CHANGING HIGH PERFORMANCE COMPUTER TECHNOLOGY IS MAKING EXPORT CONTROL MORE DIFFICULT

New designs in HPCs and systems of computers, as well as availability of more advanced and less costly processors, software, and peripheral equipment, is rendering the challenge of applying export controls to HPCs more difficult.

For certain types of computer designs, the ability to add processors or boards could increase the machine's performance beyond authorized levels. In addition, advances in computer processor communications technology have facilitated the clustering of personal computers and workstations into effective parallel computers.

The usefulness of clustered computers is application-dependent. Some U.S. Government and computer industry experts have concluded that for many problems, networks of workstations could not compete with appropriately designed high performance computers.[206] Most traditional HPCs achieve far greater efficiency than parallel machines, due to their use of custom-made components.

Foreign access to high performance computers through networks is possible because of inadequate security measures.

Vector Architectures

Vector architecture relies on custom-designed processors to move a complex problem through computer processing units in sequential stages. This type of machine is designed to handle arithmetic operations efficiently on elements of arrays, called vectors.[207]

Vector systems are especially useful in high-performance scientific computing.[208] Vector systems, also called "pipeline" architectures, work like an assembly line. They work best with many similar tasks that can be broken down into steps.

The memory interface in vector machines is custom-made, and subject to export controls.

Vector machines are useful for cryptography, modeling fluids, and in the design of weapons. In particular, vector systems are suited to problems in which data at one point influence other variables in the problem, a common situation in national security applications.[209]

It is more straightforward for a programmer to use a vector system than a system comprised of parallel processors (discussed below), since it is easier to obtain maximum performance with one or a few high-power processors than with a collection of many lower capability processors.[210]

Since one of the main concerns with any HPC system is the rate of speed with which data can be retrieved from memory, another advantage is that a vector machine has a very fast memory.[211]

Still further advantages of vector systems are that they feature high memory bandwidth and low memory latency — that is, very large amounts of data can travel to and from memory very efficiently. A related advantage is that vector systems have the ability to seek multiple memory locations at the same time. This translates into very fast computational speed.

A disadvantage of a vector machine is that vector system software is not really portable. It cannot be readily transported to other vector machines.[212]

The main disadvantage of vector systems, however, is their high cost. Significant improvements in software and hardware allow the purchase of a parallel processing system for $40,000, as opposed to $1 million for a comparable vector computer.[213]

At the Defense Department's High Performance Computer Management Office, vector systems are being phased out in favor of parallel processing systems. Out of a total of 40 HPCs in the High Performance Computer Management Office inventory, fewer than 10 are now vector systems.[214]

Parallel Processing: The Connection of Computers Into a Powerful Central Resource

A parallel processing computer is a collection of processors that are connected through a communications network.[215] The type of processor, the network configuration, and the operating system that coordinates the activities distinguish parallel processing systems.

Many national security applications involve problems that can be separated into independent variables, and it is for these types of problems that parallel processing is best suited.[216]

The fastest parallel machines are all based on commodity processors — that is, processors that are commercially available on the market.[217] This approach has been applied to virtually every area of theoretical and applied physics.[218]

Massively Parallel Processors

A massively parallel processor is a collection of computers, or central processing units, linked together.[219] Each computer that is part of the whole massively parallel processor has its own memory, input/output system, and central processing unit.[220] Massively parallel processors now use commodity processors, and can utilize commodity interconnects to communicate between the individual computers that make up the system.[221] Some massively parallel processors use custom-made, very fast interconnect switches that are not commodities and are subject to export control.[222]

An advantage of a massively parallel processor is that an unlimited quantity of processors can be incorporated into the design of the machine. In a massively parallel processor, the more processors, the greater the computing speed of the machine.[223]

Because each processor is equipped with its own memory, massively parallel processors have much more memory than traditional supercomputers. The extra memory, in turn, suits these machines to data-intensive applications, such as imaging or comparing observational data with the predictions of models.[224]

A disadvantage of massively parallel processors is that memory latency is a bigger problem because the processors have to share the available memory. Another disadvantage is that each one of the computers that is part of the system has to be instructed what to do individually.[225] This phenomenon requires specialized, extremely proficient programmers to create efficient communications between the individual computers.

The commercial availability of inexpensive, powerful microprocessors has given massively parallel processors a boost in their competition with vector machines for the supercomputer market. IBM, for example, more than doubled the number of its computers in the Top 500 list (discussed below) between November 1997 and June 1998 by introducing the SP2, which strings together up to 512 of the company's RSI6000 workstation microprocessors.[226]

If optimum speed is desired, this massively parallel configuration is the best of all HPC designs.[227] The fastest high performance computer now available is the ASCI Blue Pacific.[228] That machine is part of the Department of Energy's Accelerated Strategic Computing Initiative (ASCI) program and is located at Lawrence Livermore National Laboratory. Developed in conjunction with IBM, it is a 5,856-processor machine, boasting a top speed of 3.8 teraflops[229] (Tflops) with 2.6 terabytes (Tbytes) of memory.[230] In the next phase of the ASCI initiative, IBM will deliver a 10-Tflops machine to the Department of Energy in mid-2000.[231]

Symmetrical Multiprocessor Systems

Symmetrical multiprocessor systems use multiple commodity central processing units (CPUs) that are tightly coupled via shared memory. The number of processors can be as low as two and as many as about 128.[232]

Symmetrical multiprocessor systems treat their multiple CPUs as one very fast CPU.[233] The CPUs in a symmetrical multiprocessor system are arranged on a single motherboard and share the same memory, input/output devices, operating system, and communications path.

Although symmetrical multiprocessor systems use multiple CPUs, they still perform sequential processing,[234] and allow multiple concurrent processes to be executed in parallel within different processors.[235]

An advantage of symmetrical multiprocessor systems is that the programming required to control the CPUs is simplified because of the sharing of common components.[236]

Another major advantage is cost. A Silicon Graphics symmetrical multiprocessor system, for example, with 18 microprocessors, each rated at 300 megaflops (MFLOPS)[237] or more, and a peak speed of more than 5 gigaflops (GFLOPS), costs about $1 million, whereas a Cray C90 costs about $30 million.[238]

Even though the Silicon Graphics machine is about a third as fast as the Cray machine, it is still very popular with consumers of these types of machines. The University of Illinois Supercomputing Center reportedly likes the price, flexibility, and future promise of symmetrical multiprocessor systems so much that it plans to use them exclusively within two years. Its older Crays were "cut up for scrap" at the beginning of this year, and its massively parallel computers will be phased out by 1997.[239]

One disadvantage of a symmetrical multiprocessor system is that all the CPUs on a single board share the resources of that board. This sharing limits the number of CPUs that can be placed on a single board.[240]

Although the programming model that a symmetrical multiprocessor system provides has proved to be user-friendly, the programmer must exercise care to produce efficient and correct parallel programs. To limit latency in individual jobs, most software requires enhancement — for example, employing special programming techniques to prevent components of the computer program from competing for system resources — thereby increasing inefficiency.

For this reason, symmetrical multiprocessor systems are not good platforms for high-performance real-time applications.[241]

In a symmetrical multiprocessor system design, as is true with a massively parallel processor system, the number of CPUs determines how fast a machine potentially will operate. This fact causes a problem for export controls because it is possible to add CPUs to the boards of a symmetrical multiprocessor system, or boards to a massively parallel processor system, and push the machine over export control thresholds after the original export-licensed purchase.[242]

Clusters of Commercial Off-the-Shelf Computers and Networks

Recent advances in the process of computer-to-computer communication, or networking, allow computers to be linked together, or "clustered." Networking has allowed the clustering of personal computers and workstations into well-balanced effective parallel computers, with much higher computing capabilities than any one of the clustered computers.[243]

Four thresholds have been crossed in connecting commercial-off-the-shelf components to create parallel computers:

- **Using commercial-off-the-shelf components to create parallel computers is simple** because of the ease of hardware configuration and the availability of all necessary system software from market vendors

- **It is versatile because a wide range of possible network designs** with excellent communication characteristics and scalability to large sizes is now available

- **Clustered systems performance has now matured** to the point that network communication speed is within 50 percent of that in vendor-assembled parallel computers[244]

- **Commercial-off-the-shelf clusters are now affordable**

According to officials at the Lawrence Livermore National Laboratory, networking represents only a 10 percent additional cost over the cost of the computing hardware for large systems. Thus, up to approximately 50,000 MTOPS, the computing capability available to any country today is limited only by the amount of money that is available to be spent on commercial-off-the-shelf networking.[245]

A typical commercial-off-the-shelf networking technology contains five essential elements. They are all inexpensive and widely available. The three hardware elements are switches (approximate cost: $2,000), cables (approximate cost: $100), and interface cards (approximate cost: $1,500). The two software elements are low-level network drivers for common operating systems, and industry standard communication libraries. The hardware and software technology necessary to successfully cluster commercial-off-the-shelf CPUs into effective parallel computers is well developed and disseminated in open, international collaborations worldwide.[246]

The concept of clustering commercial-off-the-shelf computers has been a subject of open academic study for over a decade. Today, the Beowulf Consortium acts as a focal point for information on clustering technology and has links to many projects. One Beowulf project is the Avalon computer at Los Alamos National Laboratory. Avalon can operate at 37,905 MTOPS[247] and was built in four days in April 1998 entirely from commodity personal computer technology (70 DEC Alpha CPUs) for $150,000.

Although commercial-off-the-shelf networking technology has only recently become effective, it has been adopted rapidly. There currently are at least seven competing high-performance network technologies (over 100 megabytes per second or higher): Myrinet, HIPPI, FiberChannel, Gigabit Ethernet, SCI, ATM, and VIA. One network vendor reported over 150 installations in the United States and 17 foreign countries including Australia, Brazil, Canada, the Netherlands, England, France, India, Israel, Italy, Japan, the Republic of Korea, and the PRC.[248]

Gigabit Ethernet is of particular interest because it is being developed by a cooperative, worldwide industry effort called the Gigabit Ethernet Alliance. 74 companies have pledged to develop products for the open standard — that is, the source software is available openly to software developers. Foreign companies are alliance members and also participate as members of the steering committee and the certification process for compliance. Gigabit Ethernet is projected to be a $3 billion market by the year 2000, which at today's prices translates into approximately 300,000 network switches per year.[249]

On October 15, 1997, a group of experts met to discuss computer performance metrics for export control purposes. The computer and high-tech industries were represented by Hewlett-Packard, Silicon Graphics/Cray Research, IBM, Digital Equipment Corporation, Intel, Sun Microsystems, the Center for Computing Sciences, the Institute for Defense Analyses, and Centerpoint Ventures. The U.S. Government was represented by the National Institute of Standards and Technology, the Naval Research Laboratory, the Defense Advanced Research Projects Agency, the National Security Agency, Lawrence Livermore National Laboratory, the Defense Technology Security Administration, and the Department of Commerce Bureau of Export Administration.[250]

The consensus of the discussion was that commercial-off-the-shelf networking is not so significant a threat to replace HPCs as might at first appear to be the case:

> ***Networks of workstations using [commercial-off-the-shelf] networking technology differ from supercomputers. Some problems will run easily and effectively on such networks, while other classes of problems important to national security concerns will not run effectively without a major software redesign effort.*** *For many problems no amount of software redesign will allow networks of workstations to compete with appropriately designed high performance computers.*
>
> ***Even if a "rogue state" assembled such a large network of workstations by legitimately acquiring large numbers of commodity processors, the actual effort to produce the software necessary to realize the full potential of such an aggregate system would take several years.*** *During this time, the state of the art of computational technology would have increased by approximately an order of magnitude.*

> *After considerable discussion, most of the participants were in agreement that there was a fundamental difference between a system designed by a single vendor that was built as an aggregate of many commodity processors and included the software to enable these processors to cooperatively work on solving single problems of national concern, and a large collection of commodity processors not subject to export control that are externally networked together.*[251]

According to one expert, many universities have clustered systems, as they are easy to establish. For $70,000, a 12-node system with two Pentium II processors at 300 megahertz (MHz) each would produce a system with 7,200 GFLOPS.. However, the system must be properly structured to perform well, and performance will vary depending on the application, the programmer's ability, and the connection of the machines. An integrated system from Silicon Graphics/Cray will achieve between 10-20 percent of peak performance at best.[252]

An example of a powerful commercial-off-the-shelf network can be found at the Illinois Supercomputing Center. Four eight-processor and two 16-processor machines from Silicon Graphics are connected in a cluster with a peak speed of nearly 20 GFLOPS.[253]

According to one expert, it does not require any special expertise to network workstations using commercial-off-the-shelf technology. The software engineering techniques are being taught to undergraduates as part of standard courses in advanced computing, but anyone with programming knowledge should be able to create a network as well.[254]

The parallel supercomputers of today have peak speeds of over 100 billion floating point operations per second (100 GFLOPS). This is roughly 100 times the peak speed of a Cray YMP class machine, which was the standard for high-performance computing of just five years ago.[255]

However, it is difficult to achieve a high percentage of this peak performance on a parallel machine.

Whereas a tuned code running on a Cray might reach 80-90 percent of peak speed, codes running on parallel computers typically execute at only 10-20 percent of peak.[256] There are two reasons for this:

- **The first is that Cray-class computers incorporate extremely expensive, custom-designed processors with vector-processing hardware.** These processors are designed to stream large amounts of data through a highly efficient calculational pipeline. Codes that have been tuned to take advantage of this hardware ("vectorized" codes) tend to run at high percentages of peak speed.[257]

  Parallel machines, on the other hand, are generally built from much simpler building blocks. For example, they may use the same processors that are used in stand-alone computer workstations. Individually, these processors are not nearly so sophisticated or so efficient as the vector processors. Thus, it is not possible to achieve so high a percentage of peak speed.[258]

  Some parallel machines contain custom processors (TMC CM-5 vector units) or custom modifications of off-the-shelf processors (Cray T-3D modified DEC alpha chips). Even in those cases, however, the percent of peak achievable on a single node is still on the order of 50 percent or less. In parallel computer design, there is constant tension between the need to use commodity parts as the computational building blocks in order to achieve economies of scale, and the desire to achieve ever-higher percentages of peak performance through the implementation of custom hardware.[259]

- **The second reason that parallel computers run at lower percentages of peak speeds than vector supercomputers is communications overhead.** On parallel computers, the extraordinary peak speeds of 100 GFLOPS or more are achieved by linking hundreds or even thousands of processors with a fast communications network.

> Virtually all parallel computers today are “distributed memory” computers. This means that the random access memory (RAM) is spread though the machine, typically 32 megabytes at each node. When a calculation is performed on a parallel machine, access is frequently needed to pieces of data on different nodes.
>
> It may be possible to overlap this communication with another computation in a different part of the program in order not to delay the entire program while waiting for the communication, but this is not always the case. Since the timing clock continues while the communication is taking place, even though no calculational work is being performed, the measured performance of the code goes down and a lower percentage of peak performance is recorded.[260]

Domain Decomposition

“Domain decomposition” involves partitioning the data to be processed by a parallel program across the machine’s processors.[261]

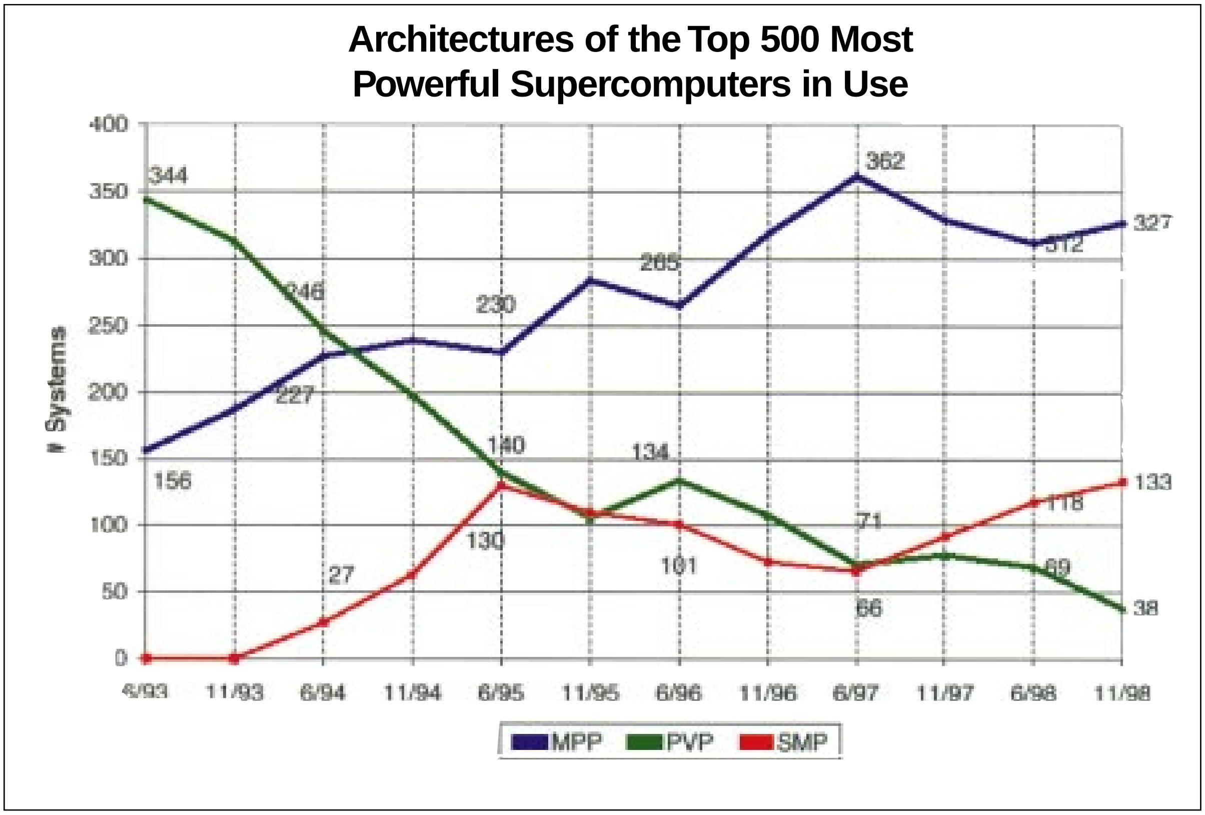
In distributed memory architectures, each processor has direct access only to the portion of main memory that is physically located on its node. In order to access other memory on the machine, it must communicate with the node on which that memory is located and send explicit requests to that node for data.[262] Figuring out the optimal domain decomposition for a problem is one of the most basic and important tasks in parallel computing, since it determines the balance between communication and computation in a program and, ultimately, how fast that program will run.[263]

Memory access constitutes an inherent bottleneck in shared-memory systems.[264]

Highly Parallel Technology

Microprocessor-based supercomputing has brought about a major change in accessibility and affordability. Massively parallel processors continue to account for more than half of all installed supercomputers worldwide, but there is a move toward shared memory, including the use of more symmetrical multiprocessor systems and of distributed-shared memory. There is also a tendency to promote scalability through the clustering of shared memory machines because of the increased efficiency of message passing this offers. The task of data parallel programming has been helped by standardization efforts such as Message Passing Interface and High-Performance Fortran.[265]

Highly parallel technology is becoming popular for the following reasons. First, affordable parallel systems now out-perform the best conventional supercomputers. Cost is, of course, a strong factor, and the performance per dollar of parallel systems is particularly favorable.[266] The reliability of these systems has greatly improved. Both third-party scientific and engineering applications, as well as business applications, are now appearing. Thus, commercial customers, not just research labs, are acquiring parallel systems.[267]



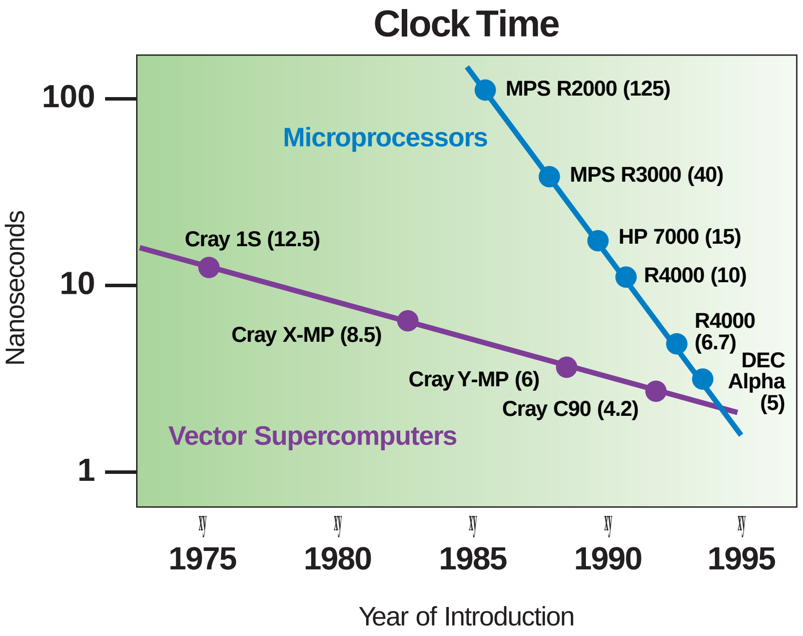
Since late 1993, massively parallel processors (MPP) and symmetrical multiprocessor systems (SMP) began to overtake vector systems (PVP) as the most powerful computer systems in use. Affordable parallel systems now out-perform the best conventional supercomputers. While cost is one reason, the reliability of such systems has greatly improved.

Twice a year the "Top 500 list," a compendium of the 500 most powerful computer systems, is published.[268] On the previous page is an example of the numbers and types of systems in the biannual list of the top 500 fastest computers. As this chart points out, massively parallel processors and symmetrical multiprocessor systems are on the rise, while vector systems are losing ground.[269]

Microprocessor Technology

While vector and massively parallel computers have been contending for the supercomputing market, an important new factor has become the availability of extremely powerful commodity microprocessors, the mass-produced chips at the heart of computer workstations.

Ten years ago, workstation microprocessors were far slower than the processors in supercomputers. The fastest microprocessor in 1988, for example, was rated at one million floating point operations per second (MFLOPS) while Cray's processors were rated at 200 MFLOPS.[270] A floating-point operation is the equivalent of multiplying



Ten years ago, workstation microprocessors were far slower than the processors in supercomputers. Today, Cray's processors have improved by a factor of 10, to two gigaflops in the new T90. But the faster microprocessor runs at 600 MFLOPS, an improvement by a factor of 600.

two 15-digit numbers. Today, Cray's processors have improved by a factor of ten, to two gigaflops in the brand-new T90; but the fastest microprocessor runs at 600 MFLOPS, an improvement by a factor of 600.

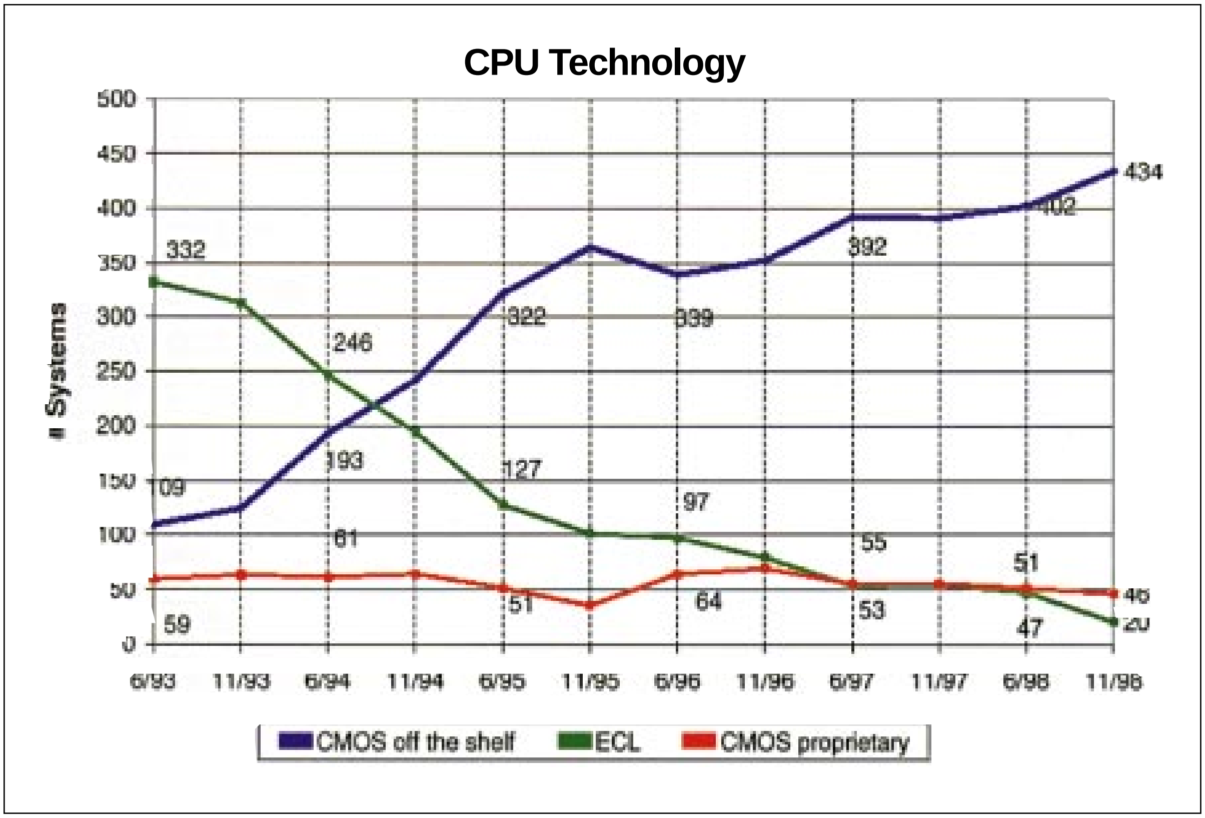
Commercial off-the-shelf microprocessor power is available for a fraction of the cost of a traditional vector processor. Unlike vector processors, which consist of complex collections of chips and are only fabricated by the hundreds each year, commercial off-the-shelf microprocessors are designed for mass production based on two decades of experience making integrated circuits. Research and development costs for each commercial off-the-shelf microprocessor are spread over hundreds of thousands of chips.[271]

Microprocessors, also known as CPUs, are integrated circuits. They can be divided into broad categories of logic family technologies. The selection of a certain logic technology in the design of an integrated circuit is made after determining an application and weighing the advantages of each type of logic family. Among these are:

- **Emitter-Coupled Logic (ECL)** is used for circuits that will operate in a high-speed environment, as it offers the fastest switching speeds of all logic families; it is the first type HPC chip. ECL, however, is power-hungry, requires complex cooling techniques, and is expensive.[272]

- **Complementary Metal-Oxide Semiconductor Logic (CMOS)** is relatively inexpensive, compact and requires small amounts of power. CMOS off-the-shelf is the standard PC or workstation chip; proprietary CMOS is custom-built, specially designed for the particular HPC and incompatible with PCs and workstations.

Realizing the differences between logic technologies gives a perspective to understanding where CPU technology is headed, and the reasons that the market is driving one technology faster than another. As the following chart illustrates, commercial off-the-shelf, inexpensive CPUs are coming to dominate the high performance computing world.[273]



Inexpensive commercial, off-the-shelf CPUs utilizing complementary metal-oxide semiconductor logic (CMOS) in their circuitry are beginning to dominate the high performance computing world, beating out CPUs using the faster emitter-coupled logic (ECL). The latter technology, however, is power-hungry and requires complex cooling techniques that make it more expensive.

Interconnect Technology

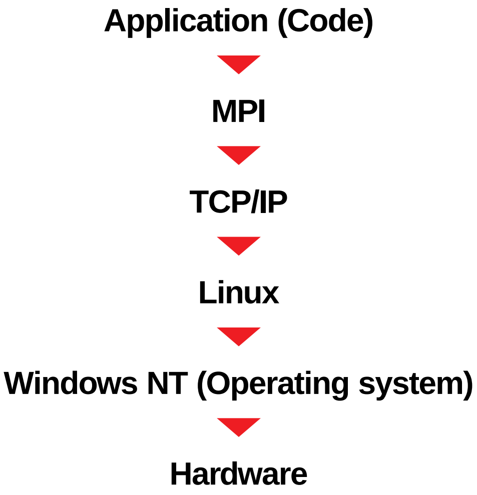
In multiprocessor systems, actual performance is strongly influenced by the quality of the "interconnect" that moves data among processors and memory subsystems.[274]

Traditionally, interconnects could be grouped into two categories: proprietary high-performance interconnects that were used within the products of individual vendors, and industry standard interconnects that were more readily available on the market, such as local area networks.[275] The two categories featured different capabilities, measured in bandwidth and latency.

Recently, a new class of interconnect has emerged: clustering interconnects. These offer much higher bandwidth and lower latency than local area networks. Their shortcomings are comparable to proprietary high-performance interconnects, including lower bandwidth, higher latency, and greater performance degradation in large configurations or immature system software environments.[276]

Message Passing Interface

Message Passing Interface (MPI) is a program containing a set of sub-routines that provide a method of communication that enables various components of a parallel computer system to act in concert. The communications protocol that MPI uses is the same utilized by the Internet. According to Dr. Jeff Hollingsworth of the University of Maryland Computer Science Department, an example of how each of the different software applications interact with the hardware would be as follows:[:277]



Application (Code)

MPI

TCP/IP

Linux

Windows NT (Operating system)

Hardware

Some software, says Hollingsworth, is sold in a version that is compatible with MPI. One example is automobile crash simulation software. This software, which is essentially code to simulate a physical system in three dimensions, is adaptable to other scientific applications such as fluid dynamics, according to Hollingsworth.[278]

Hollingsworth states that software that is not already “MPI ready” can be modified into code that can be run in an MPI, or parallel, environment. Modifying this software to enable it to run in an MPI environment can be very difficult, or quite easy, says Hollingsworth, depending on “data decomposition.” [279]

The ease of converting software that is not “MPI ready” into an “MPI ready” version is dependent on the expertise of the software engineers and scientists working on the problem. For a single application and a single computer program, the level of expertise required to convert a computer program in this way is attainable in graduate level, and some undergraduate level, college courses, according to Hollingsworth.[280]

It has not been possible to determine which, if any, commercially available software is both MPI ready and applicable to defense-related scientific work.