/gain ratio of programming will not improve in the future. On the contrary, it is expected to worsen. Memory hierarchies will continue to grow deeper. Ability to handle many levels of parallelism and locality of data will therefore be critical. Ultimately the programmer will be responsible for that. It is therefore likely that lack of adequate tools for software development will become a key issue. Furthermore, the best price versus (peak) performance will be found in systems using less powerful processors in smaller SMP configurations with constrained memory and I/O capabilities. This will further aggravate the challenge to sustained application level performance. The task of the programmer can be greatly facilitated, however, by supporting all dominant programming models across all architectures.

Debugging and performance analysis in parallel environments is, and will continue to be, difficult. This will only become worse with increasing levels of parallelism. Key research is needed in presentation techniques and in tight two-way information flow between compilers and tools. System vendors are not motivated to invest in good tools for

supercomputing applications. Open source is a promising solution in this area.

### Applications and benchmarking

Life Sciences is a major emerging area in HPC. Availability of inexpensive but "unlimited" computing capability promises to unleash major advances in the area. Life Sciences research using computers is expected to greatly reduce the cost of developing new drugs.

Visualisation will become more and more important as the amount of data being generated by computer simulation explodes in volume. Sharing, visualisation and dissemination of data will likely be provided through the Grid in the future.

Corporations increasingly use 3$^{rd}$ party applications rather than develop their own proprietary codes, because of the cost of developing and maintaining the complex applications for increasingly complex architectures. This trend will likely continue. Proprietary codes will be developed only when they will provide significant competitive advantage.

Benchmarking is becoming increasingly expensive and vendors are unlikely to do custom benchmarks except for very large systems. What is required is a standard benchmark that is widely accepted as representative key HPC areas. Such a benchmark does not currently exist.

### Grid

A Grid is a virtual computing system formed by aggregating the diverse services provided by distributed resources, to synthesise problem solving environments. The vision is to create virtual dynamic organisations (research or business) through secure and coordinated resource sharing among collections of individuals, institutions, and enterprises. Communities of users, generally with like interests, will pool resources such as compute, storage, and data. Applications will run at, or data will reside at the "best" available resource, with ubiquitous access.

Grid computing will emerge from both technical and commercial computing. Opposing forces will force it. However, there are also several unique problems related to the Grid caused by multiple autonomous domains, wide area dispersal, dynamic and unpredictable system, huge scale, and finally massive heterogeneity in architecture, software, policy, support, etc. Grids are just beginning to be used in some research environments, but many problems remain to be solved adequately for wider use, such as:

- Distributed control.
- High latency and low reliability due to large distances.
- Security and authentication.
- Wide-area resource management, resource discovery, remote resource reservation, optimal scheduling algorithms.

- Accounting and charge-back.
- Data Management, data/file system access across geographically dispersed locations.
- QOS ("Quality of Service") guarantees.

IBM believes that Grids will emerge much as e-Sourcing and e-Utilities will. IBM has begun a program at IBM Research for advanced research on systems management, resource management, data management, security, and heterogeneity related to Grids. "BlueGrid" is an internal IBM grid linking resources at multiple IBM sites across three continents. It is being used for evaluation of current Grid technologies, research into Grid-related issues, and in evaluating and grid-enabling applications. IBM has recently identified Grid Computing as a key corporate initiative and is currently evaluating options for product offerings, services, and support in the area.

IBM has recently won several Grid-related bids - including the Distributed TeraScale facility (DTF) in the US, the DAS-2 Grid in the Netherlands, and DataStore component of the UN National Grid at Oxford University. IBM is also sponsoring University research in Grid-related technology and applications. An example is the Digital Mammography Archive being built by the University of Pennsylvania which would provide secure remote access to mammography data to doctors across a Grid.

**Future computer centres**

A major trend in Information Technology is the trend toward out-sourcing. Increasingly corporations are deciding to focus on their core business and electing to out-source the delivery of IT services to hosting companies. e-Sourcing is the delivery of computing services and applications online. Customers will increasingly be drawn to this utility-like model because it would eliminate the headache of hiring a technology staff and of acquiring and managing computing equipment, and alleviate the costs and time it takes to get IT projects up and running.

Large services companies like IBM will likely form partnerships with telecommunications and hosting companies to create the Internet network to deliver services as a utility company. This model is synergistic with the trend toward Grid Computing in the HPC environment and the two will likely build on each other to bring the utility model of computing to fruition.

IBM recently announced that it is planning to spend $4 billion to add 50 hosting centres worldwide that will serve as its e-Sourcing hubs. Researchers at IBM and in the Grid community are working toward building the infrastructure and computing environments for this model. If successful, it is very likely that most corporations will get their computing capabilities from a relatively small number of mega computing centres that are managed by large services companies. Customers will get access these facilities in a secure manner on a pay-as-you-use basis - especially for expensive supercomputing capability.

## 3.5 Burton Smith (Cray)

Cray Inc. was represented by Dr. Burton J. Smith at the interview. The interview was held at NSC, June 20, 2001.

Burton Smith is Chief Scientist of Cray Inc. He received the BSEE from the University of New Mexico in 1967 and the Sc. D. from MIT in 1972. From 1985 to 1988 he was Fellow at the Super-computing Research Center of the Institute for Defence Analyses in Maryland. Before that, he was Vice President, Research and Development at Denelcor, Inc. and was chief architect of the HEP computer system. Dr. Smith is a Fellow of both the ACM and the IEEE, and winner of the IEEE-ACM Eckert-Mauchly award in 1991. His main interest is general purpose parallel computer architecture.

### 3.5.1 Summary of key ideas of Burton Smith

1. One can not foresee an end to any of Moore's laws within 10 years. The increasing transistor density will lead to "system-on-chip." However, Cray will not use chips with more than one processor.
2. Both mass-produced general purpose and low volume special purpose processors will be used in future HPC systems. Memory chips will only be off-the-shelf components.
3. Cray will continue to focus on high bandwidth, and will therefore continue to produce latency tolerant processors such as vector processors. These can not be bought as COTS.
4. Beowulf clusters are low-bandwidth systems, and will exist besides high-bandwidth architectures, such as those produced by Cray.
5. The often mentioned imbalance of processor performance versus memory bandwidth and latency is a solved problem for Cray. However, it can not be solved with COTS.
6. Cray will use optical interconnect within 5–6 years.
7. Cray do not use deep memory hierarchies since caches do not scale up well.
8. There will be more improvements in bandwidth for internal networks than for external.
9. Memory access technology used in future HPC systems will be uniform and non-uniform shared memory.
10. The number of architectures will not diminish in the next ten years, perhaps even increase a little bit. The main difference between different architectures will be in their bandwidths.
11. The size of future computer systems is difficult to tell. The largest systems within 10 years might contain 10 000 processors, perhaps even more.

12. The gap between peak performance and sustained performance will continue to exist in systems with too little bandwidth. Since bandwidth is more expensive than Flops, the best performance, for a given amount of money, will be obtained by buying as little bandwidth as possible for the relevant application.

13. Cray will move in the Linux direction in 10 years. However, HPC vendors will have to solve scalability problems of Linux. This will lead to Linux with proprietary parts or modifications.

14. Open source might even relieve hardware vendors from software development. However, a potential problem is who will develop software tools for the less common high bandwidth Cray systems.

15. Cray will produce their own SAN-like storage system. Hierarchical Storage Management (HSM) based on optical networks will be used. Tapes and disks will remain for long and short term storage, respectively.

16. Fortran, Co-array Fortran, and UPC will be used in 5–10 years for programming of HPC systems. Java will also get better for HPC. Parallelization environments will be dominated by something like OpenMP, Co-array Fortran and Shmem will also be used, while MPI will diminish in importance.

17. Computational biology will benefit from future HPC systems. Quantum chemistry will change focus more towards dynamic properties. Engineering simulations, multi-disciplinary applications, movie making and rendering might be other growing HPC application areas.

18. There will absolutely be a change in the way benchmarking of new systems will be done. New benchmarks which measure performance including bandwidth are needed, e. g., GUPS (Giga Updates Per Second) suggested by NSA.

19. Future computer centres will be based on a diversity of resources to match varying demands from different applications. Future centres will also host specialised visualisation resources. Access to the centre is provided via the net through a network attached server farm.

### 3.5.2 Interview

**Moore's law and hardware development**

One can not foresee an end to any of Moore's laws within 10 years. As a consequence, for the particular case of increasing transistor density, this will lead to "system-on-chip" where caches, routers, I/O, etc. are put on a single chip. However, Cray will not use chips with more than one processor.

The processor technology used will be CMOS. Processors will run fine-grained threads, particularly to achieve latency tolerance. Regarding the performance of processors we can expect clock rates greater than 2 GHz within 5–10 years. CMOS, plus some optics, will be the technologies used for memories.

### Computer building blocks

In general, both mass-produced general purpose and low volume special purpose processors will be used as building blocks in future HPC systems. Memory chips will only be off-the-shelf components, however.

Cray will continue to focus on high bandwidth. There will always be applications that needs it. Thus, Cray will continue to produce latency tolerant processors such as vector processors. These can not be bought as COTS. For this reason Cray will continue to design their own processors. The processors are built by others, however. For example, MTA processors are built by TSMC, and SV2 processors are built by IBM. Cray can afford the development of their own proprietary processors in spite of comparatively small series. The reason is that the cost for chip design is not increasing exponentially as the cost for FAB's are. Rather, it is determined by the cost for man-power. General purpose chips have to try to solve everyone's problems, but it can be done cheaper by specialising for a niche product. The main concern in the design of ordinary microprocessors is high single CPU performance for all types of applications. This is not the issue for Cray parallel architectures, which means that design efforts can be concentrated on issues which are important for HPC.

One influence on HPC computer architectures from high volume markets such as PC's, database engines, Web servers, e-business, streaming media etc. might be in rack storage. For example, we get more dense clusters from WWW-server needs. However, in general the commercial market provides no advantage for Cray. Of course, high bandwidth to every home would be good, since this would imply high bandwidth, low cost networks, which also could be utilised by Cray. A bandwidth of 1 Gb/s to each home would require commercially available hubs capable of 100-500 Gb/s which would be in the range needed by Cray.

It is possible that the vendor will become more of a system integrator. Cray clusters are already a bit like that. The area of Beowulf clusters is a friend of Cray. The war is over. Beowulf clusters are complementary to Cray products. Beowulf clusters are low-bandwidth systems, and will exist besides high-bandwidth architectures, such as those produced by Cray.

### Computer architectures

The imbalance of processor performance versus memory bandwidth and latency is often mentioned as an increasing problem. For Cray this is a solved problem. However, it can not be solved with COTS.

"We will continue to build latency tolerant processors to get high memory bandwidth" says Dr. Smith. Latency tolerant architectures are necessary in order to achieve high bandwidth. Cray will use optical interconnect within 5–6 years from now in order to reduce the cost of interconnect and simultaneously enable large scale, high bandwidth systems. Latency is not a major concern since the architecture is latency tolerant. However, cache improvements can only help to a small part. Cray do not use deep memory hierarchies since caches do not scale up well. Cray MTA (Multi Thread Architecture) has no cache at all, and cache is not used for vector operations on the Cray SV2.

Network latency as measured in clock cycles will increase. However, the bandwidth can be improved by better wire topology. Out-of-board wires are very costly. Therefore, we need to pack everything to shorten the wires.

There is a difference in the future development of internal versus external networks. Long range interconnects capacity is approximately 1 TB/s today, but this is not affordable on-chip. There will be more improvements in bandwidth for internal networks than for external!

Memory access technology used in future HPC systems will be uniform and non-uniform shared memory. When the bandwidth is high you need shared memory as a "delivery vehicle." Dr. Smith is a little suspicious about ccNUMA architectures for HPC — "they do not scale well."

There is an amazing diversity of machines and architectures today. There will never again be one dominating architecture. The number of architectures will not diminish in the next ten years, but probably remain the same. Perhaps even increase a little bit with more and more FPGA (Field Programmable Gate Array) architectures.

Several different types of architectures will continue to exist side-by-side. The main difference will be in their bandwidths, especially bandwidth between nodes. (For shared memory systems this is the only bandwidth.) Cray will continue to develop high bandwidth, greater than $1 M, systems for the next decade.

The Cray MTA has a shared memory, which implies that memory bandwidth is identical to network bandwidth. The same is basically true for the SV2. Cray will continue to build machines like the MTA, but also COTS cluster systems on the high bandwidth side, not yet announced. There is much software development in this area, which will help the other system areas as well.

**Computer systems**

How large systems can we expect in terms of number of processors? This is difficult to tell. Moore's law tells us about one-processor systems. Perhaps 10 000 processors in the largest systems within 10 years, perhaps even more. The upper limits for very large systems are set by different factors and depending on the bandwidth. For low bandwidth systems it is Moore's law which sets the limit. For high bandwidth systems it is the density of components and the cost for interconnects. Cooling is not a limiting factor.

Larger and more powerful systems might imply less reliable systems, in particular if the amount of money to spend is constant. However, there is nothing intrinsic which prevents us from solving this reliability issue of very large systems.

The gap between peak performance and sustained performance exists, and will continue to exist, in systems with too little bandwidth. The reason for this gap is that we do not value and pay enough for high bandwidth. Of course, the needed sustained performance varies with the application. For very high bandwidth applications peak performance is less and relevant as a measure of the actual, sustained performance. The best performance, for a given amount of money, will be obtained by buying as little bandwidth as possible for the relevant application. One must be aware not to provide too much bandwidth if only Flops are requested. The reason is that bandwidth is more expensive than Flops. The most cost effective solution at a big centre will be to have several systems to match different application needs.

**Linux and open source**

It is easier to work with the Open source community if you are a hardware vendor than if you are a software vendor. Cray will do some things together with the Open source community, and cooperates a bit already. Open source might even relieve hardware vendors from software development. There are also problems related to open source, however. For example, it is hard to see who will develop software tools such as debuggers, performance analysers etc. for the less common high bandwidth Cray systems. The programming environment (compilers, debuggers and performance tools) should be one single piece.

Cray will move in the Linux direction in 10 years. However, it is likely that only the same kernel of Linux will be used for all systems. HPC vendors will have to solve scalability problems of Linux. This will lead to Linux with proprietary parts or modifications.

**Storage**

In general SAN is a successful storage system, but Cray do not see any commercial product that will, or can solve Cray customer requirements. Cray will therefore produce their own SAN-like storage system. Hierarchical Storage Management (HSM) based on optical networks will be used. In fact, it is done already now at Cray in different ways. However, we can not scale SAN or NAS to the bandwidth we need.

Tapes will remain for storage of large data sets due to its lower cost. It is possible that storage ware-houses will become more common, not at every centre however. “I had high hopes for holographic storage” says Dr. Smith. “It will not happen within 10 years, however.” The performance of long term storage will follow the performance of tape storage.

Disks will continue to be used for short term storage. High performance disk storage is needed mainly for visualisation, which will be a driver for the development of disk storage.

### Programming

Fortran, Co-array Fortran, and UPC will be used in 5–10 years for programming of HPC systems. Java will get better for HPC. ZPL will not be used much, unfortunately.

Regarding parallelization environments something like OpenMP will be dominating 10 years from now. Co-array Fortran and Shared memory programming model (SHMEM) will also be used. MPI will diminish in importance, and something else will take over. It is very difficult to tell, since there are many alternatives. The programming models will remain the same, however.

OpenMP is incompatible with MPI. It is very hard to use OpenMP on-node, in combination with MPI between nodes. This might be some of the reasons why the usage of OpenMP is low according to the ENACTS user questionnaire.

The programming pain/gain ratio will get worse in the future, as compared to today, if performance is measured in Flops (peak performance).

### Applications and benchmarking

Several new research and technology areas will benefit from future HPC systems. Computational biology is one example. This area needs HPC to get off the ground. Quantum Chemistry will change, and have more connections to biology, and more combinations with molecular dynamics. Focus will be shifted from static properties like molecular structure, to dynamic properties like drug action, protein folding, etc. Perhaps engineering will grow, e. g., using simulation of products before production. There is also a potential for multi-disciplinary applications. Movie-making and rendering are further possible application areas which will benefit from future HPC systems.

$3^{rd}$ party code vendors act on a very competitive market, and some will go away. However, Dr. Smith does not see any trend regarding how much $3^{rd}$ party codes are used, as compared to codes developed in-house.

Visualisation storage requirements are a major driver for disk performance development. Density is going up according to a “square-root phenomena,” and the bandwidth seems to be keeping up. Data will be stored where the heavy computations for visualisation are done.

There will absolutely be a change in the way benchmarking of new systems will be done. People who procure computers do benchmarking well. However, we need benchmarks which measure computer performance including bandwidth, i. e., benchmarks which measure the actual performance of the system. One problem is that there is no good way to measure performance in a system independent way. One measure, GUPS (Giga Updates Per Second), has been suggested by NSA (National Security Agency). Linpack will continue to be used as a (poor) benchmark, since manufacturers have not educated the users enough. It is also a political issue: politicians want to see that the money they give out is spent on “fast” computers!

#### Grid

Grid means a seamless access to computer centres like NSC. The technologies are available, e. g., "credit card" access. This would enable occasional needs for computer resources. However, we can not easily define the Grid. It will be many things.

Grid will not influence computer architecture, but rather system architecture. It will generate more network activity and more storage activity. It will probably push storage.

#### Future computer centres

Visualisation will be an important role for computer centres (rather than vendors). It is cheaper to send pixels than uninterpreted data. The actual graphics will of course be done locally. The key is that the time scale to generate graphics (minutes) is much shorter than the time to generate data (days). Visualisation should be done close to the data due to bandwidth limitations. The highest bandwidth is required by reading of data for visualisation.

A schematic sketch of a future HPC centre is displayed in Figure 3.1. There is a diversity of resources with varying internal bandwidths, in order to satisfy varying needs from different kinds of applications. Access to the centre is provided by the net through a network attached server farm which handles access, authentication, batch queues etc. The server farm is new as compared to todays HPC centres, and also the specialised visualisation resource.

## 3.6 Tadashi Watanabe (NEC)

NEC was represented by Mr. Tadashi Watanabe at the interview. The interview was held on June 21, 2001, in Heidelberg, at the International Supercomputer Conference 2001. The interview had the format of an interactive view-graph presentation given by Mr. Watanabe, with open discussions and questions answered during the presentation.

Mr. Tadashi Watanabe joined NEC in 1968 where he worked in the area of architectural design of mainframe computers and supercomputers. He was the chief architect for NEC's first supercomputer, the SX-2 announced in 1983, and is recognised for his significant contributions to the architectural design of supercomputers having multiple, parallel vector pipelines and programmable vector caches. Mr. Watanabe became Assistant General Manager of the Computer Engineering Division and the EDP Product Planning Division in 1988 and 1989, respectively. Since 1990 he is the General Manager of the Supercomputers Marketing Promotion Division.

Mr. Watanabe is a member of the Information Processing Society of Japan, the Electronics Information and Communication Society of Japan, and the IEEE. In 1998 he received the prestigious IEEE-ACM Eckert-Mauchly Award for Outstanding Contributions to Computer Architecture.

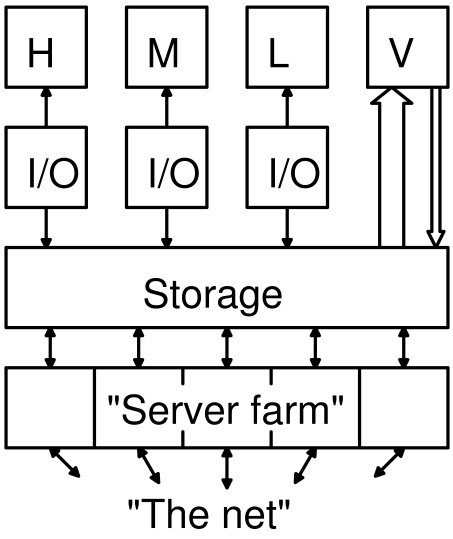


Figure 3.1: Schematic drawing of a future HPC centre according to Dr. Burton J. Smith, Cray Inc. "H", "M", and "L" indicate High, Medium, Low bandwidth computer systems, respectively. "V" indicates visualisation system. "I/O" indicates specialised I/O processor. The width of the arrows indicate bandwidth of the network connection.

### 3.6.1 Summary of key ideas of Tadashi Watanabe

1. One can expect Moore's law to hold for clock frequencies, memory bit density and microprocessor transistor density at least until 2011.

2. At NEC, the increasing transistor space will be used to put more cache space and more processors on the same chip. It is better to do more with a simple, known design, than to develop new, complicated chip designs.

3. CMOS will be the technology used for processors.

4. The price/performance trend for HPC systems will probably saturate, since most research and development of processors and other components goes into the commercial market.

5. Both COTS and proprietary building blocks will be used in future HPC systems. General purpose processors will be used for capacity computing, while special purpose vector processors will be used for capability computing. Interconnects will only be proprietary.

6. Different architectures (different kinds of memory and processor distributions) are most suitable for different types of applications. Capability computing fits best on SMP with powerful vector processors.

7. In general, there will be many computer architectures for HPC in the future. Clusters of SMP's will dominate.
8. The size of future HPC systems is limited only by the budget.
9. The price/performance factor for future systems will probably saturate. This is a result of that more money is put into development of consumer market area products, which leads to more expensive components for high-end solutions.
10. The ratio of sustained performance versus peak performance of HPC systems will decrease.
11. Linux is OK for many purposes, but NEC Unix can not be left as open source. It is too specific and not of common enough use. Proprietary Unix, or possibly special versions of Linux, will therefore continue to be used for special hardware, like NEC vector computers.
12. The need for storage is increasing, and several different storage technologies will be used (DAS, SAN, NAS, and IP-SAN). Tapes will continue to be used for long term storage. Tape technology will improve in the near future.
13. Fortran, C, C++ will be the programming languages used in 5–10 years from now. MPI, HPF and OpenMP will be the programming models used for parallelization.
14. Some application areas which will benefit from future HPC systems are genetics and climate research.
15. Benchmarks will be done as they are today. Application based benchmarks have never got off from the ground. However, the *specific* applications must also be benchmarked.
16. One scenario of the Grid is resource sharing. However, the tools to use it are not available. The Grid is too immature. The day-to-day priorities of the users are more important than to try out new things.
17. Many research disciplines will be supported within each centre in the future. One single machine never fits all!

### 3.6.2 Interview

**Moore's law and hardware development**

According to ITRS'99 (International Technology Road-map for Semiconductors 1999) one can expect Moore's law to hold for clock frequencies, memory bit density and microprocessor transistor density at least until 2011.

The road map of SIA (the Semiconductor Industry Association) predicts that by 2011 the core of a microprocessor with 5 million transistors will occupy only 0.1 $cm^2$ of a total 6.2 $cm^2$ chip area. At NEC, this

extra space will be used to put more cache space and more processors on the same chip. The philosophy behind this is that it is better to do more with a simple, known design, than to develop new, complicated chip designs.

The clock frequency of microprocessors will continue to increase exponentially, doubling approximately once every two years, as it has been doing the last several years. Clock frequency on/off chip will diverge more and more, maybe to a factor of 2-3 in 2010 (while it is the same today).

CMOS will be the technology used for processors. FP-CMOS [flexible (voltage controller etc) parameter CMOS] may be a way to go, leading to "intelligent" chips.

DRAM technology will be used for memories. However, MRAM, FE-RAM are also interesting approaches.

**Computer building blocks**

Most research and development of processors and other components goes into the commercial market. However, processors for the consumer market like mobile phones, etc. aim at low power rather than high performance. The price/performance trend for HPC systems will therefore probably saturate. This point is further discussed in the "Computer systems" section on the next page.

Both COTS and proprietary building blocks will be used in future HPC systems depending on the application. General purpose processors will be used for for capacity computing, while special purpose vector processors will be used for capability computing. See also the discussion about capacity versus capability computing in the "Computer architectures" section below on this page. The exception is interconnects which will be proprietary. It is likely that the vendor will become more of a system integrator. We are already today! However, Beowulf clusters is a foe to NEC.

**Computer architectures**

Different architectures (different kinds of memory and processor distributions) are most suitable for different types of applications. *Capacity computing* — requiring high workload and throughput, typically a very large amount of non-challenging computations — fits best on microprocessor based MPP or workstation clusters. *Capability computing* — requiring short wall clock time for large critical computational problems — on the other hand, fits best on SMP with powerful vector processors. Clustering of SMP's is a niche market. It can survive, however, since it is the only way to get performance for certain time critical applications. Both Capacity computing and Capability computing solutions are provided by NEC through the IA-64 scalable servers and the SX-5 vector supercomputers, respectively. Mostly shared memory architectures will be used in the future.

In general, there will be many computer architectures for HPC in the future. Clusters of SMP's will dominate. However, the number of

HPC vendors will decrease since many vendors will go to high-volume markets.

The imbalance of processor performance versus memory bandwidth and latency is an increasing problem. Decreasing the size of the total system will improve the latency, but not the bandwidth. The memory hierarchies will be memory caches plus registers. Hierarchies will be limited to 3 levels, since it is too complicated to have more. High internal bandwidth is needed, which means that internal networks must be proprietary. General purpose, de facto standard networks might be used for external networks, which means there will be a compromise between price and performance in this case.

Networks could be classified as "Communications networks" (for simple transfer of data using TCP/IP), and "Interconnects" (physical facilities between nodes or clusters or other system elements). Several alternative standards exist, both for "Communications networks" and for "Interconnects". The trend in the near future for "Communications networks" is 10 Gbps Ethernet which will be used 2003–2004; ATM (Asynchronous Transfer Mode) which will be used in WAN, but will not play a major role in LAN; TCP/IP protocol which may be supported in InfiniBand (10 Gbps) if it wins popularity as a network technology. "Interconnects" can be further classified in "Standard" (Ethernet TCP/IP), "New architecture" [InfiniBand, Gigabyte System Network (GSN), Myrinet] and "Original architectures" (IXS, CrayLink). Exactly which product that will be used by NEC will be a compromise between performance requirements, and (1) the stability of the company providing the network, (2) tested and certified products, and (3) standardised and popular architectures.

**Computer systems**

How large HPC systems can we expect in the future? The number of processors in the largest systems depends on the budget! There is no technological limit. However, if the number of components increase, a less reliable system is obtained. Therefore, a more powerful system might imply a less reliable system.

The price/performance factor for future systems will probably saturate. This is a result of that more money is put into development of consumer market area products, which leads to more expensive components for high-end solutions. In this respect the focus on high volume commercial products markets will have a negative impact on the development of HPC systems.

A general observation concerning performance is that the ratio of sustained performance versus peak performance of HPC systems will decrease. However, apart from performance, there are also other factors which should be considered when different systems are compared. One such important, but often forgotten factor is the TCO (Total Cost of Ownership).

### Linux and open source

Regarding open source NEC's position is as follows. Linux is OK for many purposes, but we can not leave NEC Unix as open source. It is too specific and not of common enough use. Thus, proprietary Unix, or possibly special versions of Linux, will continue to be used for special hardware, like NEC vector computers. In fact, Linux may become proprietary, like Unix did.

### Storage

The need for storage is increasing. The use of several different storage architectures is growing: DAS (Direct Attached Storage), SAN (Storage Area Network), NAS (Network Attached Storage), and IP-SAN. IP-SAN provides inter-operability between computers and storage devices through the Internet. However, a single solution does not fit everyone. Storage architecture should be selected to reach the best solution. In some cases a combination of different architectures should be used.

The importance of size and performance of storage solutions is increasing. This is due to the changing view on information as a real asset. It is also due to the remarkable increase in the amount of data.

Tapes will continue to be used for long term storage. Tape technology will improve in the near future. Furthermore, tapes will always be cheaper than disks.

It is expected that the storage capacity of hard disk drives will double each year, at least for the near future (2001–2004). During the same period the storage capacity of tapes, and the transfer rates of hard disks and tapes will have an annual growth rate of 60 %.

### Programming

Fortran, C, C++ will be the programming languages used in 5–10 years from now. Java will not be used. F−−, UPC, and Shmem are niche products. HPF, however, will be used.

MPI, HPF and OpenMP will be the programming models used for parallelization. These are and will be used a lot in the Earth Simulator Project.[5] The ENACTS user questionnaire shows that the usage of OpenMP is low. This may be due to wrong expectations. OpenMP is also difficult to use, since it is not automatic enough. In general it is likely that the pain/gain ratio of programming will not change in the future; the pain will probably increase.

### Applications and benchmarking

Some application areas which will benefit from future HPC systems are genetics and climate research. These application areas will grow. Tele astronomy will decrease.

One can not adjust architectures to software of today. Architectures must be general, not targeted towards any specific code! Thus, 3$^{rd}$ party codes, or any other codes, will not have a big impact on architecture.

Regarding sharing, visualisation and dissemination of data nothing will be radically different from today. Network performance is the key to fast sharing of data.

Benchmarks will be done as they are today. Application based benchmarks have never got off from the ground. However, the *specific* applications must be benchmarked. A procurement of a HPC computer can not be based on the top500 list. It should be noticed that benchmarking creates additional work, but no additional value, to the vendor.

#### Grid

One scenario of the Grid is resource sharing. However, the tools to use it are not available. The Grid is too immature. The day-to-day priorities of the users are more important than to try out new things. Users can not afford the time it takes to test future technologies. These are some possible reasons why the ENACTS user questionnaire shows that the experience of Grid enabling technologies is very low. The Grid will have no impact on computer architectures, but it influences software and networks.

#### Future computer centres

A future computer centre might be a "network" centre, perhaps a Grid of centres. Many research disciplines will be supported within each centre. One single machine never fits all!

## 3.7 David Snelling (Fujitsu)

Fujitsu was represented by Dr. David Snelling at the interview. The interview was held at NSC, August 22, 2001. The interview had the format of an interactive view-graph presentation given by Dr. Snelling, with open discussions and questions answered during the presentation.

Dr. Snelling is with the Fujitsu European Centre for Information Technology (FECIT), Ltd, UK. He has a long standing interest in parallel computing and parallel architectures. His primary focus is DataFlow and Multi-Threaded Computer Architectures and their execution environments. Before joining FECIT, Dr. Snelling worked at the Centre for Novel computing at the University of Manchester, and before that Dr. Snelling was a Lecturer in Computer Science at the University of Leicester and a consultant in multi-processing at the European Centre for Medium-Range Weather Forecasts. Dr. Snelling received his PhD in 1993: "The Design and Analysis of a Stateless DataFlow Architecture".

### 3.7.1 Summary of key ideas of David Snelling

The most important key ideas conveyed at the interview with Dr. Snelling, Fujitsu, can be summarised as follows:

1. The transistor "real estate" density on the processor chips is growing. These extra transistors will be utilised for, e. g., multi-core designs, and on-chip caches.

2. The memory gap is increasing. Bandwidth is going to be a more and more important issue in the future. Dr. Snelling foresees a future when we are buying bandwidth rather than Flops.

3. Moore's law for sustained performance is a perpetuated myth.

4. Fujitsu is now going down the commodity path. Actually market driven proprietary building blocks will be used. However, the HPC networks will still remain proprietary for some time to come.

5. The memory and processor distribution which will be used by Fujitsu will be clustered (large) SMP's, with modest hierarchy (2 levels of cache).

6. The sustained system performance as measured in percentage of peak performance will continue to drop.

7. It is the services you put on top of the system software that produce the commercial value. The particular kernel could be based on open source. An important issue related to open source is the availability of good compilers. One should also not forget Microsoft in this discussion.

8. Fortran, C++ and Java will be the dominating programming languages used for HPC systems in the future. Mixed mode parallelization using MPI and OpenMP in combination will be critical within 5 years from today. Programming for HPC systems are facing a serious set of challenges such as development of problem solving environments for HPC applications.

9. Benchmarks are not cost-effective for the vendors and will have to change. The IDC initiative regarding new benchmarks is probably the right way to go.

10. There are many problems related to the development of the Grid such as accounting, units to be used, "exchange rates," brokering, financing of the infrastructure, and security. Dr. Snellings personal view is that "meta computing is a ridiculous concept" while Fujitsu would disagree on this point. Distributed *non*-computing resources such as data Grid is more reasonable.

11. Future HPC centres will be very commercial and market driven. Service and support will be much more important than CPU seconds. Added value will be on the service side rather than on the system side. Computer centres may become very special purpose with special purpose hardware. The trend (in Europe) is towards fewer and bigger centres.

### 3.7.2 Interview

#### General remarks

Before the interview Dr. Snelling points out that he does not represent Fujitsu Limited with respect to strategy, policy, or commitments of any kind. He also says that his answers are likely to contain errors, and that these are his errors, not Fujitsu's. Furthermore, he says that some answers may appear vague, as he needs to avoid confidential details.

#### Moore's law and hardware development

The transistor "real estate" density on the processor chips is growing. These extra transistors will be utilised by "brute force". For example, multi-core designs, and on-chip caches will appear.

End users are not seeing the benefits from the "silicon development" like the increasing transistor density. It is the latency between the memory and the CPU that makes the difference. Furthermore, the gap between processor speed and memory bandwidth is increasing. Bandwidth is going to be a more and more important issue in the future. Dr. Snelling foresees a future when we are buying bandwidth rather than Flops.

The (peak) performance is determined at the chip design level while sustained performance is determined by the complete system. One should note that integer registers, which are important for commercial applications, tend to get the best performance (locations) on the chip, as compared to floating point registers, which are important for HPC market applications.

"Moore's law for sustained performance is a perpetuated myth," says Dr. Snelling. As soon as you start looking at sustained performance we are not even close to reaching Moore's law.

#### Computer building blocks

Fujitsu is driven by the commercial market, rather than impacted by it. The next step is that commercial market will be driven by the embedded market. Fujitsu is now going down the commodity path. Actually market driven proprietary building blocks will be used. Fujitsu still builds everything, but it is SPARC compatible so it is driven by the market. The HPC networks is a different business. They will still remain proprietary for some time to come. A trend is that the vendor will become more of a system integrator.

#### Computer architectures and systems

The memory and processor distribution which will be used will be clustered (large) SMP's, with modest hierarchy (2 levels of cache).

The sustained system performance as measured in percentage of peak performance will continue to drop. One reason is that the processor-memory gap widens. Therefore, latency hiding must be used throughout the system. The increase in clock speed will force longer

pipelines. Truly pipelined networks are unlikely, however. The market will actually discourage them.

Electric power will become increasingly important in future HPC systems. This is probably one factor which will set the upper limit for very large systems.

### Linux and open source

Regarding operating systems it is less important which particular kernel that will be put "under the hood", it could be based on open source. It is the services you put on top of the system that matters. The commercial value will come from software put on top of open source operating systems. For example, parallel NAVi will be sold on top of Linux in the future.

Fujitsu has a Linux group. However, what will survive is really a support issue. Do not forget Microsoft in this respect! For example, it is important to note that much development is done on desktops running Windows. Another important support issue related to Linux and open source is the availability of good compilers. It is not certain that such compilers will be provided by the Open source community.

### Storage

Dr. Snelling says that he is not an expert on storage systems. Regarding components for storage Fujitsu buys disks from whichever provider is considered best.

### Programming

Fortran, C++ and Java will be the dominating programming languages used for HPC systems in the future. The usage of Java will increase because programmer skills will be available mainly in Java. This is because Java is used at the bottom of the market, for low power, embedded systems such as mobile phones etc. These programmers will bring Java to science (physics etc.) when bored at mobile phones.

Programming for HPC systems are facing a serious set of challenges. Dr. Snelling does not see any development of problem solving environments for HPC applications.

Fujitsu is now making a transition from vector to scalar computer systems. However, from the programming point if view, the vector codes will provide very good starting points for data-parallel programming. Vector codes also provide a good starting point for instruction level parallelism. Parallelism is needed on all levels in order to survive in the long run.

Mixed mode parallelization using MPI and OpenMP in combination will be critical within 5 years from today. There are very complex interactions when using these paradigms. No new paradigms will appear, apart from perhaps invisible parallel level underneath Java.

### Applications and benchmarking

Regarding applications which will use HPC resources in the future: The ENACTS user questionnaire indicates little or no change, according to the users!

Concerning visualisation Dr. Snelling is being somewhat provocative saying that he is "not sure if it is a tool or toy." Of course, visualisation is important to display large data sets such as weather simulations. It could also form the basis for collaboration by means of collaborative environments etc.

Linpack SPEC benchmarks are next to pointless since they are directly proportional to clock speed. SPEC codes are tightly optimised, and compilers recognise them. In general benchmarks are not cost-effective for the vendors. It is a very expensive way not to win a deal, and should be made differently. Benchmarks should form the basis for vendor selection and contracts. The IDC (International Data Corporation) initiative regarding new benchmarks[30] is probably the right way to go. However, the selection of benchmark codes will determine which research groups that will get the best performance!

### Grid

Results in the ENACTS user questionnaire shows that most users are interested in Grid solutions in order to get access to the fastest CPU available. Everyone is going to go for the fastest systems.

There are many problems related to the development of the Grid. For example, accounting is a critical issue. What units should be used? What should the "exchange rates" be? How should brokering be handled? Who will pay for the infrastructure? How can a uniform security model be achieved?

"Meta computing is a ridiculous concept!" says Dr. Snelling, but stresses that this is his personal view, and Fujitsu would disagree. Solutions like SETI@Home is OK, but e. g., Gaussian will never run behind people's screen savers. Meta-computing is also side-stepped by, e. g., the developers of Unicore.[15] (However, the follow on Unicore projects, Unicore Plus and EuroGrid, are no longer side-stepping meta-computing.) Distributed *non*-computing resources is more reasonable and is what data Grid is all about. Other distributed non-computing resources, like experimental facilities, telescopes, etc., is OK, but this is already available.

### Future computer centres

Future HPC centres will be very commercial and market driven. Service and support will be much more important than CPU seconds. The service you provide will be homogeneous. Added value has to be on the service side rather than on the system side. For example, DWD (Deutcher Wetterdienst) will start selling meteorological services to the rest of Europe. Partnership with vendors will also be more common.

What do you sell as a computer centre? CPU seconds are not acceptable. Instructions executed $\times$ memory is a better measurement. Higher levels of abstraction is another selling quantity. Brokering of Gaussian is a good example.

Computer centres may become very special purpose with special purpose hardware such as, e. g., IBM's Blue Gene computer used for genomics. Fewer and bigger centres is the trend in Europe right now.

A very large centre can run several procurements at the same time and always work with more than one vendor at each given time. This would be an advantage of the centre, since it opens up the possibility to combine vendors with different profiles.

Distributed centres are possible, but not for single distributed applications. On the other hand, centres can collaborate and take turns on making procurements and reach the economy of scale.

# Chapter 4

# The HPC Roadmap and the Grid

***Addendum by J-C Desplat (EPCC), 26$^{th}$ February 2002***

Surveys of leading edge technologies have the annoying habit of getting outdated within short time-scales of their publications. Unfortunately, the reports published by the ENACTS consortium are no exception to the rule and particular issues uncovered by our reports may require further clarification shortly after their publication. In the current study, the "HPC Technology Roadmap" by NSC and CSCISM, the most noticeable change is the rapidly evolving attitude of vendors about Grid-related technologies.

So, have vendors actually changed their opinion on Grid Computing? Well, the short answer is "yes and no!" Although recent months have seen major players such as IBM and Sun Microsystems getting increasingly involved in standardisation efforts (e. g., as part of fora such as the Global Grid Forum (GGF)), one may argue that it is principally because the concepts encompassed by Grid Computing, and indeed their relevance to the vendors' core businesses, are now significantly overlapping. Let us consider the reasons for this shift in more detail.

## 4.1 Making the Grid relevant to business and industry

In January 2002, Steve Tuecke (Argonne National Lab), one of the main architects of the Globus toolkit, presented at the UK National e-Science Centre the future technology underpinning Globus 3. This new technology, baptised the 'Open Grid Service Architecture' (OGSA), stems from the integration of Grid technologies (as characterised by Globus 2) and Web services such as found in IBM's WebSphere or Sun's Java2 Enterprise Edition (J2EE). OGSA will permit the creation of composite Web services that link sites that are in different administrative domains

or companies, e. g., for business-to-business (B2B) transactions, real-time supply chain integration, or improving interoperability following company mergers for instance. This convergence is in itself a revolution as it results in making the development of "the Grid" highly relevant to vendors and developers and raised the interest of major corporate companies (such as Oracle and Microsoft) until now only marginally associated with Grid Computing.

The last GGF held in Toronto proved a real milestone with the creation of a number of Working and Research Groups focusing on topics dear to vendors and application providers. The best example is probably the newly-formed group on "Databases and the Grid" looking at issues such as "Data requirements for the Grid", "Databases and the Grid", and "Database access and integration services and the Grid" [3, 14]. Anyone aware of the technical challenges and economical stakes posed for businesses by (safe) database integration in a distributed heterogeneous environment will easily perceive the importance of this new development. It is particularly pleasing to point out the active participation of European researchers and developers within these new Working groups, as exemplified by the lead role taken by the Grid database taskforce of the UK e-Science programme [3, 14].

Further information about the OGSA architecture can be found in [9].

## 4.2 Impact on the vendors' strategy

Restricting our analysis to two of the most influential vendors in HPC, namely IBM and Sun Microsystems, the impact of the convergence of Grid and Web Services technologies reported in the previous section can be summarised as follows.

Although scientific Grid computing (generic computational Grids and data Grids) is still supported by IBM, e. g., through its support to set up the Dutch Grid [4] or the Distributed Terascale Facility [10], this involvement remains marginal and is mostly motivated by the potential applicability of some of the underlying technology to business and industry. As Dave Turek (IBM vice-president of emerging technologies) stated himself about the above activities [43]:

> "We expect the technology to find its way rapidly into more conventional commercial development."

Since the advent of the OGSA though, IBM's commitment to Grid Computing has been considerably reaffirmed. For instance, IBM announced that $4 billion worth of Grid-related projects will be developed at IBM Armonk (N.Y.) over the next few years [29]. The increasing relevance of Grid computing to business and the industry has been emphasised by the same Dave Turek [29]:

> "It [the Grid] especially has attracted a lot of interest by large companies that can use it to connect geographically diverse offices and unify the supply chain[. . . ]. It can also be used to

> create virtual databases that pool the content from different offices, and take on large engineering and modelling projects beyond the scope of any single computer."

Another evidence of IBM's substantial involvement is exemplified by the lead role played by Jeff Nick, one of their senior technical staff, in the specification of OGSA [27]. Shortly before GGF4, Irving Wladawsky-Berger, IBM's vice president of technology and strategy, announced that IBM and the Globus Team are expected to propose a number of distributed protocols centred around security, authentication, identification and collaboration. It is also expected that IBM's other strategic initiatives, such as their support for the Linux operating system and the eLiza project [6], are likely to play a role in their overall Grid strategy. For instance, recently announced work on toolkits enabling users to "grid-enable" their applications will be developed for both AIX and Linux.

Testbeds for Grid-based Web services are already being set up. The most famous example is probably the Grid developed around the University of Pennsylvania which will connect the servers and databases of hospitals around the US to share information on mammogram procedures and research [47].

Sun Microsystems' interest in Grid technology has also considerably strengthened over the past year. First, by actively supporting the peer-to-peer (P2P) model through the development of its JXTA technology [12]. P2P is indeed tailored for B2B commerce, as companies can use P2P to order from suppliers and serve customers. As reported in [23], a variety of companies, including Intel, GlaxoSmithKline, Raytheon, the law firm Baker & McKenzie, the accounting firm Ernst & Young and First Union, are discovering that this technology can help them do business faster, better and cheaper.

Sun is also extending the functionalities of its portfolio of Grid-related software solutions like the Sun Grid Engine, Grid Engine Enterprise Edition, iPlanet Portal Server, Sun Management Centre, HPC ClusterTools, Jini and Jxta, and SunCluster. They have also recently announced the availability of Sun Grid Services, a combination of their Grid computing software with the Sun ONE Web services [13]. As Wolfgang Gentzsch, Engineering Director for Grid Software at Sun Microsystems, explains [20]:

> "We [Sun Microsystems] offer basically two ways of combining our two software stacks, Sun Grid and Sun ONE. The first solution is via the industry-proven iPlanet Portal Server for secure, remote and transparent access to Sun ONE Web services and to Sun Grid computing services. This obviously leads to one global access point for businesses, which first want to process data in their grid, and make the results available as a service to employees or customers.
>
> The second solution is via Sun ONE Connectors to major applications, like ERP and CRM, or simply via JSP or EJB Containers that wrap applications inside a Java Bean. Once

> integrated with our Java Beans technology, the grid application is easily integrated into a Web service. This technology allows our customers to build their own scalable compute services and offer them through Sun ONE, as services on demand. In such a combined environment, Sun ONE manages users, customers, communities, policies, and even access to grid services, while the grid manages computing, data processing, storage, efficient utilisation of resources, collaboration, and more."

Sun Microsystems has been particularly engaged in promoting the concept of commercial enterprise Grid and claims to have already helped numerous customers deploying such Grids for over a year. Based on its Grid Engine Enterprise Edition, Sun's approach strives to provide economies of scale, access to one common HPC service for all departments, manageability, reliability and quality of services, reduced hardware and software costs, reduced operational cost, and increased productivity.

Finally, Sun welcomes the advent of OGSA and plans to exploit their experience of Grid services to contribute to its development. Wolfgang Gentzsch stated that:

> "Sun is absolutely convinced that standards and open source are fundamental strategies for the success of Grid computing for suppliers and users alike. This is why Sun was the first systems supplier to place key grid technology – the Grid Engine product suite – into open source, and why we've worked hard to establish Global Grid Forum standards initiatives like DRMAA, the Distributed Resource Management Application API. We welcome and support the OGSA proposal.
>
> We are very pleased to see that the OGSA architecture will be based on the same W3C standards that also our industry-proven Sun ONE Web Services are based upon. Sun will be providing input as an active member of the OGSA working group based on our experience with the large Sun ONE install base we have today."

In a recent interview [20], Wolfgang Gentzsch estimated the total number of software developers currently involved in Grid-related activities to exceed 2,000!

Besides IBM and Sun Microsystems, vendors such as Platform Computing, Hewlett-Packard and Compaq have announced their intention to incorporate the OGSA technology into their products within the near future [38]. Since the release of the Globus Toolkit 2 (GT2), commercial interest in the software has increased dramatically, with companies such as Microsoft expressing their commitment to making it the standard for distributed collaboration.

This partnership between the Globus Project and Microsoft is also worthy a mention. Microsoft is providing $1 million in funding and support for the development of a version of the Globus Toolkit for Windows XP [41]. As pointed out by Ian Foster (Argonne National Lab), one of

the Globus Project managers, this type of partnerships will eventually ensure the adoption of Grid technology by the wider public:

> "The partnerships that the Globus Project has established with both IBM and Microsoft lay the foundation for a far broader adoption of Grid computing, by taking the first steps towards its establishment as a core technology for e-business. Grids have already been adopted broadly for e-science; in the next 2-3 years, we expect to see significant uptake in industry, first for science and engineering applications, and then as part of the increasingly dynamic and integrated e-utility fabric that will power industry in the future."

## 4.3 Disclaimer

The views expressed in this addendum chapter only reflect the author's own perception and analysis.

# Chapter 5

# Summary and conclusions of the technology watch report

## 5.1 Introduction

Who can look in to the future? We all know this is impossible but nevertheless it is necessary. In order to plan our activities we need to have some model of what we think that the future brings.

The interview material we have at hand for this report are made by highly acclaimed, knowledgeable representatives from the most prestigious and well known vendors of HPC systems. It is a very interesting and exciting material to read. All interviewed have stressed that this is their own personal view of future HPC and not necessary that of the vendor, so we have unique personal opinions of HPC future based on experience and knowledge from rich sources, some which is only available to people within the corporations.

Many forward looking statements are extrapolations of today's trend. All material from the interviews are made public so it comes as no surprise that vendor representatives can not share all of their knowledge, especially major alterations in strategies. Nevertheless, having access to such a complete set of in-depth interviews gives a unique snapshot of where the industry believe it is heading.

However, it is hard not to get lost in all the details, especially in an area where there is so much rapid progress as in the computer industry. Everything is moving ahead at a very rapid pace, and some areas faster than others. Below we try to summarise and interpret the interviews keeping the focus on what the implications will be for for the end users and the computer centres.

In section 5.2 we summarise the results from the interviews, trying to be as neutral as possible. Section 5.3 is where we give our views on what we think are the most important trends and developments for the next decade.

## 5.2 Summary of the interviews

### 5.2.1 HPC market

If we start by looking at how market is viewed, HPC today has a rather small share of the total computer market whereas the commercial (enterprise) and consumer markets are much larger. This trend will continue. This also means that a majority of the investments will be in non-HPC markets. There is a difference of opinion on the implications of this however. Either this market trend will force future HPC systems to be integrated directly from commercial components (processors, networks or even complete systems), or proprietary components can still be developed cost effectively for HPC specific solutions since what the consumer and commercial market develops and invests in can be used also for HPC, especially the facilities and methods for making chips.

There is also general agreement that a majority of the investment will move (if it has not already moved) to the consumer market, but again there are different opinions on how this will effect future HPC system architectures. Some believe this means that future systems are going to based mainly on consumer market components even though this might require a whole new programming paradigm, while others believe it will enable new technologies that can be used in more traditional HPC environments.

Constant price pressure and the presence of low budget solutions continue to put pressure on profit margins. Vendors need to look at where they can cut costs. Software is expensive to develop and maintain and the Open source movement is viewed as a way of sharing and thus cutting cost. People intensive tasks such as extensive benchmarking is another area where vendors are looking to cut cost.

### 5.2.2 HPC systems

Future system architecture actually has both similarities and large differences between the interviewed. From a very general point of view one could actually say that all share a view of future systems with:

- Clustered SMP nodes giving a scalable parallel architectures.
- The SMP nodes can be anything from a single processor up to thousands processors.
- These systems are only limited in size by:
  - available money
  - cost of building space and utility (power)
  - reliability
- General purpose systems (but special purpose systems is not ruled out).
- Non uniform memory access (but with very large variations in “NUMAness”).

This is also reflected by a general agreement on MPI and OpenMP as the parallel programming models for the present as well as the future.

On the other hand though, underneath this very general common view almost everything else can differ: the processors, memory subsystems, node interconnects, compilers, operating systems, power consumption, cooling, square foot usage etc. and it equally fair to say that we really are talking about very different computer systems.

As a complement to general purpose systems, special purpose systems will emerge. These systems will be based on consumer product type of components, giving very high price/performance ratio but requiring very different programming models. Initially they will be targeted towards niche markets with suitable applications such as life sciences.

In addition to vendor specific architectures, low cost cluster HPC systems (Beowulf) has opened up a whole new sector for HPC. From being a rather limited solution it has matured significantly over the last years and continue to evolve, particularly in the high end. Most of the vendors claims that Beowulf is complimentary to their offering. Many vendors now also offer their own Beowulf clusters. It is yet unclear how this market will evolve and who the main players are going to be. It is likely however that Beowulf clusters will blend into the HPC world more and more seamlessly. Just as with traditional HPC we also believe there will be a multitude of various HW and SW configurations for Beowulfs.

### 5.2.3 Building blocks [processors , memory (bw, latency), network, storage, graphics etc. ]

Silicon will continue to be used for chips for the next 5–10 years. Moore's law is believed to continue. Important to note however is that this is for the "original" Moore's law, the doubling of transistors every 18 months. Several questions can be derived from this fact: How are all new transistors going to be used? How is this translated into sustained performance improvements? Answers to these questions were expressed differently from different persons.

Number of processor architectures continue to decrease, but there is no general agreement as to if this means death to special purpose processors. High volume general purpose processors has to be very general purpose so performance for HPC applications will not be optimal. On the other hand is the question open if the cost for developing special purpose processors will be prohibitive.

Memory subsystems will continue to be a differentiator between vendors. Microprocessor performance will continue to increase much faster than memory performance. Latency is getting to the point where the latency (in terms of clock cycles) will actually start to increase. Instead latency reduction techniques such as prefetch, multi-threading, deep memory hierarchies etc. will be applied, and this is an area where many different techniques will be developed. Codes with regular memory patterns are likely to benefit more than others by many of the latency reduction techniques. Vectorisation is only actively pursued by

two vendors but could also be viewed as a latency reduction technique. Bandwidth is the main focus for some vendors. It is likely to become one main differentiator between different systems and could replace the whole vector/no-vector debate.

Mass storage systems will increasingly be networked rather than directly attached. Disk density will double every year and tape density somewhat slower. I/O rates are not likely to keep up with capacity increase. No consensus if disks will overtake tapes for long term storage. Alternative mass storage solutions are still far from being commercially ready.

Graphics was touched on in most interviews and there is strong consensus that increased bandwidth and at the same time very limited increase in display resolution will change the way some visualisation is done. Centralised graphics servers will be used with thin clients for display only, even at large distances.

### 5.2.4 Parallel programming models

No one is predicting any major change in the parallel programming models, most vendors seems be of the opinion that whatever the hardware looks like it must be possible to program in either MPI or OpenMP. There are also some voices for co-array Fortran and UPC but is very unlikely that they will be as dominant. The hybrid MPI+OpenMP programming model is favoured by some but it remains very much unclear if this is going to be widely used. HPF seems to to have a future in Japan only.

### 5.2.5 Programming languages

No surprises here: Fortran, C, C++ and maybe Java continue to dominate.

### 5.2.6 Software tools

There seems to be a potential for a conflict when it comes to the necessary software tools for debugging and performance analysis of very large systems. Vendors are more and more shying away from developing the necessary SW due to cost, instead directing customers to commercial $3^{\text{rd}}$ party software or open source. At the same time, even though the parallel programming models are the same across different platforms, the hardware underneath might be very different and thus requiring system specific tools to be able to fully understand and utilise the most out of the system.

### 5.2.7 Pain/gain

The pain versus gain is not going to be improved, probably get worse if not much worse to be able to tame the biggest and most powerful

systems. Almdahl's law is a stern task master, especially with processors counts reaching 10 000's. Maintaining very large systems will also require exceptional efforts. Housing, powering, maintaining and scheduling system with ten's of thousands of processors is not for the faint of heart. It is no coincidence that public presentations of new very large systems spend a significant time talking about the building size, power required, pictures of cables etc.

### 5.2.8 Operating systems

Linux is the driving force around open source operating systems in the HPC world. No one can afford to stay outside of Linux. How HPC specific features are going to make it into Linux is not yet clear and it seems like the initial euphoria has settled down a bit when vendors are starting to realise that working with the Open source community is very different from doing your own development. One way or the other we still believe that Linux will be the main HPC operating system 5 years from now.

### 5.2.9 Price performance (peak vs sustained)

There is general agreement that price/performance will continue to improve at current rates, which is approximately Moore's law. Important to note however that this is for peak speeds. When it comes to sustained performance, the picture is much more vague and most answers indicate that there will be an even stronger dependence on the type of application being used. Highly parallel applications with low bandwidth requirements will increasingly see a much better price/performance than traditional memory intensive HPC applications that are not trivially parallel.

### 5.2.10 Benchmarks

Almost unanimous agreement that benchmarking for customer procurements will have to change. The cost is too high for this in today's market where margins are getting smaller all the time. Requirements for better general benchmarks is there but the awareness of current efforts such as IDC's was low.

### 5.2.11 Grid

Most believed the Grid is still immature today but will become important in the future, but the opinions differed substantially in how and what way this will happen. Direct experience seemed low and one could argue whether the response might have been different if the people interviewed had more experience of the Grid. Interestingly to note was that no one thought that the Grid and Grid development would have any impact on future system architecture. It is rather something that their systems are used for and an opportunity to sell systems!

### 5.2.12 Future computer centres

HPC computer centres will have a place in the future according to the interviewed. Many different reasons were given but the fact that future very high performance will require even more expertise in various fields such as hardware, applications, performance analysis/tuning, logistics, procurements etc is reason enough to believe in a future for HPC centres. Application specific centres were predicted by many.

## 5.3 Our conclusions of the technology watch report

If anyone believed that future HPC systems all would move towards a a unified computer architecture they will be disappointed. New advances in all relevant computer technology made possible by an increasing market as well as the Open source movement instead make the possibilities even larger than before.

We see five major trends for the next 5–10 years:

1. We believe that the increasing mass market for computer and consumer systems will provide the basis for a continued exponential technical development in all areas of computer systems development. This will mean many new and exciting possibilities also for HPC computer systems and we will continue to see a proliferation of systems instead of a convergence towards some common platform. A substantial part of future HPC systems will be provided by non- traditional HPC vendors. A major driving force behind this will be Linux clusters used for HPC. Off the shelf components means much lower profit margins which will create new and different business models than what we have today. New HPC vendors will appear, some will merge and some will disappear.

2. We see future HPC systems are parallel, scalable computing architectures that are based on the notion of clustered SMP:s. Scaling is possible by either adding nodes and/or increasing the nodes themselves. A significant portion of the components are based on COTS but proprietary components play an important role. Operating systems is Unix with Linux as target. SMP nodes can be anything from a single processor up to thousands of processors and will have large variations in processors, compilers, memory subsystems, node interconnect and mass storage. The upper limit in size is set by the prize, not the technology.

3. Peak performance as well as price/performance will continue to improve according to Moore's law. This is not true for sustained performance however where there will be an even more profound dependence than today between performance and application characteristics. Applications that easily can take advantage of advances in consumer product technology might even get a "super"

Moore's law performance increase. Life sciences are in that category.

4. Parallel programming models and programming languages will be on an evolutionary path rather than a revolutionary. MPI and OpenMP for parallel programming and Fortran, C, C++ and Java for programming languages. This is good news for the user. The bad news is that to achieve high performance, detailed knowledge about the application as well as the underlying hardware is still required. Getting to extreme performance requires even larger efforts than today due to the sheer size of systems that will be available. Lack of adequate tools for software development will likely become a key issue for HPC.

5. The Grid is evolving fast and we strongly believe it will play a key role in the way HPC is used. In the context of this report, we see one new key role will be to serve as a layer between the user and the HPC computer system. Hiding details of not only geographical location but (maybe more importantly) system architectures and usage will in many cases enable the continued proliferation we see in HPC system architectures. We believe HPC vendors should consider this when developing new systems. This development should be very straightforward as long as we are talking about users of $3^{rd}$ party codes and straightforward compile and run applications. Many high performance user however still needs to be able to be close to the system and this is an area where the Grid will take longer time to develop.

For HPC centres we believe that they will be needed but the rapid development of ever faster networks, Grid and Grid middle-ware, increased pain/gain ratio, increased system complexity and the increased infrastructure cost for logistics (power,housing) etc will mean fewer but larger centres. HPC centres will have to compete with other centres and to be successful we have the following recommendations:

- Collaborations and teaming up with other centres. Resource sharing.
- Expertise in many different areas of system architecture.
- Involved in open source development.
- Continued emphasis on MPI and OpenMP.
- Performance expertise, increasing. Vendor support decreasing due to leaner profit models.
- Expertise in Grid infrastructure, middle-ware.
- Application specific centres.

# Part II

# Case study report: Implications for molecular sciences

by Filippo De Angelis, Francesco Mercuri, Marzio Rosi, Antonio Sgamellotti, Francesco Tarantelli, and Giuseppe Vitillaro

# Chapter 6

# Molecular sciences

## 6.1 Overview

From the interviews to the computer companies representatives, the following scenario appears to characterise the near-future development of computer hardware and architectures:

Almost without exception, the experts interviewed concur that Moore's law will continue to describe accurately the pace of further chip integration and the resulting trend in theoretical peak performance of computers, at least for the next 5 years. Similarly, there is wide agreement on the (short-to-medium term) exponential increase in communication bandwidth. However, among the most important consequences of this for high performance computing is that memory — and more generally communication — latency as measured in CPU cycles will increase rapidly and impede most applications from reaping fully the benefits of Moore's law. The principal architectural features proposed to partly alleviate this problem aim at hiding latency behind an increased depth of data storage hierarchies and/or a moderate-to-high level of hardware multi-threading.

Increased transistor integration will also have the likely consequences that, on one hand, more room for extra logic will be available on-chip, allowing a widespread development of relatively low-cost special purpose processors. On the other hand, high integration will probably mean that not all transistor will be able to sync to one clock and multiple-clock chips will appear.

It is likely that, with various shifts of balance, we will continue to witness the development of both relatively low-cost Commercial-Off-The-Shelf (COTS) computers and proprietary hardware, the latter delivering higher performance at a much higher cost. SMP clustering will presumably be the dominant architecture. The total number of processors comprising a HPC machine will be in the 10000's, with consequent problems of reliability. These will have to be addressed by redundancy and statistical algorithms, eventually merging into true AI management systems.

Disk drives for I/O systems will likely continue to increase in ca-

pacity at a much faster pace than I/O bandwidth, so that the latter will become the limiting factor. Networked I/O systems, based on SAN, NAS and Internet SCSI will become widespread.

The most popular programming languages currently used (Fortran, C, C++, Java) will continue to dominate software development. Interestingly, it is generally agreed that the programmer's pain/gain ratio is destined to increase, signalling that compiler and RTE technology development will not keep pace with hardware complexity and speed.

It can be safely assumed that Unix-like operating systems will continue to dominate HPC. Linux will rapidly increase its market share, although proprietary OS will be phased out very slowly, if at all.

Based on the above assumptions we have projected two probably typical HPC computers 5 years from now, within the price range allotted, one exemplifying a proprietary SMP cluster and the other being the natural development of current Beowulf clusters. Based on the interviews and current trends, we have assumed that, within the projected time span, the price per processor will remain roughly unchanged.

## 6.2 State of the art and evolution of prototype hardware

The main state-of-the-art features of two typical HPC systems are reported in Table 6.1. Single processor and node characteristics pertaining to two representative systems are shown; a proprietary system based on IBM Power4 processors and a COTS-based systems, in which Intel Pentium III-Pentium 4 processors are considered.

| | Proprietary Hardware (IBM Power 4) | Hardware COTS (Intel Pentium 3-4) |
|---|---|---|
| Processor | | |
| clock frequency | 1 GHz | 1-2 GHz |
| peak power | 5 Gflops | 500Mflops/1Gflops |
| memory bandwidth | 5 Gb/s | 800Mb/s-1.5Gb/s |
| price | 12,500 Euro | 2,500 Euro |
| Node | | |
| max SMP | 32 | 4/8 |
| max memory | 256 Gb | 4/8 Gb |
| dimensions | 5-10 units | 4-5 units |

Table 6.1: State-of-the-art performances of proprietary and COTS hardware

In Tables 6.2 and 6.3 the 5-years projections of the prototype computing systems reported in Table 6.1 are shown. Such a projection has been obtained by barely applying Moore's law over the time span of 5 years (60 months) which results in a Moore factor of 3.3. This corresponds to a general performance increase of $2^{3.3} \approx 10$. Moreover,

in Tables 6.2 and 6.3, the features of two different computing systems are shown, corresponding to a cost of 12 and 3 million Euro (MEuro), roughly representative of available resources of a big computing centre and a departmental structure, respectively.

| | | |
|---|---|---|
| Processor | | |
| clock frequency | 10 GHz | |
| peak power | 50 Gflops | |
| memory bandwidth | 20 Gb/s | |
| price | 12,500 Euro | |
| Node | | |
| max SMP | 64 | |
| max memory | 512 Gb | |
| dimensions | 5-10 units | |
| | 12 MEuro option | 3 MEuro option |
| 64-processors nodes | 16 | 4 |
| 4-nodes racks | 4 | 1 |
| processors | 1024 | 256 |
| RAM per node | 256 Gb | 256 Gb |
| memory | 4Tb | 1Tb |
| switch | yes | yes |
| peak performance | 50 Tflops | 12 Tflops |
| 1 cache re-use | 5 Tflops | 1 Tflops |

Table 6.2: IBM SP 5-years predictable evolution

## 6.3 Quantum chemistry methods

The study of the electronic properties and of the reactivity of complex molecular systems by means of computer simulations, is an area of particular interest in modern chemistry, both from a theoretical and practical point of view. In particular, the potential applications of computer simulation techniques in the field of material sciences, catalysis and molecular sciences in general, have been recently extended to the study of biological systems. Modern computational techniques and increasing computer power, together with theoretical advances, allow nowadays the study of complex molecular systems of large dimensions at a very high degree of accuracy.

In particular Density Functional Theory (DFT) can provide an accurate description of the electronic properties of a wide variety of systems, and the generality and potential of DFT have been recognised by the international scientific community. The Nobel prize for chemistry to Walter Kohn and John Pople has definitively established the general value of DFT for chemical applications. Moreover, the possibility of performing DFT-based classical Molecular Dynamics (MD) simulations of complex systems, through the Car-Parrinello (CP) method, and the

| | | |
|---|---|---|
| Processor | | |
| clock frequency | 10 GHz | |
| peak power | 10 Gflops | |
| memory bandwidth | 10 Gb/s | |
| price | 2,500 Euro | |
| Node | | |
| max SMP | 16-32 | |
| max memory | 16-128 Gb | |
| dimensions | 4-5 units | |
| | 12 MEuro option | 3 MEuro option |
| 32-processors nodes | 150 | 37 |
| 8-nodes racks | 19 | 5 |
| processors | 4800 | 1184 |
| RAM per node | 64 Gb | 64 Gb |
| memory | 9.4 Tb | 2.3 Tb |
| switch | yes | yes |
| peak performance | 48 Tflops | 12 Tflops |
| 1 cache re-use | 4 Tflops | 1 Tflops |

Table 6.3: Beowulf Intel 5-years predictable evolution

computational efficiency of the scheme proposed by the two authors, together with parallel computing, allow us to accurately investigate the electronic, structural and reactivity properties of systems of increasing complexity. This represents the state of the art in modern chemistry.

### 6.3.1 Schrödinger equation

The electronic properties of a quantum system may be in principle calculated by solving of the Schrödinger equation. Such equation is a differential equation which, for stationary states, has the form:

$$\hat{H}\Psi = E\Psi \tag{6.1}$$

The operator $\hat{H}$ is the Hamiltonian of the system, which, in absence of external fields and for a non-relativistic system, is composed by electronic and nuclear kinetic energy terms and by potential energy terms including nucleus-electron attraction, nucleus-nucleus repulsion and electron-electron repulsion contributions. $E$ is the energy of the system and $\Psi$ is the wave function, which contains all the informations about the system. The wave function is the central quantity in quantum mechanics and is a complex function of all the spatial and spin coordinates of the particles constituting the system. The wave function is defined in such a way so that $|\Psi|^2$ corresponds to the probability density of the system in the space defined by spatial and spin coordinates. $\Psi$ is required to be normalised and anti-symmetric with respect to fermion exchange.

After factoring off the nuclear motion according to the Born-Oppenheimer approximation, the standard non-relativistic Hamiltonian for the electronic problem reads (in atomic units):

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_i \sum_\alpha \frac{Z_\alpha}{r_{i\alpha}} + \sum_{i<j} \frac{1}{r_{ij}} + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}$$

where $\nabla^2$ is the Laplace operator, $Z_\alpha$ are the nuclear charges, $r_{ij}$ and $R_{\alpha\beta}$ are the inter-electronic and inter-nuclear distances, respectively.

Unfortunately, the Schrödinger equation admits exact solutions only for one-particle problems, because of the Coulomb pairwise interaction terms in the Hamiltonian. Therefore solving such an equation for a many-electron system requires the definition of an approximate form for the wave function. The usual approach consists in expanding the wave function in a suitable and typically very large set of linearly independent basis functions, either directly as a linear combination (Configuration Interaction approach, CI) or as an exponential ansatz (Coupled Cluster theories). Thereby, the differential equation problem is essentially translated into an algebraic one.

### 6.3.2 The Hartree-Fock method

The Hartree-Fock method represents, historically, one of the first rigorous approaches to the calculation of the electronic structure of molecular systems. It is the basis of the molecular orbital (MO) approach, in which the many-electron wave function is simply expressed as an anti-symmetrised product of one-electron functions (the orbitals). Such a form for the wave function, called a Slater determinant, represents a severe approximation, for it is based on the idea that the electrons are essentially uncorrelated, each interacting with an external field produced by all the others. The true many-electron wave function must in fact be a superposition of Slater determinants, the combination coefficients of which are such that the total energy is stationary.

In the Hartree-Fock picture, the molecular orbitals $\phi_i$ are obtained as solutions of a one-electron pseudo-eigenvalue equation

$$\hat{F}\phi_i = \epsilon_i \phi_i$$

where the Fock operator $\hat{F}$ contains the potential felt by an electron due to the nuclei and all the other electrons (self-interaction is automatically cancelled) in addition to its kinetic energy. The inter-electronic part of the potential is a sum of Coulomb terms each referring to one orbital in the system. A part of this potential, called *exchange*, is non-local, arising ultimately from the anti-symmetry of the wave function. It is clear that, since the operator $\hat{F}$ depends on the orbitals themselves, the solutions to the Hartree-Fock equation must be sought self-consistently.

The most common way to tackle the Hartree-Fock problem is by expansion of the orbitals as a linear combination of known, usually

non-orthogonal, one-particle functions. This casts the equation as an algebraic generalised eigenvalue problem

$$\mathbf{FC} = \mathbf{SCE}$$

whose size is given by the basis set size. **F** is the matrix representative of the Fock operator in the basis and **S** is the basis set overlap matrix. **C** is the eigenvector matrix, whose columns collect the combination coefficients expressing the molecular orbitals. Since **F** depends on **C** via the potential, the eigenvalue equation is solved again and again until self-consistency is achieved.

It is to be noted that the Fock matrix **F** depends in turn on the matrix representation of the Coulomb and exchange operators over the basis set. These are built by reduction over two induces of 4-index matrices of Coulomb and exchange six-dimensional integrals which must be evaluated by numerical quadrature. Computing these integrals, storing them, and building the Fock matrix, is thus an $M^4$ problem, where $M$ is the basis set size. By contrast, the matrix diagonalisation itself is an $M^3$ problem. In typical production applications, $M$ exceeds the number of electrons in the systems by an order of magnitude, rapidly making even the simple Hartree-Fock problem quite difficult to handle.

### 6.3.3 Electron correlation

As we discussed in the previous section the Hartree-Fock model lacks a proper description of inter-electronic correlation: each electron is described by a separate one-particle wave function and its interaction with the other electrons is seen as an external potential. In order to overcome such limitations it is essentially necessary, either explicitly or implicitly, to express the total wave function not as a single determinant but as a superposition of many determinants. Often the underlying one-particle orbital basis is chosen as the Hartree-Fock orbital basis, but it can also be re-computed, again self-consistently, for a given form of the wave function.

The computational approaches to the electron correlation problem are wide-ranging. The conceptually simplest one is the so-called Configuration Interaction method. Using a fixed MO basis set (e.g. the Hartree-Fock basis), a large number of determinants, in principle exponential in the number of electrons, can be constructed. The Hamiltonian matrix is built (although never stored explicitly) over this set of determinants and a number of desired eigenvalues and eigenvectors are computed by subspace iteration techniques. The Hamiltonian matrix is sparse, because a matrix elements is zero unless the two determinants differ by less than three orbitals, but the size of the determinant basis set is typically very large and, as we said before, explosively growing with the size of the system. In addition, its matrix elements are a function of the 4-index Coulomb integrals over the molecular orbitals, which must be computed starting from the corresponding integrals over the underlying one-particle basis. This basis transformation is in itself an $M^5$ problem, although it consists basically of a sequence of double matrix multiply operations (and a double $M^4$ sorting problem) which can

usually be performed rather efficiently on modern computers. Configuration lists can exceed nowadays $10^6$ in size.

The computational problems in other methods, either of the Coupled Cluster type or perturbation theory-based like Moller-Plesset perturbation theory are largely similar. One additional complication is found in Multi-Configuration Self-Consistent-Field theories (MCSCF), where, in addition to the linear combination problem described above, also the molecular orbitals are optimised for a given wave function expansion, by coupling a non-linear Fock-like problem to the linear configuration mixing problem. In this way, the basis transformation calculation must also be repeated several times. Of course, substantially smaller configuration lists can be handled in MCSCF than in CI.

### 6.3.4 Density Functional Theory

This method, which has reached maturity much more recently than the conventional methods described above, is today perhaps the most widely used and intensely studied one. The Density Functional Theory (DFT) is essentially based on a theorem stating that the total energy of the ground state of a molecular system must be a (unknown) functional of the one-electron density $\rho(\mathbf{r})$. Thus in DFT, the search for the wave function, a complex function of all the position variables of the system and of spin, is replaced by the search for the density, which, independent of system size, is a function of three variables and spin.

Most modern practical implementations of DFT for electronic structure calculations, make use the Kohn-Sham equations. These are based on the observation that any realistic electron density can be written exactly as a sum of fictitious independent particle densities (square orbitals), leading to equations for these orbitals similar to the Hartree-Fock equations but with the added advantage that no non-local part of the potential is present. Thus, in principle, the Kohn-Sham theory affords an exact solution of the many-electron problem at the price of an independent particle model. Of course, the true functional form of the potential in terms of the density is not known, but effective approximations have been devised over the years starting from simple exactly solvable models.

The resolution of the Kohn-Sham equations is usually less computationally expensive than that of the corresponding Hartree-Fock analogues; with respect to the $M^4$ scaling of Hartree-Fock, DFT methods offer a more favourable scaling of $M^3$, typical of diagonalisation problems and allow the study of larger systems at a high degree of accuracy. Moreover iterative diagonalisation techniques and alternative minimisation schemes which avoid straightforward diagonalisations can be devised, further reducing the power of M.

## 6.4 Impact of HPC development

Based on the hypotheses shown in section 5.2 we briefly review the basic techniques and algorithms of relevance in Molecular Sciences trying

to predict the impact that the evolution of computer hardware and architecture will have on computer simulations of chemical phenomena.

The considerations reported in this study are necessarily based on a projection of a known past into a predictable future. It is reasonable to assume that, as already happened in the past, the improvement of the present architectures will imply the development of software libraries (linear algebra, communications, programming languages, operating systems, tools) able to provide the required means to the applications, in order to fully use the resources derived from the new hardware. Moreover, we can reasonably suppose that the improvement of the hardware will imply the study of more complex applications. The development of new memory hierarchies, more articulated computing and communication hierarchies, has implied and will imply the use of more complex paradigms (from BLAS to PBLAS, from LAPACK to SCALAPACK, etc.). Predictions on completely new computational techniques are out of the scope of the present study, both from a technical and an application point of view. All the present technical predictions are based on the idea that the available architectures in the next 5/10 years will be based on the model of the deterministic Touring machine. Very recent developments on the application of fundamental principles to computer science may suggest that new computational models and paradigms will be able in the future to afford problems that in theory cannot be solved by deterministic machines.

Current computational chemistry techniques allow the simulation of systems of variable size, ranging from a few tens to millions of atoms. The parameters which rule the reliability of the simulation reside on the accuracy in the definition of the inter-atomic potential and on the dimensions of the investigated system, since a higher accuracy usually corresponds to an increased computational overload which in turn limits the dimension of the system under study.

In particular, three main approaches to the definition of the inter-atomic potential can be outlined, each corresponding to a different degree of computational complexity and accuracy:

i) Model potentials based on Force-Field (FF) parameterisations[44, 40].

ii) DFT (Density Functional Theory) derived potentials[39, 31, 19].

iii) *ab initio* potentials, directly derived from the solution of the Schrödinger equation of the investigated system[35, 36, 42, 48, 33].

i) The interaction potential is usually expressed as the sum of *a-priori* parameterised van der Waals and Coulombic contributions, the latter showing a quadratic scaling, $O(M^2)$, with the dimensions of the system (here generally indicated with $M$), due to the double sum on the atomic effective charges. The maximum accuracy of this class of methods is typically of the order of 20 kcal/mol, usually too limited for the description of phenomena of chemical interest. However, FF-based methods, implemented according to the fast-multipole (FM) expansion

of the Coulomb potential, have been applied to the approximate description of systems containing up to a few million of atoms and have found a wide success in the investigation of biological systems, surface science and material science (e.g. protein science, material fractures, liquid crystals). FF-based algorithms usually offer excellent scaling performances on massively parallel architectures; indeed, FM expansion of the Coulomb potential can be implemented in such a way as to reach a linear scaling with the dimensions of the system [22]. The aforementioned class of problems will certainly benefit from the hardware and architecture developments outlined in the previous section, allowing the simulation of larger and larger systems in a reasonable time due to the inherent parallelism of FF-based algorithms, and the predicted diffusion of COTS hardware will not reduce the performances of the actual codes.

ii) DFT methods [39, 31, 19] allow to derive the interaction potential from first principles, and lead in principle to the exact solution known the exact exchange-correlation (XC) potential. They usually offer an $O(M^3)$ scaling with the dimensions of the system, typical of direct diagonalisation techniques. More favourable scaling, of the order of $M^2 log M$, can be achieved by using a plane wave (PW) expansion as a basis set and pseudo-potentials (PP) for the description of core electrons [28]. The typical accuracy of these methods, of the order of 3-7 kcal/mol, makes them suitable to the study of chemical interesting problems, and have been successfully applied to the investigation of chemical reactivity and complex material in systems composed by up to a few hundreds of atoms. Moreover, recent developments [24], have allowed to couple the evaluation of a DFT potential to a classical molecular dynamics (MD) scheme, introducing time as a further degree of freedom to explore. The basic algorithmic features of DFT-based MD methods reside in the use of efficient fast Fourier transform (FFT) techniques to compute the different contributions to the total energy (kinetic energy, Coulomb, XC, PP) and its derivatives, the latter task being particularly computationally intensive. The large number of PWs, typically of the order of $10^5$-$10^6$, necessarily translates in large memory requirements; moreover, the parallel implementation of this class of algorithms requires an extremely efficient communication network, due to the particular implementation of the parallel FFT which requires global data exchange. To give a measure of the memory and CPU requirements, let us consider that a system composed by 350 atoms can require up to 24 Gbyte of memory and, to be executed in a reasonable time, 32 IBM power 3 processors. This class of methods will therefore benefit from both increased computer power, communication and memory bandwidth, even if it will be probably limited to run on proprietary hardware, due to the reduced performances of COTS communication devices. Moreover, an adequate development of scientific libraries (FFT and linear algebra) is needed to retain high performance for this class of algorithms.

iii) *ab initio* methods [35, 36, 42, 48, 33] represent the higher level of description of the inter-atomic potential and allow, in principle, the exact solution of the Schrödinger equation without the introduction of

any parameters. However, they usually offer a particularly unfavourable scaling with the dimensions of the system, $O(M^8)$, typical of many-body problems, which makes them applicable to systems composed by a limited number of atoms, usually of the order of 10–20. However, they can be extremely accurate, up to 0.5 kcal/mol, and are still successfully applied in the field of atmospheric chemistry and elementary chemical processes, in which high accuracies are needed. From a computational point of view, the most intensive tasks are represented by the analytic or numerical evaluation of 2-electron integrals, and integral transformations from an atomic orbital to a molecular orbital base. Conventional algorithms require the storage on disk of an enormous amount of data, the semi-transformed integrals, of the order of tens of Gbytes, as well as large memory storage, bandwidth and latency. This class of applications can be defined as memory-bound and is indeed bound to the efficiency of the memory access on a single processor; moreover, the extremely involved data connectivity, makes these algorithms of difficult implementation on parallel machines requiring more and more efficient single CPUs. Therefore, we expect this class of applications to be strongly limited by the reduced memory latency which will not probably grow as much as CPU clock and memory bandwidth. Partial relief to this problem will probably come from deeper and deeper memory hierarchies and increased super-scalar CPU technology.

## 6.5 Conclusions of the case study report

At the actual state of theory and algorithms, we can try to project the impact on applications in the field of Molecular Sciences that the three approaches to the definition of the inter-atomic potential outlined in the previous section will have in 5 to 10 years. In particular:

i) Model potentials based on Force-Field (FF) parameterisations will continue to dominate the scene in the field of liquid crystals, ferroelectric nematic materials [21] and protein-folding. Indeed, due to the large dimensions of the systems under study (up to a few million of atoms) and to the very long time-scale of the dynamical phenomena relevant in such fields (up to milliseconds), FF-based methods will still represent the only viable simulation tool up to 10 years.

ii) In 5 (10) years, DFT methods will probably allow the accurate computations of electronic, structural and dynamical reactive properties of systems containing 3000 (10000) atoms. We can therefore predict that DFT methods will substitute FF parameterisations in chemiometric applications, in which a large number of medium-size calculations is needed. This will have a direct impact in pharmacology; indeed, the design of a new drug usually requires a pre-selection operated by computer simulations and data analysis; the advantage of a much higher accuracy in the description of the investigated molecular systems and properties directly translates into a high selectivity of the target system with

a significant reduction of the number of laboratory tests, up to a factor of 10.

DFT methods will also allow the accurate simulation of small protein systems, or of realistic portions of them, with particular impact on the comprehension of the action mechanism of metalloenzymes, where a reduced model usually neglects the fundamental underlying interactions. To understand the importance of such a field, it is sufficient to mention that both respiration and photosynthesis involve metallo-organic active centres constituted by several thousand atoms; comprehension of the action mechanism of such systems will allow to device efficient synthetic bio-mimetic analogues of the natural systems, with a high impact in the field of energy storage and molecular sensors.

Moreover, we can predict that DFT-based methods will allow the accurate simulation of nano-scale systems with a high impact in the design of molecular engines, quantum computation devices and chemical storage of data [18].

iii) *Ab initio* potentials will reach such a high standard accuracy (ca. 0.2 kcal/mol) so as to allow the realistic simulation of elementary reactions of relevance in the field of atmospheric chemistry ( e.g. ozone depletion) [37] or combustion ( e.g. nitrogen oxides chemistry) [46], integrating and partially replacing existing experimental techniques. *Ab initio* simulations will allow the accurate computation of reactive cross sections and rate constants of elementary systems composed up to 10–50 atoms, under extreme conditions (high temperature and pressure) on a quantitative basis; moreover, the study of elementary reactions in the interstellar space [16, 45, 34], e.g. the synthesis of organic compounds from small molecules and atoms (C, N, O, H) which is not usually directly accessible to the experimentalists, will be of great help in the design of new space-aircrafts materials and ultimately in the comprehension of the origin of life.

However, according to the previous analysis, the computational resources required for such demanding applications can be achieved exclusively by large-scale computing facilities. In this view, the 12 MEuro platform option discussed above might represent only a starting point for the creation of a large transnational European super-computing resource.

# Appendixes

# Appendix A

# Interview specific questions

## A.1 Conventions

The following sections (A.2–A.7) presents the questions asked during the interviews. The questions are divided in sections [(Grid) Awareness, User Profile, Application profile, Infrastructure, Security & Service, and Future needs] following the ENACTS user questionnaire[25], as explained in section 2.

The following conventions apply:

C1. Comments (and some questions) directly related to the questionnaire results are denoted by "C".

Q1. Further "specific questions" are denoted by "Q".

## A.2 (Grid) Awareness

C1. Why low experience of Grid enabling technologies?

C2. Could you describe some (in your opinion) likely future scenarios of Grid based solutions related to HPC?

Q1. How large is/will be the impact of Grid on computer architectures?

## A.3 User Profile

Q1. New ways of doing research: what areas of research will benefit the most from future systems?

Q2. How will sharing, visualisation and dissemination of data be performed in the future?

## A.4 Application profile

C1. 4.1: Why is the usage low for OpenMP?

Q1. Parallel programming models:

   A. MPI is 8 years, OpenMP is 3 years, but both are based on similar proprietary ideas from 15 years ago. Is this what we will have 10 years from now?

   B. What about Co-Array-Fortran/UPC, SHMEM, HPF?

   C. Does new hardware imply new programming models?

Q2. Which programming languages will be used in the future?

Q3. Do you foresee any changes in the way benchmarking of new systems will be done?

Q4. Do $3^{rd}$ party codes influence the system architectures? (See also results in section 7.)

Q5. Do you see any change in how much $3^{rd}$ party codes are used, as compared to codes developed in-house?

## A.5 Infrastructure

Q1. Impact from other (commercial) markets: PC, data base engines, Web servers, e-business, streaming media. These are high volume markets and will have an impact on COTS architecture. How does this influence the performance for HPC applications?

Q2. Will there be an increased or decreased number of architectures for HPC? Will there be one dominant architecture?

Q3. Beowulf clusters: friend or foe?

Q4. Operating systems:

   A. The future of Linux?

   B. Proprietary Unix: will it survive?

Q5. Role of the vendor:

   A. Becoming a system integrator?

   B. How to work with the Open source community?

## A.6 Security & Service

Q1. Pain/gain ratio for users to use HPC systems and get performance has not improved (some would say has gotten worse). Will this change in the future?

Q2. Does more powerful imply less reliable systems?

## A.7 Future needs

C1. Future expectations and requirements reflect what is already used today(?) To what extent should the research and development by the vendors be influenced by user expectations and "wishes"?

C2. 7.6: Why are not highly parallel systems considered more important?

Q1. Future computer systems:

A. Summary description
   i. Programming model(s).
   ii. Building blocks (COTS vs proprietary) and interconnects.
   iii. Memory and processor distribution.
   iv. Scaling to very large systems:
      - How large systems can we expect in terms of number of processors?
      - What is setting the upper limits for very large systems?
   v. Sustained system performance compared to peak (and compared to todays ratio)?
   vi. Price/performance trend for future systems.
   vii. How to get the best performance?

B. Processors:
   i. What type of technology?
   ii. Mass-produced general purpose versus low volume special purpose.
   iii. Transistor "real estate" growth: what will the growth development look like and what to do with all the transistors?
   iv. Performance.

C. Memory:
   i. What type of technology?
   ii. Shared? Distributed?
   iii. Memory hierarchies. Performance.
   iv. Application memory size.
   v. What will/should be done to decrease the imbalance of processor performance vs. memory bandwidth and latency?

D. Networks (Scaling):
   i. Bandwidth, latency.
   ii. Will there be any differences in the future development of internal versus external networks?

E. Storage:
   i. Storage Attached Networks (SAN)? Network Attached Storage (NAS)? Hierarchical Storage Management (HSM)?

ii. Disk versus tapes for long term storage.

iii. Performance:

- short term storage (during a calculation)
- long term storage

F. Visualisation.

G. Moore's law: For what components and for how long?

Q2. Future computer centre, what will it look like?

# Appendix B

# List of acronyms

**ACM** Association for Computing Machinery

**AI** Artificial Intelligence

**API** Application Program Interface

**ASP** Application Service Provider

**ATM** Asynchronous Transfer Mode

**B2B** Business-to-business

**BLAS** Basic Linear Algebra Subprograms

**CAVE** CAVE Automatic Virtual Environment

**CI** Configuration Interaction

**CMOS** Complementary Metal Oxide Semiconductor

**COTS** Commercial Off-The-Shelf

**CP** Car-Parrinello

**CPU** Central Processing Unit

**CRM** Customer Relationship Management

**CSCISM** Center for High Performance Computing in Molecular Sciences

**DAS** Direct Attached Storage

**DFT** Density Functional Theory

**DRAM** Dynamic Random Access Memory

**DRMAA** Distributed Resource Management Application API

**DSP** Digital Signal Processing

**DTF** Distributed Terascale Facility (IBM HPC installation in the US)

**DWD** Deutche Wetterdienst

**EJB** Enterprise JavaBeans

**ENACTS** European Network for Advanced Computing Technology for Science

**EPCC** Edinburgh Parallel Computing Centre

**ERP** Enterprise Resource Planning

**FeRAM** Ferrorelectric RAM

**FF** Force Field

**FFT** Fast Fourier Transform

**FM** Fast Multipole

**FP-CMOS** Flexible Parameter CMOS

**FPGA** Field Programmable Gate Array

**GGF** Global Grid Forum

**GSN** Gigabyte System Network

**GT2** Globus Toolkit 2

**GUPS** Giga Updates Per Second

**HPC** High Performance Computing

**HPF** High Performance Fortran

**HSM** Hierarchical Storage Management

**IA-64** Intel Architecture 64 bit

**IDC** International Data Corporation

**IEEE** Institute of Electrical and Electronics Engineers

**IP** Internet Protocol

**ISV** Independent Software Vendor

**ITRS** International Technology Road-map for Semiconductors

**J2EE** Java2 Enterprise Edition

**JSP** Java Server Pages

**JXTA** JuXTApose

**LAN** Local Area Network

**LAPACK** Linear Algebra PACKage

**LTC** Linux Technology Center (at IBM)

**MCSCF** Multi-Configuration Self-Consistent Field

**MD** Molecular Dynamics

**MPI** Message Passing Interface

**MPP** Multiple Parallel Processing

**MRAM** Magnetoresistive RAM

**MTA** Multi Threaded Architecture (Cray HPC system)

**NAS** Network Attached Storage

**NAVi** Network Animated View

**NSA** National Security Agency

**NSC** National Supercomputer Centre

**NUMA** Non-Uniform Memory Access

**OGSA** Open Grid Service Architecture

**OpenMP** Open Multi Processing

**P2P** Peer-to-peer

**PBLAS** Parallel Basic Linear Algebra Subprograms

**PC** Personal Computer

**PP** Pseudo Potential

**PW** Plane Wave

**QOS** Quality Of Service

**RAM** Random Access Memory

**RFP** Request For Proposal

**RTE** Run Time Environment

**SAN** Storage Area Network

**SCSI** Small Computer System Interface

**SCSL** Source Code Software Licensing

**SHMEM** Shared Memory (access library)

**SIA** Semiconductor Industry Association

**SMP** Symmetric Multi Processing

**SPEC** Standard Performance Evaluation Corporation

**SPP** Special Purpose Processor

**Sun ONE** Sun Open Net Environment

**TCO** Total Cost of Ownership

**TCP** Transmission Control Protocol

**TSMC** Taiwan Semiconductor Manufacturing Company

**UMA** Uniform Memory Access

**UPC** Unified Parallel C

**W3C** World Wide Web Consortium

**WAN** Wide Area Network

**WDM** Wavelength Division Multiplexing

**XC** Exchange Correlation

**ZPL** Z (level) Programming Language

# Bibliography


[1] The accelerated strategic computing initiative (asci) web site. http://www.llnl.gov/asci.

[2] The cactus code web site. http://www.cactuscode.org.

[3] Database access and integration (DAI). http://umbriel.dcs.gla.ac.uk/nesc/general/esi/events/dai.html.

[4] DutchGrid: Large-scale distributed computing in the netherlands. http://www.dutchgrid.nl/.

[5] Earth simulator research and development center. http://www.es.jamstec.go.jp/eng/menu.html.

[6] The eLiza project web site. http://www-1.ibm.com/servers/eserver/introducing/eliza/.

[7] The ENACTS project. http://www.enacts.org.

[8] The eurogrid web site. application testbed for european GRID computing. http://www.eurogrid.org.

[9] Globus project homepage: Towards globus toolkit 3.0: Open grid services architecture. http://www.globus.org/ogsa/.

[10] IBM to build world's most powerful computing grid. IBM News: See http://www.ibm.com/news/us/2001/08/092.html.

[11] The new productivity initiative web site. http://www.newproductivity.org.

[12] Sun microsystems: Project JXTA. http://www.sun.com/p2p/.

[13] Sun open net environment (Sun ONE). http://www.sun.com/software/sunone/overview/.

[14] Uk database task force (DBTF) and GGF databases and the grid (DAG) working group meeting. http://umbriel.dcs.gla.ac.uk/nesc/general/esi/events/dbtf.html.

[15] The unicore web site. http://www.unicore.de.

[16] L. J. Allamandola, A. G. G. M. Tielens, and J. R. Barker. Interstellar polycyclic aromatic hydrocarbons: the infrared emission bands, the excitation-emission mechanism and the astrophysical implications. *Astrophys. J. Supp.*, 71:733, 1989.

[17] G. Allen, W. Benger, T. Goodale, H. Hege, G. Lanfermann, A. Merzky, T. Radke, E. Seidel, and J. Shalf. The cactus code: A problem solving environment for the grid. In *Proceedings of 9th IEEE Intern. Symposium on High Performance Distributed Computing (HPDC9)*, page 25, Los Alamos/CA, 1999. IEEE, IEEE Computer Society Press.

[18] C. W. Bauschlicher, A. Ricca, and R. Merkle. Chemical storage od data. *Nanotechnology*, 8:1, 1997.

[19] C. W. Bauschlicher, A. Ricca, H. Partridge, and S. R. Langhoff. Chemistry by density functional theory. In D. P. Chong, editor, *Recent Advances in Density Functional Methods, Part II*. World Scientific Publishing Co., Singapore, 1997.

[20] A. Beck. HPCwire: Building a better grid: an interview with wolfgang gentzsh. http://www.tgc.com/hpcwire.html, article 102306, March 2002.

[21] R. Berardi, M. Ricci, and C. Zannoni. Ferroelectric nematic and smectic liquid crystals from tapered molecules. *ChemPhysChem*, 2:443, 2001.

[22] J. A. Board, J. W. Causey, Jr. Leathrum, J. F., A. Windemuth, and K. Schulten. Accelerated molecular dynamics simulation with parallel fast multipole algorithm. *Chem. Phys. Letters*, 89:198, 1992.

[23] J. Burton. DSstar: Peer-to-peer grows up and gets a real job. http://www.tgc.com/dsstar/, article 103165, June 2001.

[24] R. Car and M. Parrinello. Unified approach for molecular dynamics and density functional theory. *Phys. Rev. Letters*, 55:2471, 1985.

[25] J.-C. Desplat, Judy Hardy, Mario Antonioletti, Jarek Nabrzyski, Maciej Stroinski, and Norbert Meyer. Grid service requirements. Technical report, EPCC and PSNC, Edinburgh, January 2002. http://www.enacts.org.

[26] Jack J. Dongarra and David W. Walker. The quest for petascale computing. *Computing in Science & Engineering*, 3(3):32–39, May/June 2001.

[27] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The physiology of the grid: An open grid services architecture for distributed systems integration. http://www.globus.org/research/papers/ogsa.pdf.

[28] G. Galli and A. Pasquarello. First-principles molecular dynamics. In M. P. Allen and D. J. Tildesley, editors, *Computer Simulation in Chemical Physics*. Kluwer Academic Publishers, The Netherlands, 1993.

[29] J. Jackson. IBM sees future in grid computing. *Washington Technology*, 16(12), September 2001. See http://www.washingtontechnology.com/news/16_12/business/-17133-1.html.

[30] Earl Joseph, Christopher G. Willard, Michael Swenson, and Debra Goldfarb. A new HPC technical computing benchmark: "the IDC balanced rating". *IDC Bulletin #W24824*, June 2001.

[31] W. Kohn, A. D. Becke, and R. G. Parr. Density functional theory of electronic structure. *J. Phys. Chem.*, 100:12974, 1996.

[32] Raymond Kurzweil. The singularity is near. http://www.-kurzweilai.net/articles/art0134.html?printable=1, 2001.

[33] Ed. Langhoff, S. R. *Quantum Mechanical Electronic Structure with Chemical Accuracy*. Kluwer, Dordrecht, 1994.

[34] A. Leger and J. L. Puget. Identification of the 'unidentified' ir emission features of interstellar dust? *Astron. Astrophys.*, 137:L5, 1984.

[35] J. P. Lowe. *Quantum Chemistry*. Academic, Boston, 1993.

[36] D. A. McQuarrie. *Quantum Chemistry*. University Science Books, Mill Valley, CA, 1983.

[37] M. J. Molina and F. S. Rowland. Stratospheric sink for chlorofluoromethanes-chlorine atom catalyzed desctruction of ozone. *Nature*, 249:810, 1974.

[38] Dan Neel and Ed Scannell. InfoWorld: Grid project could net web services tools. http://www.computerworld.com/storyba/-0,4125,NAV47_STO67960,00.html, February 2002.

[39] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford, New York, 1989.

[40] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114:10024, 1992.

[41] C. Rogala. HPCwire: Bringing the grid to industry. http://www.tgc.com/hpcwire.html, article 102195, March 2002.

[42] III Schaefer, H. F. *Quantum Chemistry: The Development of Ab Initio Methods in Molecular Electronic Structure Theory*. Clarendon Press, Oxford, 1984.

[43] S. Shankland. CNET news.com: Distributed computing gets a corporate twist. http://news.com.com/2100-1001-270988.html?legacy=cnet, August 2001.

[44] U. C. Singh and P.A. Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $ch_4cl$ + cl exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7:718, 1986.

[45] J. Szczepanski and M. Vala. Laboratory evidence for ionized polycyclic aromatic hydrocarbons in the interstellar medium. *Nature*, 363:699, 1993.

[46] P. Weilmuster, A. Keller, and K. H. Homann. Large molecules, radicals, ions and small soot particles in fuel-rich hydrocarbon flames. part i: Positive ions of polycyclic aromatichydrocarbons (pah) in low-pressure premixed flames of acetylene and oxygen. *Combust. Flame*, 116:62, 1999.

[47] T. Weiss. InfoWorld: Grid computing to aid breast cancer treatment, research. http://www.computerworld.com/storyba/-0,4125,NAV47_STO66115,00.html, November 2001.

[48] Ed. Yarkony, D. R. *Modern Electronic Structure Theory. Parts I and II.* World Scientific Publishing Co., Singapore, 1995.