The Power of

"HIGH PERFORMANCE computing and networking (HPCN) is likely to represent one of the foundations of the technological and economic infrastructures of the Community for the next century." This statement by Professor Carlo Rubbia, the Nobel laureate for physics and chairman of an advisory committee on HPCN, highlights the importance of this area of information technology.

HPCN encompasses many topics - communications, advanced computing technologies, parallel processing. This article looks at just one aspect: the evolution of supercomputing and its replacement by parallel computing. Subsequent articles will deal with other important issues: the application of HPCN in industry and commerce, requirements for networking and communications, and the potential for exploiting embedded systems.

As an example of the relevance and usefulness of HPCN, consider its application to the car industry. Along with economy, safety has become a critical issue in the design of modern cars. Unfortunately, safety and economy result in conflicting requirements: safety necessitates strength, and hence bulk, while fuel-efficiency dictates that the vehicle be as light as possible. The use of advanced materials, such as carbon fibre, is one effective way of reaching an acceptable compromise. However, it is also a very expensive way. An alternative is to use more sophisticated body structures, trimming material and weight without losing structural integrity or strength. Unfortunately, it isn't possible to predict exactly what will be the ultimate effect of even minor changes to the structure, especially in the case of a collision. This is why cars are crash tested.

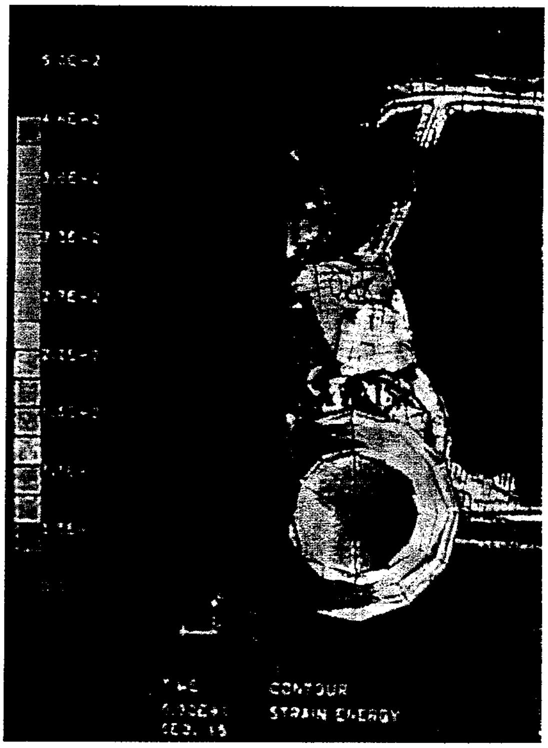
Obviously, this is expensive, and especially so when it is part of an iterative process of design, test, and re-design. Fortunately, it is also possible to simulate the effect of a collision on a newly-designed structure (see illustration on page 6), resulting in a very cost-effective and time-efficient car design procedure. These simulation packages require very high performance computers and limitations in the amount of available computing power constrain the sophistication of the simulation. The advent of the new computing architectures described in this article provides the opportunity to increase this power by orders of magnitude at an affordable cost, facilitating highly accurate high-resolution simulations which could be generated in seconds or minutes rather than hours or days.

Once this becomes an integral part of the manufacturing process, design times and costs will fall significantly along with an accompanying reduction in the so-called time-to-market: a key to competitiveness.

Supercomputers

Traditionally, high performances in computing have been achieved using supercomputers: machines which are architecturally the same (in broad terms at least) as conventional everyday computers, but which are faster and much more expensive. That is to say, supercomputers have a central processing unit (or maybe a very small number dedicated to specific tasks) and a memory subsystem (or, again, maybe a very small number of memory subsystems).

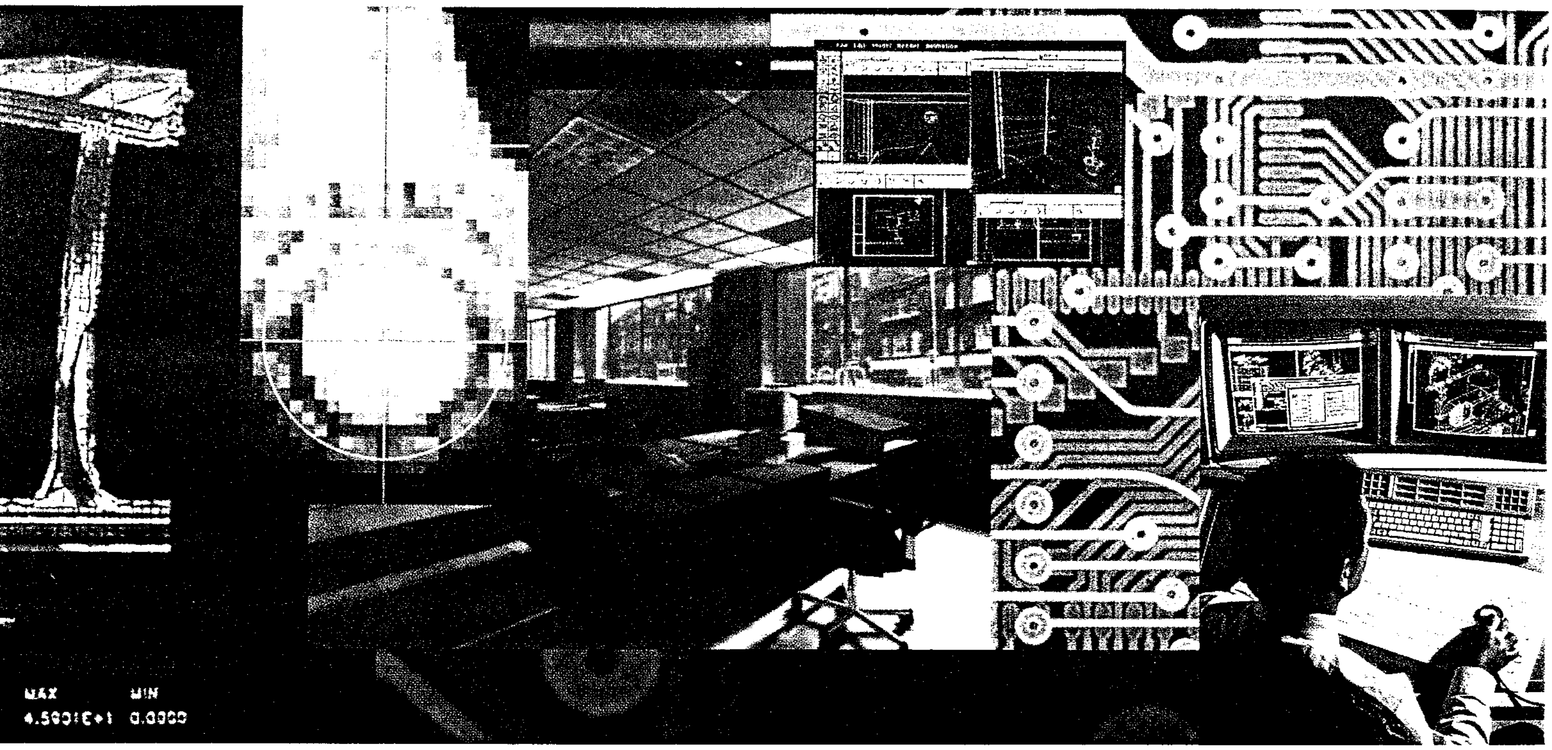
These supercomputers, like their conventional cousins, are effectively sequential in nature. The instructions which are to be executed are held in the memory subsystem, as are the data which are to be operated upon. Instructions are fetched from memory, executed, and data is shifted to and from memory, and altered, according to the current instruction. In some supercomputers, a certain amount of the to-ing and fro-ing of data and the processing



in that data can occur simultaneously, but the speed which the supercomputer offers is achieved more because of the extremely sophisticated and fast electronic technology in which the supercomputers are implemented. This sophistication, and speed, does not come cheaply. And there are some fundamental upper limits to which the technology can be pushed; these are "hard" limits because they are imposed by the physics of the materials, rather than the ingenuity of the computer architecture engineers.

The main limitations of supercomputer technologies are, naturally enough, the speed of the processor (how fast it can switch states when computing), the speed of the memory (how fast an entry can be accessed), and the

Parallelism



speed of the communications between memory and processor (how fast data can be transferred). While in each of these three areas substantial progress has been made, the effective processing power of supercomputers has grown approximately 20 fold over the past decade from less than 200 Mflops(1) in 1982 to over 4 Gflops in 1992. Even being optimistic and allowing for a continuation in these trends, conventional supercomputing performance is not likely to break the 100 Gflop barrier before the end of the century. For the demanding applications of HPCN, performances of at least 100 Gflops and, more probably, 1 Tflop (one million million floating point operations per second) will be needed.

Parallel computers

The solution to the problem is not to depend on a single (or a few) very powerful processors to do all the work piece by piece, sequentially, but to do it in parallel. Parallel computing offers equally high, and potentially much higher, performance by harnessing the power of several (tens, hundreds, or thousands) of individual computing elements - processors - and by getting each one of them to contribute to the solution of the problem at hand at the same time: in parallel. This, of course, necessitates that the task can be "parallelized" and that communication between processors can be handled efficiently and effectively.

The successors of supercomputers will undoubtedly be based on parallel computing. In the September 92 issue of *IEEE Spectrum*, it was noted that "the conventional supercomputer now seems poised for an indefinite but inexhorable decline. On the verge of a gradual takeover, industry analysts believe, is the massively parallel processor." One only has to look at the present policy of computer manufacturers to be

The car industry provides one of the best examples of the use of advanced simulation via high performance computing. Along with economy, safety has become a critical issue in the design of modern cars.

convinced of this. For example, Cray Research, Digital Equipment Corporation, IBM, Thinking Machines, and Intel have all announced new products based on parallel architectures. The DECmpp 12000 series is in effect a DEC interface to a MasPar computer with a nominal performance of 1.2 Gflops, while Intel's Paragon XP/S system uses upwards of 1000 processors and offers, in principle, 5 to 300 Gflops.

The foundations for success in Europe in this critical area of information technology have already been laid through programmes such as Esprit. One of the most notable series of events in the area concerns the development of the T800 Transputer by Inmos and its successor, the T9000, which will soon be available. Building on this success, several European companies such as Meiko, Parsys, Parsytec, and Telmat now successfully market transputer-based parallel computing machines. For example, earlier in 1992 Meiko completed the sale of two supercomputers to Israel. One of these was a Meiko Computing Surface incorporating transputers exclusively; the other was a heterogeneous machine exploiting SPARC and Intel i860 processors and Inmos transputers. Meiko have also delivered a substantial system to Cray Research Corp. in the USA for eventual re-deployment in a US nuclear facility. They have also shipped a very significant order to Toyota Motor Corp. in Japan. This is a computer of the same type that is being delivered to Cray and to Israel (i.e. the heterogenous transputer/i860 machine) but is larger than either, with more than 64 nodes and is rated at 10 Gflops. This Meiko machine is a high performance simulation system and is being used in Toyota's engineering design process. Toyota invested in this capability in the expectation of a consequent six month reduction in the design and product tooling cycle for a new automobile; a reduction which can only but help increase their competitiveness in world markets.

Parsytec, whose main business is in low-cost parallel machines for industrial and embedded applications, have also had several commercial successes in very large parallel systems; for example, they recently delivered a massively-parallel 1024 processor system - the Parsytec GCel-3/1024 - to the Paderborn Centre for Parallel Computing. The peak performance of this system is more than 4.5 Gflops while the sustained performance exceeds 2 Gflops and it will be used for a variety of supercomputing applications.

The rate at which this emerging market will grow depends on several key issues. There must be adequate software support of parallel systems - languages and operating systems - which make it as easy to use and program parallel machines as it is with conventional sequential machines. The software should be as "standardized" as possible to ensure the portability of applications and a broad market. Environments or "work-benches" will be needed to ease the process of porting existing applications from conventional computers to the new parallel computers and to facilitate the development of new applications. Finally, as the number of processors which these parallel machines use reaches the thousands, occassional failure of individual processors is inevitable and parallel machines must have an in-built fault tolerance to allow automatic recovery from these faults.

The foundations for success in Europe in this critical area of information technology have already been laid through programmes such as Esprit.

Supercomputing is going through a fundamental change. We are witnessing a switch today away from the traditional "exotic" (and expensive) supercomputing technologies to new ones which exploit the power of parallelism - thousands of processors doing together what a single processor could not possibly manage alone. It is clear that, in terms of both R&D and market presence in parallel computing systems, European companies are poised for considerable success in an evolving and dynamically changing marketplace.

It is worth remembering, in the context of emerging markets, that the evolving parallel high-performance computing technology will not be confined to the traditional supercomputing market. Since it is based on less "exotic" technology than supercomputers heretofore, parallel high-performance computing technology will be taken up by the mainstream computing community. Here, parallel computing will include not just scientific and engineering applications - stand-alone, networked, or in embedded systems - but also commercial applications - banking, airline reservations, databases, etc. It now remains for Europe to exploit the power of parallelism. ■

David Vernon, *DG XIII*

[1] *A flop (floating point operation – an addition, subtraction, division, multiplication or other arithmetic operation on a real number) is the basic unit of computation for most HPC applications. One megaflop (Mflop) is equal to one million flops; one gigaflop (Gflop) is equal to one thousand million floating point operations and one teraflop (Tflop) is equal to a million million floating point operations. These figures all refer to the number of operations which the computer can execute in one second.*

DECEMBER 1992 No. 9

XIII MAGAZINE

EUROPEAN COMMUNITY POLICY FOR INFORMATION TECHNOLOGIES AND INDUSTRIES AND TELECOMMUNICATIONS

VALUE ADDED TELECOMS SERVICES

COMMISSION OF THE EUROPEAN COMMUNITIES
DIRECTORATE-GENERAL XIII