EL/UK/RAL

European Research Consortium for Informatics and Mathematics

High Performance Computing and Networking: an ERCIM View

Monday, February 15, 1993

EXECUTIVE SUMMARY

The following paper produced by the ERCIM partners supports the main areas of activities proposed in the High Performance Computing and Networking Report but proposes a different strategy for achieving the goals.

ERCIM supports

ERCIM supports the HPCN report's main conclusions:

(1) The strategic importance of parallel computing

(2) The strength of European research in parallel computing

(3) The poor record Europe has of turning research results into products

(4) The need for a high speed networking infrastructure

ERCIM concerns

ERCIM is concerned that:

(1) The linking of High Performance Computing with High Performance Networking is inappropriate as HPC is not the only demanding challenge for networking. Any HPN initiative should be based on a broader base of applications.

(2) The application area breakdown in the HPCN report is inappropriate. An HPC initiative should concentrate solely on those activities of benefit to Europe and its industry that require centralised high performance computing rather than attempt to cover the whole area of parallel processing.

(3) The proposed management structure is essentially similar to existing structures that have been found lacking. A new approach is required.

ERCIM proposes

ERCIM proposes an alternative strategy:

(1) High Performance Networking will be a necessary infrastructure for normal economic life within Europe. It should be established independent of a High Performance Computing initiative and should be available to all new initiatives.

(2) A High Performance Computing initiative should focus on the industrial and environmental applications of simulation requiring centralised high performance computing. Any such initiative should be aware of the need for it to be a tool in the engineering design process.

(3) A separate initiative aimed at the use of Affordable Parallel Processing within a generic environment aimed at an industrial sector of prime importance to Europe should be established.

(3) Notice should be taken of the comments in the ERCIM Strategic Research report concerning the management of such programmes and the assignment of IPR.

1. INTRODUCTION

The Report of the High Performance Computing and Networking Advisory Committee (Volume 1) was published in October 1992 (it is available from the Commission of the European Communities, DG XIII, 200 Rue de la Loi, B-1049, Brussels, Belgium). The Report title will be abbreviated to HPCN for the remainder of this document. The HPCN Report is addressed to all those whose future is affected by High Performance Computing and Networking in the European Community. It proposes a large and focused programme aimed at establishing a European industrial and scientific lead in the combined area of High Performance Computing and Networking.

Also, in October 1992, the European Research Consortium for Informatics and Mathematics (ERCIM) published a Report entitled *Strategic Research: A major Focus for the Fourth Framework Programme* (it is available from ERCIM Central Office, Domaine de Voluceau, Rocquencourt, BP 105, F-78153 Le Chesnay, France). The Report title will be abbreviated to ESR (ERCIM Strategic Research) for the remainder of this document. The ESR Report emphasised the need for a major focus on strategic research and proposed a set of key areas that should be supported. It also commented on the management structure appropriate to such a programme.

While the two reports identify similar areas of key importance, the shapes of the proposed programmes differ as does the proposed management structure. This report identifies the differences between the two approaches and concentrates on those parts of the ESR report covering the same ground as the HPCN report. It indicates why ERCIM believes its approach is more valid.

2. HIGH PERFORMANCE COMPUTING AND NETWORKING

2.1 Major Points

The HPCN report makes a number of key points which ERCIM strongly supports:

(1) The emergence of parallel computing systems offers the potential for massive increases in performance which are essential to meet the

increasing demands of the industrial, commercial and academic communities within Europe.

(2) European R&D programmes have led to a world-class understanding of parallel systems and to the industrial development of hardware and software. However, Europe has a poor record in widely disseminating its knowledge and bringing to market its results on an appropriate scale.

(3) High speed networking is crucial to the development of an effective European infrastructure.

The ESR report states that, by the year 2000, most systems will be based on scalable affordable parallel systems increasing in size and power at the rate of a factor of 2 per year at the same cost. It points out that no architecture will be the best in all areas and that it is essential, therefore, that the problems arising from writing software that is scalable, and will run on a range of hardware, be solved.

High speed communications and the need to distribute computer power to where it is required will require systems that are appropriate to such a distributed, heterogeneous environment. A mixture of communications, media and distribution strategies must be supported.

The ESR report supports the conclusion that the European industry is relatively strong in the area of parallel architectures and should be supported.

A large part of the ESR report concentrates on the problem that many research projects in Europe fail to produce commercial products even when there are clear possibilities. It makes a number of proposals that would improve the current framework for research within Europe.

2.2 Major Concerns

The major concerns that ERCIM has with the HPCN report are:

(1) The linking of High Performance Computing and Networking.

(2) The three broad classes chosen to break down applications.

(3) The management structure proposed for the programme.

The ESR report makes the strong statement that:

it is misguided to relate the need for high performance communication with the requirement for high performance computing at a set of selected sites with a dispersed user base. The challenges of high speed communication should be aimed at areas where fast communication of

data, graphics and images is essential to the business on an almost real time basis. This covers a wide range of challenges which have a greater significance to European industry as a whole. A strategic research focus based on ***affordable parallel systems used in a distributed environment*** *presents greater challenges and a wider opportunity for success.*

The HPCN report proposes a breakdown of applications relevant to the programme into three broad classes:

(1) Simulation and design in engineering and science

(2) Information management

(3) Embedded systems applications

While recognising the need to break a large programme into manageable parts, this particular breakdown seems particularly inappropriate. Design is intimately concerned with information and its reuse. A separation of information management from design suggests a lack of understanding of the design process and will inevitably lead to an environment that is inappropriate for manufacturing industry in Europe and will fail to recognise the need to integrate information management and design into a single environment. The third focus on embedded systems is one that ERCIM would support but is clearly inappropriate for a programme that sees its main emphasis on linking **High Performance** computing and **Networking**. This particular application area has neither an interest in high speed networking as described in the HPCN report nor does it have a requirement in most cases for the top-end high performance computing. This strand is much closer to the **affordable parallel processing** thrust of the ESR report.

The following sections will address these concerns and propose alternative strategies which ERCIM believe would achieve the laudable aims of the HPCN proposal but in a wider context and with better foci.

3. HIGH PERFORMANCE NETWORKING

ERCIM strongly supports the HPCN proposal for a programme aimed at building a Gigabit/sec network across Europe. ERCIM has sympathy with the strategy which proposes experimental high speed (Gigabit/sec) test beds between lead sites. At the same time, an interim network infrastructure should be established between major hubs across Europe with at least 34 Mbits/sec between them initially.

ERCIM's main concern is that the need for such a network infrastructure is independent of a High Performance Computing initiative. Clearly, some HPC activities would benefit from such a network infrastructure (for example, remote working if that is considered desirable). However, some applications of HPC do not require it and many applications that are not

part of the HPC programme both need it and provide at least as demanding uses. (There is some evidence in the HPCN report itself that applications not intimately connected to HPC have been included just to make the case of joining HPC and HPN more plausible.)

Today's networking has evolved largely independent of its application. The challenges of distributed multimedia cooperative working, data retrieval, security, remote control, etc all require a coming together of networking, domain expertise, data transfer protocols, guaranteed performance and cached storage within the network. It will be necessary to explore new distributed multimedia broadband applications together with end users in a number of different application areas (for example, the distributed university, the distributed office, the distributed design environment, the distributed control centre, etc).

Multimedia user-interface paradigms will be significantly different from single media ones and they will have profound effects on both protocols and networking topology. For example, there will be a need for higher layer protocols for multimedia applications (for example, packet-video/audio, support for distributed applications), internetworking between wired and wireless networks, and lower level protocols for high speed (ATM, DTM etc) and wireless (IR, PR, cellular) networks. For example, SICS has a Stockholm-wide dark fibre network being used as an ATM testbed and a large FDDI-ring. Testing of a DTM network is about to start and there are experiments with wavelength multiplexing and optical switching. Similar experiments are in process at other ERCIM partners.

Any European network infrastructure will be required to perform appropriately in a range of application areas. Consequently, if one area is to be chosen as the basis for a demonstrator, it must be sufficiently broad to encounter the most demanding problems. By linking the implementation of the networking infrastructure to a highly specific focus such as High Performance Computing as perceived in the report, there is the real danger that the networking infrastructure will only satisfy the limited requirements of that activity. High Performance Computing is neither the most demanding application nor the richest in terms of functionality. Its networking demands are often just the need for a large hose of unstructured data.

ERCIM believes the right approach is to define a topology for the long-term network and an implementation plan that takes into account the various requirements of different application areas. A High Performance Computing initiative should be one user of such a network but not the only one. There would be a benefit in having other initiatives such as a generic application of fundamental importance to Europe and which has greater scope within industry. The pilot network should link sites with the broadest range of existing applications.

4. HIGH PERFORMANCE COMPUTING

The HPCN proposal is, to a great extent, oriented towards large scale simulation, and is dominated by the needs of computational scientists doing complex simulations of processes and phenomena. Such simulations of physical, chemical and biological processes and phenomena tend to be extremely challenging with respect to both computing power and local data transfer. It is clear that research in this area is important in that the present peak performance will be affordable in some year's time. However, the peak performance market is limited in size and already occupied by other players. The challenge is to balance Europe's contribution such that it feeds into the requirements of affordable computing at the correct time.

The proposed breakdown of HPC applications is:

(1) Simulation and design in engineering and science

(2) Information management

(3) Embedded systems applications

The HPCN report proposes this breakdown on the basis that:

(1) Simulation and design will be concerned with exploring the largest and most challenging problems currently envisaged. Examples given are environmental modelling where individual simulations may take several months on a High Performance Computer, Computational Fluid Dynamics simulations of large man-made or environmental constructs etc. The implication is that there are major challenges for Europe concerned with large simulations using a small number of dedicated centres with remote access to view results and steer future batch calculations.

(2) Information management will be primarily concerned with intelligent searching of large databases such as patent information, legal cases, and medical records. In the long term, the searching of data involving voice, graphics and images will provide additional challenges.

(3) Embedded systems will have fixed programs with the ability to tune the hardware and software to achieve high performance and thus a competitive edge. The main example quoted is automatic address recognition in a postal system. Others mentioned are HDTV, image processing, virtual reality, compression, robot control, etc.

The major thrust of any new initiatives should be at exploring a set of avenues where European industry can obtain a competitive edge by the timely use of parallel processing techniques. The rapid increase in

performance of systems over the next decade indicates that a range of new applications will become tractable each year and it will be possible to demonstrate the cost effectiveness of these systems commercially. For example, while it is not possible to do interactive design and simulation of complex engineering artefacts today on affordable parallel systems, within the next decade some areas will become tractable and such areas are where SMEs can achieve market penetration. Examples of such areas are building design, electromagnetic design, fluid dynamics associated with simple parts, etc. On the other hand, large scale simulations of complex fluid dynamics systems will remain batch calculations for the foreseeable future. In consequence, ERCIM would prefer a separate focus on Affordable Parallel Processing rather than putting the focus only on those problems which are grand challenges now and for the foreseeable future.

Similarly, in the area of information management, specialised parallel processing hardware has already been cost effective in parallel searching of large numbers of similar records (Telephone Directory Enquiry, for example). Making such facilities available to a design engineer could shorten the design cycle, make products with wider applicability, reduce the costs and allow products to be sold in different markets with different constraints. Information management is rarely an end in itself. It is part of a much larger activity. Separating it from its application will be counter productive. The industrial applications of simulation for design purposes should not be considered as something different or decoupled from information systems. When applied to the design process, simulation has to be an integral part of the total knowledge reuse process.

The third strand, embedded systems sits uneasily with an initiative focusing on centralised high performance computing. Such systems are likely to produce their competitive edge through richness of functionality, adaptability, ability to handle a changing environment, usability, etc. A market edge based on tuning hardware or software for minimal cost may not be the only one feasible. While recognising the importance of embedded systems and the challenges to be faced, these are much better addressed in a separate programme

There is a danger in the HPCN report in that it appears to believe that Europe can compete effectively with Japan and the USA on *all* aspects of HPC. This is a naive view. Europe must focus on a set of activities important to Europe and one based on a significant software component. The HPCN report also implies that the rate of development of HPC systems has created a unique window of opportunity for Europe. That same window of opportunity is open to Japan and the USA and with their programmes already in place, they will continue to have the edge if the European programme is just a carbon copy of their programmes. *The window of opportunity only exists for Europe if it can target an area that is important but not well served by the other programmes.*

The main advantage of highly parallel systems is that they make processing power available at the most effective point whether it is local or remote. The major bottleneck is the lack of enabling technology to allow users to access such systems easily. Ease of programming, parallel software development methods, parallel operating systems etc are where Europe needs to invest effort. Until these issues are solved, very few applications will run effectively on HPCs and the demand for such systems will remain on the fringe of the market place.

The requirements of the users of High Performance Computing have not been sufficiently analysed in the HPCN report. There is a fundamental difference between academic and industrial users. The academic user is interested in high performance and will adapt and optimise his application program for any new architecture. The quality of the programming environment is not of major importance. On the other hand, the industrial user is committed to using third party application software in many cases. Such software must be available and supported by the developer for any architecture under consideration. The third party application developer needs standards for the programming environment and it must be sophisticated. Also, the end user needs good software for the management of the architecture in a production environment. There must be good linkage between, for example, the application code and the visualisation system.

5. PROPOSED MANAGEMENT STRUCTURE

The HPCN report breaks the work programme into eight key areas distributed across four separate action lines:

(1) Applications
- Existing
- Novel
- Tools

(2) R&D
- Key Technologies
- Technology Demonstrators

(3) Infrastructure
- Networking
- Support Nodes

(4) Education and Training

As has already been made clear, ERCIM does not believe one of the major planks of the proposal that *the HPCN programme should be the driving force behind the trans-European backbone and the high-speed test beds.* In consequence, the case for establishing 28 supercomputer centres over a

five year period as the basis for the European networking infrastructure is not one that ERCIM believes has been made. Similarly, the case for a further 20 centres whose main aim is to provide network access to those facilities is also not supported. ERCIM is in favour of the provision of a set of HPC centres and that these should be major nodes on the European network infrastructure. The major concern is that the HPC part of the programme is being given the lead role in defining a European infrastructure. The use of external centres by major industrial companies will not occur unless the environment provided is sufficiently attractive to provide some of their other needs as well. The proposed infrastructure is not targetted towards the third party application developer who is predominantly an SME.

ERCIM has less concern with regard to the other major areas of the programme. Putting 30% of the HPC budget into application support split between existing applications, new applications and development tools seems the minimal sensible share of the programme. Similarly ensuring the continuance of R&D in this general area of parallel processing is also supported. However, the major problem with producing usable highly parallel systems at the moment is the lack of the basic infrastructure needed to develop software. A stronger emphasis on the basic tools environment would be appropriate.

The management structure proposed for handling this ambitious project is a Programme Management Group (PMG) responsible for selecting projects, coordination, monitoring etc. This together with a Steering Committee responsible for supervising the programme will report to a HPCN Board containing the major participating organisations. The HPCN Board will be responsible for defining the overall objectives of the programme. The PMG will have some staff but most of the management will be subcontracted to specialised companies.

The specific projects will be defined in much the same way as existing ESPRIT Projects. Calls for Proposals in the Application areas will be made for the 36 application projects to be awarded each year together with a further 18 tools projects loosely connected to the application projects.

The ERCIM ESR report has raised a number of concerns with regard to the current method of awarding projects and monitoring that takes place in ESPRIT. It gives a number of reasons for the failure to produce products that are specific to ESPRIT:

(1) ***No Independent Assessment:*** *assessment of research results is carried out by advanced users integrated and paid for within the project. In consequence, they become part of the project with in-depth knowledge and are not an independent evaluation arm. Also the number of organisations that can benefit from and assess the prototype products is limited. Advanced users that are part of the project should not be responsible for independent assessment.*

(2) ***Unavailability of Assessment Results:*** *because the experiments carried out by partners in a consortium are considered to be the property of the consortium, the results are not made available to a wider audience.*

(3) ***Discouragement of Real Innovation:*** *the market-driven focus of ESPRIT has resulted in less innovative projects, and the number of novel experiments has decreased. The tight workplans and the way some projects interpret them can preclude genuine innovative approaches that do not fit. (The fact that the work plan is formulated much earlier inhibits truly innovative approaches.)*

(4) ***Dominance of Large Companies:*** *the earlier sections have indicated the need for flexibility and speed in making strategic and innovative advances. These are often not the characteristics of large organisations. Such companies tend to enter the market once it has been established. In consequence, the dominance of large companies rather than SMEs in the strategic rather than near-market part of the programme builds in a greater chance of failure to exploit.*
In some instances, large companies regard the Commission as a customer (rather than a partner) requiring contract work to be performed. Assuming the funding is appropriate, work will be carried out on the basis that this is what the Commission wants rather what the company needs. There is little intention of marketing the end results. This leads to an unquestioning attitude to the goals of the research programme and an inability to change direction when the situation changes.

(5) ***Proposal Selection:*** *the experts selecting projects largely come from industry. While having a good knowledge of the current market, they are less able to see possible new strategic directions or innovative solutions in existing areas.*

(6) ***Maintaining Balance:*** *the undue emphasis on a balanced programme in terms of the participation of community countries in the programme often puts unreasonable constraints on research projects.*

(7) ***Insufficient Support for Product Development:*** *the extent of the SPRINT and VALUE programmes is insufficient and not integrated with the main research programmes.*

(8) ***Heavy Project Management:*** *the approach adopted by project officer and project manager often results in a large management overhead in current CEC programmes, causing SMEs to shy away from acting as Project Managers. This ensures a significant presence of large manufacturers in the role of project managers. With the points made in (4) above, this can lead to less innovative approaches than might be feasible.*

ERCIM believes many of these objections apply to the HPCN project as it is defined. The ESR report suggests a number of ways in which the current structure could be improved. These are equally relevant to the HPC programme.

6. CONCLUSION

ERCIM welcomes the HPCN report in that it focuses on a number of key issues that are important for Europe to address. However, the programme as stated, is not the most effective way to attack those issues nor does it have the cohesion necessary for a programme of this size. The major points are:

(1) High Performance Networking is vital for normal economic life in Europe in the future and should be tackled on a broad front independent of a High performance Computing initiative.

(2) A High Performance Computing initiative should concentrate on the most demanding industrial and environmental applications of simulation requiring centralised High Performance Computing.

(3) A separate initiative concentrating on the use of Affordable Parallel Processing within a generic environment aimed at an industrial sector of prime importance to Europe should be established.