From Megaflops to Total Solutions

The Changing Dynamics of Competitiveness in Supercomputing

Boelie Elzen   Donald MacKenzie

Introduction

In the latter part of the 1980s, headlines in U.S. newspapers suggested that the American supercomputer industry was on the brink of collapse. United States science and technology was losing its competitive advantage because American scientists and engineers did not have sufficient access to supercomputer power.

In this chapter[1] we will examine the validity of these claims and conclude that they are unfounded. To this end, we will first give an historical sketch of the development of supercomputers over the past three decades. We will describe the development of the machines in conjunction with the development of the network of social relationships connected to them. We will trace both how technological development affected social relations and how social relations affected subsequent technological development, creating a coherent "sociotechnical network"[2] which is difficult for outside competitors to break into. In a concluding section we argue that reports on the "bad shape" of the American supercomputer industry are unduly pessimistic because they neglect the coherency of this network and are based on too narrow a view of what is and has been going on.

Supercomputers can be defined pragmatically as the fastest computers at any given time.[3] This definition suggests that the development of supercomputers is primarily driven by a quest for higher speeds. Indeed, the early history of supercomputing shows that certain factors that were of prime importance in the development of commercial mainframe computers were subservient to speed considerations in the development of supercomputers. This led to the emergence of a class of machines that, despite their high cost (of the order of $10 million or more), found their way to quite a number of users. The coming into being of a network of relationships between supercomputer manufacturers and these users created a mutual dependency that affected the design of subsequent generations of machine. Paradoxically, one of the effects of the development of this network is that speed has become a less dominant factor in the development of supercomputers.

Control Data Corporation

Around 1960 the performance and reliability of computers leapt forward as the result of the introduction of a new component technology: the transistor. Early commercial transistorized computers were delivered by IBM and the Control Data Corporation (CDC), notably the IBM 7090 and the CDC 1604.[4] IBM, of course, was already a well-established company, while CDC was incorporated only in 1957. The 1604 was CDC's first product, designed in three years by a team directed by Seymour R. Cray.[5]

Seymour Cray had studied electrical engineering and applied mathematics at the University of Minnesota. In 1950 he was recruited by Engineering Research Associates (ERA) of St. Paul, Minnesota.[6] With its origins in wartime cryptanalytic work, ERA was one of the pioneers of digital computing in the United States, though the secrecy of the code-breaking task (its continuing primary market) meant that the firm's work was much less well known than that of J. Presper Eckert and John W. Mauchly in Philadelphia. In May 1952, however, ERA was sold to Remington Rand, which already owned Eckert-Mauchly, and in June 1955 Remington Rand merged with the Sperry Corporation to form Sperry Rand.

There are few known details of Cray's work for ERA and Sperry Rand, though the young Cray quickly won considerable responsibility, notably for Sperry Rand's Naval Tactical Data System (NTDS) computer. He was thus already a figure of some importance to his first start-up company, the Control Data Corporation, formed when Cray and eight others, most famously William C. Norris, left Sperry Rand in 1957. Cray was the chief designer of Control Data's first computer, the CDC 1604, announced in October 1959. Built from transistors, rather than the previously pervasive vacuum tubes, the highly successful 1604 moved Control Data into profit and launched it on a path that was to enable it briefly to challenge IBM's dominance of the computer industry, a dominance that was already hardening by 1957.

The NDTS, and especially the 1604, were considerable achievements, and secured Cray's growing reputation as a computer designer. Yet neither was the stuff of legend, nor—beyond the beginnings of anecdotes concerning his preference for simple designs and intolerance of those he considered fools[7]—is there much evidence of a distinctive "Cray style" in their development.

The origins of both legend and style, and the first clear manifestation of what was to become Cray's distinctive sociotechnical strategy, can first be seen unequivocally in discussions within Control Data on what to do to follow the company's success with the 1604. The obvious step was to build directly on that success, offering an improved machine, but one compatible with the 1604 (so users of the latter could run their programs unaltered). While the 1604 had been oriented to the demands of "scientific" users, such as defense contractors and universities, there was a growing sense within Control Data of the need to orient at least equally to business data processing, where arithmetic speed was of less concern than the capacity to manipulate large data sets. Compatibility and business orientation were not necessarily at odds. By adding new instructions, specially tailored for commercial usage, to the instruction set of the 1604, business demands could be catered to without sacrificing compatibility with the previous machine.

This emerging strategy was perfectly sensible. It was indeed similar to, if less ambitious than, that to be announced in 1964 by IBM, with its famous System 360. This was a series of compatible machines, some oriented to the business and some to the scientific market, but all sharing the same basic architecture, and with an instruction set rich enough to cater to both sorts of demand. Cray, however, disagreed with all elements of the strategy—compatibility with the existing machine, orientation to the commercial as well as scientific market, a complex instruction set.

His alternative strategy prioritized speed: in particular, speed at the "floating-point" arithmetic operations that were the dominant concern of defense and scientific users. In that prioritization, Cray did not wish to be constrained by choices made in the development of the 1604. Compatibility was to be sacrificed to speed. As one of his famous maxims has it, he likes to start the design of a new-generation machine with "a clean sheet of paper." He had no interest in business data processing, and abhorred the complexity that arose from trying to satisfy both scientific and business users.

The 1604 was making a lot of money for Control Data, and so it seemed possible to pursue both strategies simultaneously. One group of designers went on to develop a series of complex-instruction-set computers compatible with the 1604 (the Control Data 3600 series), with a primary orientation to the commercial market. A second group, led by Cray, set out to develop the fastest machine they could conceive of, which became the CDC 6600.

Cray's status as the chief designer of the corporation's first and most successful computer, and the threat (possibly explicit) that he would leave,[8] enabled him to negotiate in 1961/1962 a remarkable arrangement with Control Data chairman Norris. He was allowed to move, with the small team working on the 6600, a hundred miles away from Control Data's headquarters in Minneapolis-St. Paul, to a newly built laboratory on a plot of country land, owned by Cray personally and close to his house, in woods overlooking the Chippewa River. Cray thus won a remarkable degree of autonomy from corporate control. Even Norris had to seek Cray's permission to come to the Chippewa laboratory, and Cray visited Control Data headquarters only every few months.

The technical and social aspects of Cray's strategy were closely related. Chippewa-style isolation would have been incompatible with successfully building a series of compatible, general-purpose computers, which required determining the needs of different kinds of users; balancing one technical characteristic against another; giving attention to software as well as hardware; keeping different projects connected together; harnessing all the different parts of a growing corporation to a common, but diffuse, set of tasks; "committee design."[9] By the move to Chippewa, Cray created a geographical and social barrier between his team and all this negotiation and compromise.

The instruction set for the computer designed at Chippewa, the Control Data 6600, is emblematic of Cray's sociotechnical strategy. It contained only 64 instructions, at a time when a hundred or more were common. When the attempt is being made to satisfy a variety of different user concerns, the easiest means of harmonization is to satisfy vested interests by adding instructions. The IBM Stretch computer, designed in the late 1950s, is an extreme example. An intensely ambitious project, intended to combine extreme speed with an attempt to straddle the scientific, cryptanalyst, and business markets, Stretch had an instruction set of no fewer than 735 instructions.[10]

A simple instruction set for the 6600 permitted most of those instructions to have their own hardware support, tailor-made for speed. Ten independent processing units separately handled logical operations, multiplication, division, and so on. Input/output operations were hived off to ten independent peripheral processing units (PPU) (see Fig. 1).

Striking though its overall design is, the 6600 by no means emerged, Athena-like, simply from the brain of Seymour Cray. The instruction-set simplicity of the 6600 became architectural complexity, for example, as a sophisticated "scoreboard" unit had to be designed to keep the different independent processors working harmoniously. Even Cray could not on his own master all the details, so even within his small team an internal division of labor was needed. James Thornton, in particular, took responsibility for much of the detailed design.

The combination of Cray's intense personal involvement and the laboratory's isolation lent coherence to the project: "A team spirit developed and carried over into sporting and recreational events in the community."[11] Developing the 6600,

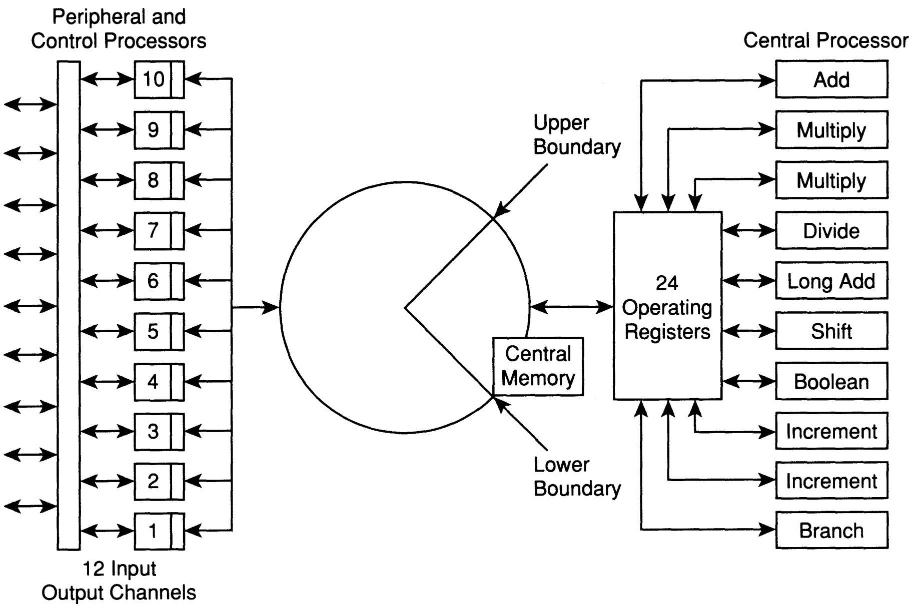

Figure 1. Block diagram of the Control Data 6600. (*Source:* James E. Thornton, "The CDC 6600 Project," *Ann. His. Comput.*, Vol. 2, no. 4, October 1980, pp. 338–348, 346)

however, involved far more than the sociotechnical work of leading the team at Chippewa. Cray did not attempt to develop the basic components for the machine: developing an innovative configuration or "architecture" for them was work enough. This placed his team on the horns of a dilemma. A conservative choice of components would reduce risks, but might not give the speed that was necessary. The other Control Data teams were no slouches, despite the range of needs they were seeking to satisfy, and to justify itself within Control Data, much less find a place on the market, the 6600 project had to be a lot faster than the 3600 series. Components at the state of the art, or just beyond it, would give the edge in speed, but would place the fate of Cray's project in the hands of their developers, over whom he had no control.

Cray's preferred approach was conservative—"keep a decade behind" is one of his sayings on display at Boston's Computer Museum—and his team began by trying to wring a 15- to 20-fold speed increase over the 1604 without a radical change in components. They found this impossible to achieve. Fortunately, a new silicon transistor, manufactured by Fairchild Semiconductor, appeared on the market in time to salvage the project, and design was begun again with that as its basis, though the speed goal of the delayed project had to be increased relative to the 3600 to make up for the lost time.

The problematic relationship between computer designer and component supplier is a theme that was to recur in the history of supercomputing. So is another issue that came to the fore during the development of the 6600. Like almost all other computers, the 6600's operations were synchronized by pulses in the control circuitry; the intervals between those pulses were its "clock cycle." The target clock cycle for the 6600 was 100 nanoseconds, one ten-millionth of a second. In such a tiny interval of time, the finite speed of electrical signals became a constraint. If the wires were too long, a signal would not arrive at its destination within one cycle of the clock. So the circuitry of the 6600 had to be packaged very densely. With the silicon transistors they got a 10-fold density improvement, but dense packaging meant intense heat generation.

This heat had to be removed. Cray grasped the centrality of what others might have considered a menial aspect of computer design, and superintended the design of a special cooling system, with Freon refrigerant circulating though pipes in the machine's structure to remove the heat.

To produce a machine of the 6600's daunting complexity was no easy task. The wider computer industry had already started applying its own products in more and more automated design and production systems. Cray took a step in the opposite direction. The most sophisticated computer of its day was, in effect, handcrafted. Cray was even reluctant to turn it over to the Control Data production facilities in the Minneapolis suburb of Arden Hills, so the first few 6600s were built at Chippewa. Finally, the transition was successfully made, but it was no simple matter of handing over blueprints. The production process, and integrating the large network of suppliers whose parts went into the 6600, required the most careful attention, though not from Cray himself. His habit was to delegate the task to others in his team, but he was always fortunate in the people to whom he delegated it. Les Davis, who was Cray's chief engineer for almost three decades, played a particularly crucial role in making Seymour Cray's ideas work.

This connection to the outside world was not the only way the boundary around Chippewa had to be made permeable. Potential customers also had to have access, so while users in general were kept at arm's length, a few select people passed relatively freely between Chippewa and sites where the 6600 might be used. The most crucial such site was the nuclear weapons laboratory at Livermore, California. As it turned out, the Livermore Director of Computing, Sidney Fernbach, had easier access to Cray's laboratory than Norris did, and close liaison developed between the two sites. Cray, widely considered a "technical dictator," was prepared to listen to Fernbach's advice about how to shape the 6600 to make sure it met Livermore's unique needs for computer power.[12]

As a result, the 6600 provided a quantum leap in the computing power available to Livermore and its competitor nuclear laboratory at Los Alamos. It even gave the United States a temporary lever in its attempt to control French nuclear weapons policy, the American government in 1966 blocking the export of a Control Data 6600 destined for the French bomb program (though the requisite calculations were performed surreptitiously on an apparently civil 6600).[13] Although the term was not yet used, the 6600 was indeed a supercomputer, enjoying a significant advantage in arithmetic speed over all other machines of its day, worldwide.

Even Control Data, which had had to proceed much more on faith than had Fernbach—Cray's reports to his employers were delightfully succinct and uninformative[14]—was repaid. The largest sales achieved by any previous supercomputer, IBM's Stretch, was eight. Soon after its introduction in 1964, the 6600 became a big success and, before the decade was out, had orders exceeding 100, at around $8 million a machine. IBM was rattled, the 6600 prompting a famous acerbic memo from chairman Thomas J. Watson, Jr., to his staff, enquiring why Cray's team of "34 people—including the janitor" had outperformed the computing industry's mightiest corporation.[15]

"Big Blue," as the rest of the industry called IBM, bent its efforts to developing, out of the basic multipurpose Series 360 architecture, "top-end" machines, to compete with Cray. Although IBM controlled vastly more resources, and had at its call considerable talent (including Gene Amdahl, a computer designer of great skill who was to become almost as famous as Cray), it failed. Ultimately, IBM was not prepared to sacrifice compatibility for speed; nor, perhaps, were the "social" aspects of Cray's strategy replicable within a giant organization with a notoriously conformist corporate culture. The 1967 IBM 360/91 surpassed the 6600, but Cray was a moving target, and had his successor machine, the 7600, ready by 1969. It was 1971 when IBM, with the 360/195, caught up. IBM sought to compensate for its lag by announcing these machines well in advance of their readiness. Control Data responded with the computer industry's most famous law suit, using the U.S.'s antitrust laws to charge IBM with illegal monopolistic practices.

Yet all was not entirely well with Cray's strategy, and IBM's countermoves were only part of the problem. The 7600 was built very much according to the same priorities as the 6600, but while the 6600 offered a speed advantage of as much as 20 over previous-generation computers, the 7600 was only four times faster than the 6600. Was it worth spending a further $10 million and significant amounts of time modifying programs, to obtain that degree of speed-up? Some 6600 users, notably the nuclear weapons laboratories, answered in the affirmative. But many others said "no." Sales of the 7600, while still healthy, were only half those of the 6600. In particular, universities, a large sector of the 6600's market, failed to upgrade.

Initially, Cray seemed unperturbed. He and his team began designing a further machine, the 8600. This involved a bigger break from the 7600 than that between the 7600 and 6600. Despairing of achieving large enough speed increases by new components and incremental changes to architecture, Cray moved to embrace the much discussed, but as yet little practiced, principle of parallelism. The 8600 was to have four central processing units working simultaneously, while all previous Cray machines (and nearly all previous computers of whatever kind) had but one. Again, an idea that seems simple in principle turned out complex in practice. Ensuring adequate communication between the processors, and preventing them from contending for access to the computer's one memory, were formidable problems.[16]

While Seymour Cray and his team were working on the 8600, another team within Control Data, led initially by Cray's former deputy James Thornton, was working on a rival machine: the STAR-100. Central to the STAR-100 was an idea at least as novel as multiple central processors: vector processing. In a vector processor one instruction can be used to perform a certain operation, not just on one or two pieces of data (as in a conventional "scalar" computer), but on large ordered sets ("strings" or vectors). An example would be an instruction to add two strings each of 100 figures to give one string of 100 figures as a result. If the data could be organized in this way (and many of the problems of interest to the weapons designers at Livermore appeared, at least at first sight, to have this kind of regularity), considerable gains in speed could be achieved without the complexity of multiple central processors.

With the backing of Fernbach and Livermore, Control Data began work on the STAR-100 (a STring ARray processor with the goal of 100 million results per second), but at the Arden Hills site, not Chippewa. The strategy differed from Cray's. Speed was indeed a priority, but users of different kinds had to be catered to. The cryptanalysts of the National Security Agency persuaded the STAR's developers to add hardware support for what those developers referred to as "spook instructions": data manipulations of particular interest to cryptanalysis. Control Data management (which had much greater access to the design process of the STAR than to that of the Cray machines) saw the STAR as the centerpiece of an integrated, compatible set of computers analogous to, but more advanced than, the IBM 360 Series. The result was a very large instruction set of over 200 instructions. For a long period, too, leadership of the project was ambiguous, the machine becoming an "engineers' paradise" in which everybody could have some novel ideas incorporated, but in which communication and coordination was poor. Only determined action by Neil Lincoln, who eventually replaced Thornton as project leader (Thornton left Control Data to set up his own firm, Network Systems Corporation) finally achieved delivery of the STAR to Livermore in 1974, four years late.[17]

The STAR was—indeed still is—a controversial machine. Its adherents point out that it did achieve, indeed, surpassed, its impressive goal of 100 million results per second, and note that the vector processing pioneered on the STAR was to dominate at least the next two decades of supercomputing. Its detractors point out that it only approached its top speed on special programs that allowed extensive use of its vector capabilities. In scalar mode, in which one instruction produces one result, the machine was slower than the CDC 7600 of five years previous. The judgment of the market at the time was with the detractors: only three STARS were sold.

The very existence of the STAR project in the late 1960s and early 1970s, however, added further to the internal troubles of the 8600 project. Both factors were compounded by a change in direction at the top level of Control Data. Diagnosing "a great change" taking place in the supercomputer market, William C. Norris, CDC's President and Chairman of the Board, said that Control Data's high-speed scientific computers had developed to a point where customers now needed little more in the way of increased speed and power. Instead, said Norris, supercomputer users were demanding service and software to help them get more effective use of the speed they already had. "In other words," he concluded, "the emphasis today shifts to applying very large computers as opposed to development of more power." Although Control Data would continue to build and market large-scale scientific computers, investment in their research and development would be curtailed.[18]

Control Data's new corporate plan allowed for supercomputer development, but not at the pace Cray wanted—a completely fresh, significantly faster machine, every five years. The increasingly tenuous ties between Cray and Control Data were severed in 1972: "Since building large computers is my hobby, I decided that with this shift in emphasis, it was time for me to make a change," Cray said. He left to start his own company, taking four colleagues with him.[19]

Cray Research

It says a lot about the respect in which Cray was held that the split was surprisingly amicable. Control Data's Commercial Credit Company even invested $500,000 in the new Cray Research, Inc. (CRI), adding to $500,000 of Cray's own money and a total of $1,500,000 from 14 other investors: the computer business had been sufficiently profitable that Cray had several personal friends within it who were able to put in as much as $250,000 each. Cray was both president and chief executive officer. Finally, he could give full expression to his sociotechnical strategy, without even the residual encumberances of a large corporation.

The strategy was of breathtaking simplicity, but it also made sense. Cray Research would build and sell one machine at a time, and each machine would be a supercomputer. There would be no diversified product range, no attempt to make money (as Control Data very successfully did) primarily from the sale of peripherals like disk drives and printers, no dilution of the commitment to build the fastest possible machine. By delimiting the goal, and keeping to a single development team of perhaps 20 people under the undistracted command of Seymour Cray himself, costs could be kept small. A selling price sufficiently above cost would be set to cover research and development expenditures.[19]

That the customer base for the world's fastest computer was small—Cray estimated it at 50—did not disturb him. Indeed, it was to his advantage that he already knew who his potential customers were—those purchasers of the 6600 and 7600, such as the nuclear weapons laboratories, whose demand for speed was still unsatisfied and was perhaps insatiable. They would be prepared to pay the by now traditional supercomputer price of around $8 million to $10 million per machine. The economics of his enterprise was, indeed, clear. If he could achieve the high-performance computer industry's traditional margin of a selling price of three times manufacturing cost, the proceeds of a single sale would recoup the entire initial capital investment.

Cray could not afford to take any risks with component technology, and certainly could not afford to develop it within Cray Research. He chose a very simple, but reliable, integrated circuit. It was, however, by no means fast enough on its own to give anything like the increase in speed needed to establish his new machine, to be called simply the CRAY-1.

It was in the choice of the Cray-1's architecture that Cray displayed a flexibility often absent in single-minded, dominant technical entrepreneurs.[20] He abandoned the multiple processor approach that had failed on the CDC 8600, and adopted that of its rival, the STAR-100—vector processing.[21] However, he had the advantage over the STAR's designers that he had the failings of a real machine (or at least, an advanced development project) to learn from. Like others, he concluded that the STAR had two interrelated flaws.

First, the STAR's scalar performance was far slower than its vector speed, so if even a small part of a program could not be made suitable for vector processing, the overall speed of running that program would be drastically reduced. Cray therefore decided to place great emphasis on giving the CRAY-1 the fastest possible scalar processor. Second, the full vector speed of the STAR was achieved only if data could be packaged into regular vectors of considerable size. This was partly attributed to the fact that the STAR processed the vectors directly from memory and then sent the results back to memory, in effect using a "pipeline" that was very fast when full, but which took a relatively long time to fill. Cray decided instead to introduce a small intermediate storage level ("vector registers"), built from very fast, but extremely expensive, memory chips.

Other differences between the CRAY-1 and the STAR were predictable consequences of their very different circumstances of development. Cray did not worry about compatibility with any other machine, whether designed by him or anyone else. Once again, he had his "clean sheet of paper." The CRAY-1's instruction set is in fact more complex than the 6600's—it may be that Cray's old links to cryptanalysis came into play—but the great elaboration of the STAR was avoided. The old issues of physical size and cooling were once again central. The STAR's "memory-to-memory" pipeline, and relatively slow scalar unit, permitted a physically large machine. Cray's fast scalar unit, and vector register design, did not.

His goal was a clock cycle of 12.5 nanoseconds, well below the 40 nanoseconds of the STAR. In the former time interval, even light in free space travels no more than 4 meters, and an electric signal in a wire is slower. This goal influenced several technical decisions. Along with the continuing fear of placing himself in the hands of others whom he could not control, it persuaded even Cray to override, as far as memory design was concerned, his motto about keeping "a decade behind." In previous machines he had always used magnetic core memories. Under the impact of competition from the new semiconductor memories, however, the manufacturers of core memories were concentrating on the cheap, low-performance end of the market. Cray therefore decided to opt for slower, but physically smaller and reliably available, semiconductor memory chips.[22]

Shrinking the machine also intensified the familiar problem of heat. The CRAY-1 generated about four times as much heat per cubic centimeter as the 7600. Cray's team therefore developed a new cooling scheme for the CRAY-1. Its integrated circuits were mounted on boards back-to-back with copper plates built onto vertical columns of "cold bars"—aluminum blocks containing stainless steel tubes through which Freon coolant flowed. The complete machine consisted of twelve columns arranged in a 270-degree arc, thus giving the machine its now famously elegant C-shaped horizontal cross-section[23] (see Fig. 2).

The CRAY-1 was as much a *tour de force* as the 6600. All but a few adherents of the STAR accepted that the CRAY-1 was, by a large margin, the world's fastest machine when it appeared in 1976. It is interesting to note, however, that it was a triumph of Cray's general sociotechnical strategy and insight into the weak points of previous designs, rather than of specific invention. Among the few parts of the original design that were patented were the cooling system and the vector registers.

It was nevertheless a technical triumph. This time around, IBM did not even try to compete: corporate pride was outweighed by the memory of past failures and by a sense that supercomputing was merely a market niche of limited size, rather than the flagship of all computing. Control Data tried, with a reengineered, improved version of the STAR, the 1981 Cyber 205. It was a strong technical rival to the CRAY-1—faster on long vectors, though still not as fast on short ones—but it was too late. By 1985, around 30 Cyber 205s had been sold—not bad by the standards of the 1960s, but as we shall see, not good enough by the new standards of the 1980s, standards set by Cray Research.

The Transformation of the Cray Strategy

What made the chase ultimately fruitless was not any entrenched speed advantage of the CRAY-1 and successor Cray Research machines. The Cyber 205, and its successor, the ETA^{10}, were indeed faster on some measures than the Cray machines. Similarly, the Japanese industry's entry into supercomputing during the 1980s led to machines that were faster than Cray Research's on some criteria, and the American suppliers of "massively parallel" computers regularly quote performance figures far higher than those of Cray Research's modestly parallel vector processors. Rather, Seymour Cray's sociotechnical strategy was quietly transformed. The resulting "sociotechnical network," consisting of designers, users, and machines, ultimately had no place for Cray himself. Still, it has to date proved remarkably durable.

The transformation began with the second CRAY-1 to be sold. The first sale was the classic CRAY linkage of raw speed with the needs and resources of a nuclear weapons laboratory (Los Alamos, this time, not Livermore, since Fernbach had committed that laboratory to a second STAR 100), together with the classic Cray confidence. Los Alamos had not budgeted for a new supercomputer before 1977, and by

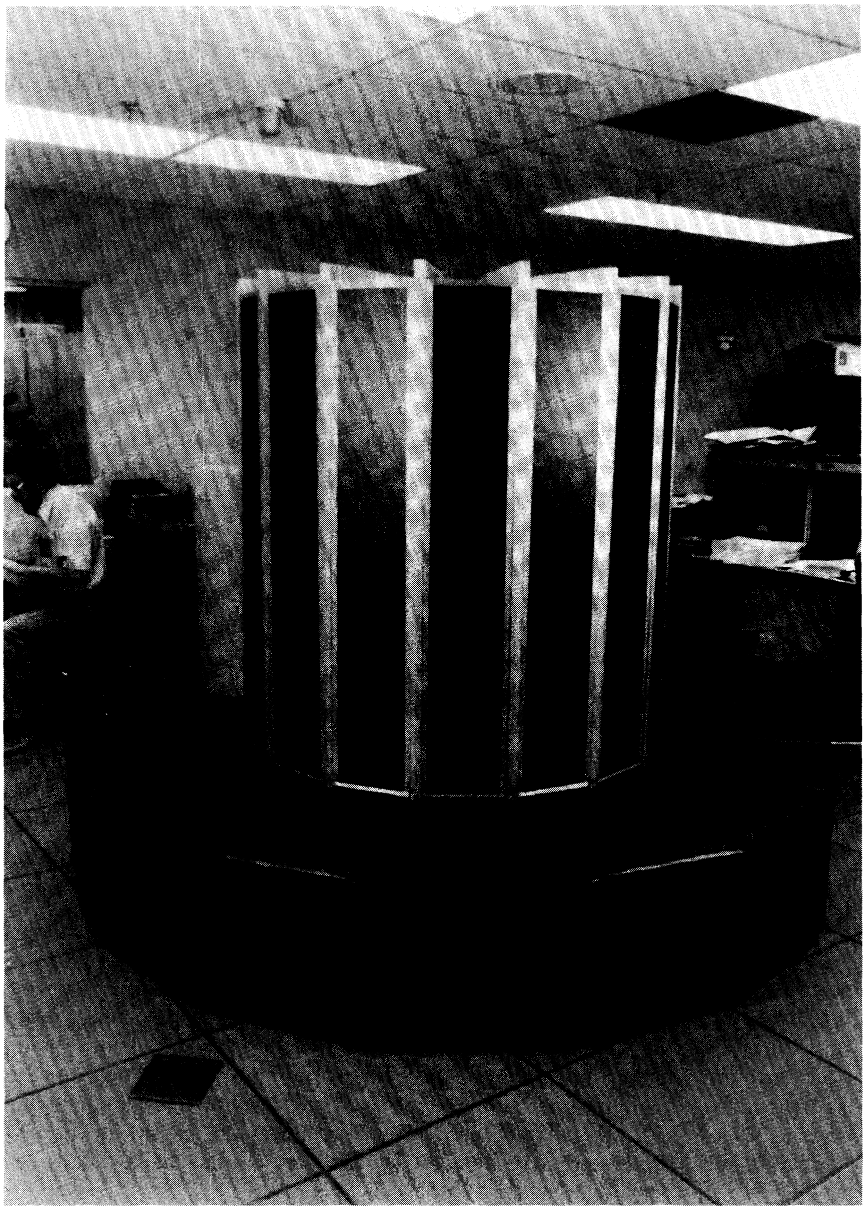

Figure 2. The CRAY-1 computer. (*Source:* David E. Lundstrom, *A Few Good Men from Univac*, M.I.T. Press, Cambridge, Mass., 1987).

the mid-1970s the weapons laboratory's computer acquisition process had become much more bureaucratized than in the 1950s and 1960s. Cray Research, without any sales four years after it had been set up, had to establish itself, and few customers other than Los Alamos would have the interest and the resources to accept a new machine that was almost devoid of software and was incompatible with its existing computers. Cray gambled and offered Los Alamos CRAY-1 Serial 1 on loan. "If the machine hadn't performed, Cray Research wouldn't have continued as a company."[24]

The customer identified for the next CRAY-1, Serial 3,[25] was not a nuclear weapons laboratory, but the National Center for Atmospheric Research. Although attracted by the CRAY-1's speed, the Boulder, Colorado, meteorological bureau refused to buy the machine unless Cray Research supplied the systems software as well.

Cray's strategy of building "the fastest computer in the world" and letting the users worry about software was put to the test. By 1976, the management of Cray Research was no longer solely in Seymour Cray's hands. Cray could raise money from friends, but he was not the man to negotiate details with Wall Street. Another midwesterner, also an electrical engineer, but with a Harvard MBA, had been hired in 1975 as chief financial officer: 34-year-old John Rollwagen, who organized the successful 1976 public flotation of Cray Research. Rollwagen and others concluded that a more accommodating approach to users had to be taken, and committed the company to supplying the National Center for Atmospheric Research with systems software—an operating system and a Fortran compiler—as well as hardware, thereby initiating a major software development effort (located, significantly, in Minneapolis, not Chippewa) to parallel Cray's hardware development. In July 1977 CRAY-1, Serial 3 was shipped to Boulder, and by the end of the year three more computers had been sold. The company made its first net profit—$2 million—and John Rollwagen became president and chief executive officer.[26]

Although the sales of the CRAY-1 were just beginning, thoughts within Cray Research had already begun to turn to what to do next. There was discussion in the company about whether or not to move into the "low end," that is, to make smaller machines (what in the 1980s would come to be called minisupercomputers), derived from the CRAY-1, that would be cheaper and aim at a larger market. Seymour Cray disagreed, explicitly turning his back on the lure of growth. In the 1978 Annual Report he wrote:

> I would rather use the corporate resources to explore and develop newer and more unique computing equipment. Such a course will keep the Company in a position of providing advanced equipment to a small customer base in an area where no other manufacturer offers competitive products. This course also tends to limit the rate of growth of the company by moving out of market areas when competitive equipments begin to impact our sales. I think it is in the long-term best interests of the stockholders to limit growth in this manner and maintain a good profit margin on a smaller sales base.[27]

As always, Cray put his technological effort where his mouth was, starting work in 1978 on the CRAY-2, with the goal of a sixfold increase in speed over the CRAY-1—nearing the tantalizing target of the gigaflop, a thousand million floating-point arithmetic operations per second.

No one defied the founder by beginning work on a minisupercomputer. Nevertheless, Cray Research's development effort was not restricted to the CRAY-2, and extensive efforts were made to make the existing CRAY-1 attractive for a broader range of customers. The CRAY 1S, announced in 1979, kept to the founder's intention by offering improved performance: an input–output subsystem to remove bottlenecks that slowed the original CRAY-1, and a larger memory. In 1982, however, it was followed by the CRAY-1M, which offered the same performance as the 1S, but, through the use of a different, less expensive component technology, at a significantly lower price—$4 to $7 million, rather than the $8.5 to $13.3 million of the machines in the 1S series. The CRAY-1M was no "mini," but it could be seen as a step in that direction.

Changing the hardware on offer was only one aspect of the new, different strategy that began to develop at Cray Research. Systematically, different categories of users (not just weapons laboratories and weather bureaus, important though those remained) were cultivated. Their needs were explored and, where necessary, technology altered or developed to meet them. The oil industry was first, with four years of cultivation between first contacts and the first sale, in 1980. That industry's most relevant computational need was for the processing of seismic data, in quantities far exceeding even the weapons laboratories or weather bureaus. A specially developed additional memory—the Solid-State Device (SSD)—helped, but a crucial (and symbolically significant) step was Cray Research's investment in the development of a link between IBM magnetic tape equipment and a Cray supercomputer, together with software to handle tapes that had suffered physically from the rigors of oil exploration.[28]

Aerospace, the automotive industry, and chemicals were further targets for successive waves of focused cultivation and "network building." Although physical devices had sometimes to be developed to cement links, software development was far more important. It was no longer sufficient, as it had been at the sale of Serial 3, to supply an operating system and a Fortran compiler. Compilers for other programming languages were developed, and particular attention was devoted to "vectorizing" compilers, which would enable the users of Cray supercomputers to take advantage of their machines' vector speed without having to invest large amounts of time in rewriting programs (as the Livermore users of the STAR-100 had had to do). Cray Research even took upon itself the conversion for use on its machines of specific computer packages that were important in areas being targeted for sales, such as the PAMCRASH automobile crash simulation program. By 1985, 200 major applications packages had been converted to run on Cray machines.[29]

The result of all this effort was that during the 1980s Cray Research's expenditure on software development came to equal that of hardware development. Seymour Cray should not be thought of as opposed to this. Ever flexible about the means to achieve the goal of speed, if not about the goal itself, he could even be seen to be in the lead of this effort, abandoning, for the CRAY-2, Cray Research's own Cray Operating System (COS), and moving instead to UNIX, which was rapidly becoming a standard operating system, and which was much better at handling time-sharing and interactive computing than Cray Research's original system. Cray's flavor of UNIX was christened UNICOS.

But his readiness to make the shift also indicated a widening chasm. The change was in part sensible because the CRAY-2 differed so radically in architecture from the CRAY-1 that converting the Cray Operating System was scarcely any easier than shifting to UNIX. Once again, Seymour Cray had started with a clean sheet of paper in his search for speed. But with the company paying closer attention to the priorities of users disinclined to invest in rewriting programs, and with the growing Cray Research investment in software, the clean sheet of paper was beginning to seem a problem, not an advantage, to others in the company.[30]

Here, furthermore, the unpredictability of trying to implement a strategy in an only partially tractable world, rather than rational choice between strategies, played a decisive role, and added to the centrifugal forces already pushing Seymour Cray away from the center of the company he had founded. To gain his desired speed increase, Cray was pursuing an ambitious design that combined new, faster chips (developed by Motorola and Fairchild in cooperation with Cray), multiple central processing, and a processor architecture significantly different from that of the CRAY-1. Despite their respect for Cray, others in the firm questioned whether he could succeed in all these innovations simultaneously, and difficulties and delays in the CRAY-2 project reinforced their fears.

Les Davis, Cray's chief engineer, began the process by which an alternative to the CRAY-2 emerged. He suggested that these new chips be used in a multiprocessor design, but one in which the basic processor was essentially the same as in the CRAY-1. Such a machine would enjoy greater software compatibility with the CRAY-1, and even if no product emerged, it would allow the increasingly important software developers to experiment with parallel processing in advance of the CRAY-2. On Davis's instigation a second design team was formed, headed by a young Taiwan-born engineer, Steve Chen, to pursue a low-level, cheap effort along these lines.

The effort greatly surpassed the hopes of its instigator. John Rollwagen, concerned that Cray Research's future was being staked too exclusively on the CRAY-2, decided to sell the Davis–Chen machine as a product. The CRAY X-MP, announced in 1982, offered up to five times the performance of the CRAY-1S, nearly as much as the CRAY-2's as yet unmet goal, but with the advantage of substantial software compatibility with the CRAY-1.[31] The X-MP fitted its niche beautifully, in both a physical sense (the use of more advanced chips meant that multiple processors could be fitted into a cabinet very similar to that of the CRAY-1) and a commercial sense, given the importance of software to Cray Research and its users. It became more popular than any previous supercomputer, with almost 160 sales of different versions by the end of 1989.[32]

Seymour Cray managed to snatch partial success for the CRAY-2 from what was rapidly beginning to look like a path to oblivion: an advanced design route to a new computer offering little, if any, speed increase on the X-MP, which was fast becoming an established product. He managed to differentiate the CRAY-2 from the X-MP by using slow but relatively inexpensive chips to offer a massive memory of up to 256 megawords, two orders of magnitude more than existing machines at the time. He compensated for this slowness of this massive memory by attaching small, fast memories to each of the CRAY-2's four processors. This tactic worked only partially: very careful management of memory resources by users of the CRAY-2 is needed to prevent the slow main memory from becoming a bottleneck. But a sufficient number of users wanted a massive memory (by the end of 1989 twenty-four CRAY-2s had been sold)[32] for there to be reasons other than sentiment for Cray Research to market its founder's design.

The most immediate reason for the CRAY-2's dangerous delay (it came on the market only in 1985, three years after the X-MP) was problems with cooling, and the various approaches that were taken to solve these problems are a further interesting indicator of the specificity of Cray's preferred sociotechnical style. The CRAY-2's components were packaged even more closely than in previous machines, and Cray, despite repeated efforts, could not make his existing approach to cooling, based on circulating Freon refrigerant, work well enough:

> You don't know you are going down the wrong road on a design until you have invested six months or a year in it. I had about three of those false starts on the CRAY-2. The cooling mechanism in those designs didn't work.[33]

In a "last desperate attempt" Cray tried a completely new cooling approach in which the whole machine would be immersed in a cooling liquid that, by its forced circulation, would remove the heat produced. When he proposed the scheme nobody took it seriously. "[E]veryone on the project laughed, in fact they rolled in the aisles. Because everybody knew the boards would swell up and it just wouldn't work."[34] Indeed, liquid immersion cooling had been tried before, and a variety of known coolants had been investigated, all of which in a short time damaged the printed circuit boards. But Seymour Cray took a different approach, selecting liquids not primarily because of their cooling properties, but on the basis of their known inertness. One of the liquids he subsequently tried was a substance that was used in artificial blood. This finally worked, and allowed Cray to build a machine that was very densely packed but that could still be cooled[35] (see Fig. 3).

An Amicable Parting

The great success of the X-MP, and partial success of the CRAY-2, contributed to Cray Research's continuing growth. Despite its specific market, the company's sales began to nudge it into the ranks of the dozen or so leading computer suppliers in the world, albeit still far behind the giant IBM. That success, however, masked the deepening of the divide between its founder and the bulk of the company he had founded. In 1989, four years after the CRAY-2 went on the market, Seymour Cray left Cray Research.

During the latter part of the 1980s, Seymour Cray's strategy, and that dominant in Cray Research, had continued to diverge. The Cray Research strategy was to build on the success of the X-MP and the growing user base that made that success possible, seeking systematically to develop new fields of application and strengthening relations with existing customers. The purchaser of an X-MP (or the improved, but compatible, Y-MP) is now buying not raw speed, but access to extensive software resources and services. Cray Research will help customers new to supercomputing to plan their installations, and provide on-site support for the life of the installation. The firm guarantees that should a Cray Research supercomputer fail anywhere in the world, a Cray engineer will be on site within two hours to solve the problem.

Far from keeping users at arm's length, Cray Research seeks to bind them ever more tightly to the company. A Cray User Group has been set up, which holds meetings for all users of Cray supercomputers every year. These meetings are attended by

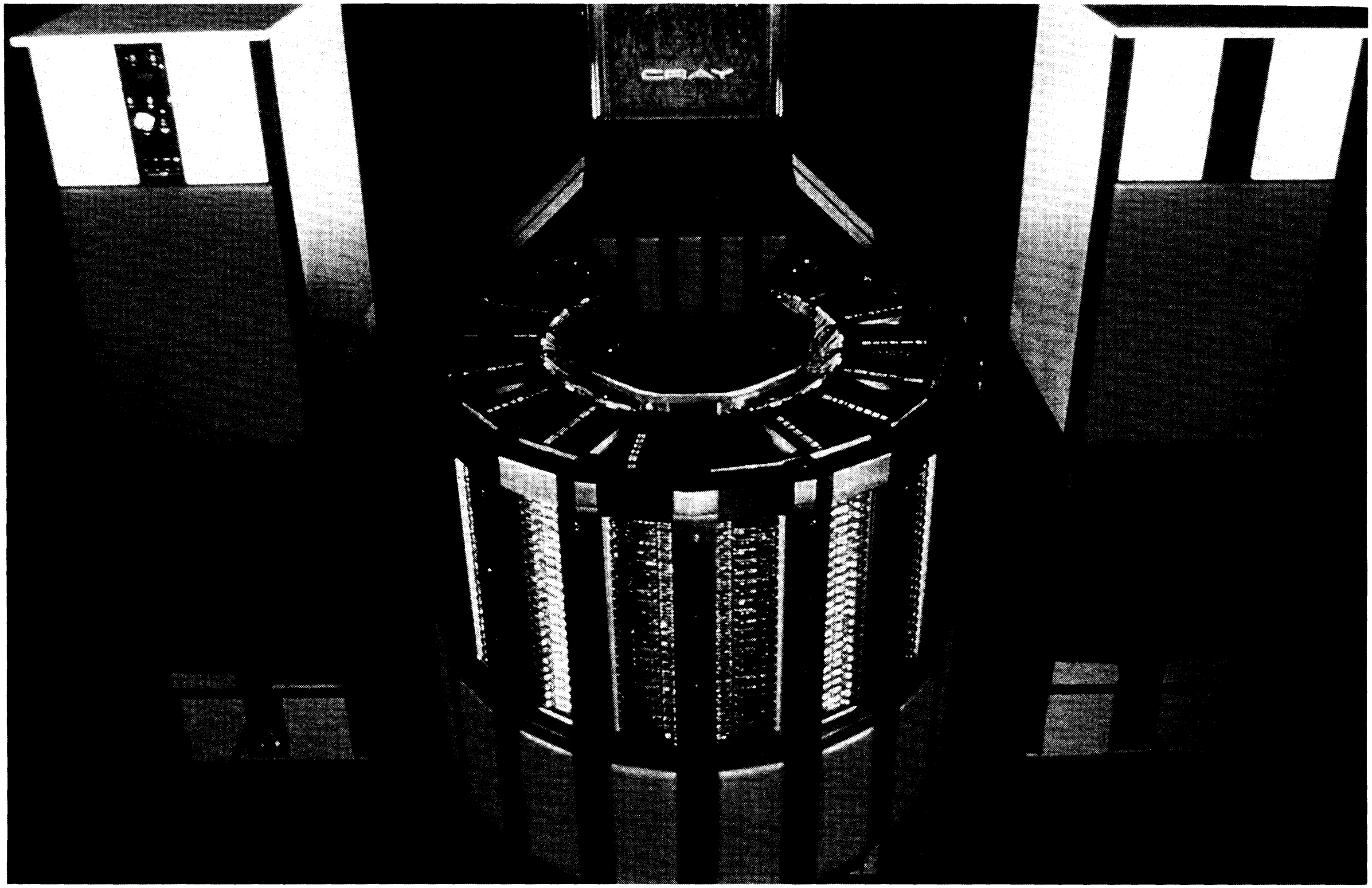

Figure 3. The CRAY-2 computer. (Photograph by Paul Shambroom; reprinted with permission from Cray Research, Inc.)

Cray representatives, enabling the company to respond quickly to any problems or desires that may emerge from these meetings. Cray Research is also paying increasing attention to links with other suppliers, as well as to users. The announcement in 1990 of a "network supercomputing" strategy for Cray Research made this explicit: the supercomputer is no longer to be seen as an artifact standing on its own, but as a central part of complex companywide networks, many parts of which will be supplied by companies other than Cray.[36]

If this strategy begins to sound a little like the statements by Norris that were instrumental in Cray leaving Control Data, one crucial difference should be emphasized: Cray Research is still very concerned with speed. The CRAY Y-MP, which came onto the market in 1988, is about two to three times faster than the X-MP, though comparison is difficult because both machines have been made available in many different processor numbers and memory sizes to suit different kinds of customers. The successor to the Y-MP, generally referred to as the C-90, will be a further expression of this technological approach: it will be compatible with its predecessors, and of broadly similar design, but will be significantly faster, embodying improved component technologies, more central processors (16 rather than the 8 of the top-of-the-range Y-MP), and a larger memory.[37]

Simultaneously, however, the heretical step "down" toward the minisupercomputer has also been made. In 1990, Cray Research announced an air-cooled version of the Y-MP series: slower, but, at $2.2 million and up, significantly cheaper, though still more expensive than most minisupers. Cray Research launched the Y-MP/EL late in 1991, configured with up to four CPUs, with each CPU costing $300,000. Each CPU costs about 1.5 percent of the price, in real terms, of the original CRAY-1, launched in 1976, but with the same performance.[38]

Seymour Cray's commitment to speed, on the other hand, remained much more obvious. His goal for the CRAY-3 is a twelvefold increase in speed over the CRAY-2, to sixteen thousand million floating-point arithmetic operations per second.[39] The machine will be compatible with the CRAY-2, but it will not be compatible with the X-MP, Y-MP, or C-90.

In one crucial respect, however, the CRAY-3 is a departure from Cray's previous strategy. A fourfold improvement in speed over the CRAY-2 is expected to come from using 16 processors compared to four on the CRAY-2, but that leaves a factor of at least 3 to come from faster components. As we have seen, Cray's preference in all previous machines was to remain well within the state of the art in component technology—to avoid both risk and also a complex division of labor. But, for the CRAY-3, he concluded that the state of the art would not sustain the sought-for increase in component speed, and so for the CRAY-3 he has taken the "high-technology approach" that he eschewed for the CRAY-2.

Nor is the form taken by this approach the silicon very large-scale integration (VLSI) path, which, though new to Cray, had many analogues in the wider computer industry. Instead, Seymour Cray became the first computer designer to commit himself to using processor chips made out of gallium arsenide rather than silicon. Gallium arsenide had long been discussed as a faster substitute for silicon (Cray himself had investigated but rejected it for the CRAY-2), but there were known to be daunting difficulties: in the number of circuits that could be implemented on a gallium arsenide chip, in manufacturing the chips, and in their reliability. Around the computer industry, the prevalent jibe was that "gallium arsenide is the technology of the future, always has been, always will be."[40]

With the CRAY-3, Seymour Cray gambled that the future was about to arrive, and that he could manage a more complex process of technological development, involving not just his own design team but also groups outside the company developing gallium arsenide components for the project. Reports suggest that he has been successful. The CRAY-3 project has, however, met with serious delays. The problems have arisen, it seems, not from the gallium arsenide, but from the continuing miniaturization needed to sustain ever-increasing speed. The CRAY-3 is planned to operate with a 2-nanosecond clock period, handling one scalar instruction every clock period. For this to be possible, given the finite speed of electrical signals, the maximum allowable wire length is 16 inches. So the CRAY-3 modules are unprecedentedly small for a supercomputer. Seymour Cray's design calls for cramming 1,024 chips into a module measuring just 4 by 4 by 1/4 inches—so delicate a job that special robots will be needed. Such a module has nine functional layers that are interconnected vertically with up to 15,000 twist pin jumpers that are inserted through holes in each layer. The whole CRAY-3, meaning 16 processors and 512 Mwords of memory, will use 336 of those modules and will be packed in a box less than a cubic foot in size.[41]

Sustaining both the CRAY-3 and C-90 projects, together with the growing range of other development activities (especially in software) that Cray Research was becoming committed to, began to place strains on the company. During the first quarter of 1989, research and development expenses were about 35 percent higher than in the first quarter of the preceding year, even after the cancellation of another ambitious project, the MP project led by Steve Chen. At the same time, Cray Research's annual growth of revenues had slowed from about 50 percent in the early 1980s to about 10 percent in 1988.[42] The share price of Cray Research, long the darling of Wall Street, was beginning to slump drastically. Either the CRAY-3 or C-90 had to go.

Which was it to be? With the delays in the CRAY-3, it and the C-90 were expected to appear at roughly the same time, 1990/1991, and were going to be similar in speed and memory sizes. The main difference was in compatibility. The CRAY-3 would be compatible with the CRAY-2, of which about 25 machines had been sold. The C-90 would be compatible with the X-MP and Y-MP line, with an installed base of around two hundred. The logic was clear. The C-90 had to be preserved and the CRAY-3 canceled.

Yet it could not be as simple as that. John Rollwagen knew that "not choosing Seymour's machine would have torn the company apart." "After six months mulling over alternatives," Rollwagen proposed to Cray yet another "amicable divorce."[43] A new company would be incorporated to undertake the further development of the CRAY-3. The new company would at first be a wholly owned subsidiary of Cray Research, with Cray Research transferring to it $50 million worth of facilities and up to $100 million in operating funds over two years. For Cray Research, with a revenue of $750 million a year, this was a lot of money. However, Cray Research's shareholders would be rewarded by receiving shares in the new company on a tax-free basis, Cray

Research would now be able to concentrate its efforts and resources on the one project, and failure of the CRAY-3 would not endanger Cray Research's existence.

While making it clear that the split was not his idea, Seymour Cray agreed to it: "I don't mind this role. I kind of like starting over."[44] Once again, he was working for a small start-up company, this time called the Cray Computer Corporation. This time, however, it was not in Chippewa. Prior to the split, the CRAY-3 team had moved to Colorado Springs, in the foothills of the Rocky Mountains. Nor is the world the same as in 1957, or in 1972. In 1972, in particular, Cray Research's commitment to building the world's fastest computer was unique. It had no direct competitor, the less than wholehearted effort at Control Data aside. At the start of the 1990s, Cray Research, the Japanese supercomputer companies, and Steve Chen (who left Cray Research when his MP project was canceled, to form a new firm backed by IBM) are all pursuing speed, as are a number of other companies that have developed or are developing machines of a quite different structure to those characteristic of "mainstream supercomputing."

Growing Networks Limit Room to Move

In this newly competitive marketplace, Cray Research still enjoys a dominant position, not through speed alone, but more through the diverse, entrenched links it has built with users. Over the years a network of relationships has evolved in which the technology under development binds users and manufacturers together. For both, this network limits their room to move, unless they are prepared to put at stake what has been built up. In the 1960s and early 1970s relationships between users and manufacturers of supercomputers were close only in the period of the actual sales. Subsequently, to put it baldly, the two parted with a "see you around in five years" (when a new supercomputer would be ready for sale). Especially during the early 1980s, with the broadening of the customer base, relationships intensified considerably. As noted earlier, Cray Research opened new markets by having extensive talks with new customers, trying to understand their problems, and subsequently trying to solve these problems by undertaking the development of specific hardware and software. Once a deal was made, Cray Research provided its customers with a variety of support services, from preinstallation site planning to on-site support for the life of the installation.

Concerning software, Cray Research established applications groups within the company that deal with the various fields of applications, keep in touch with the people in the field, try to develop new applications, or to develop solutions to emerging problems. Taking on software development implied not only a shift in its activities, but also an extension of the network of relationships surrounding supercomputers. Rather than starting software applications development from scratch, Cray Research developed relationships with a variety of third-party software vendors in order to optimize the existing codes for Cray machines, thus creating a triangular relationship with these vendors and (potential) users of Cray supercomputers. As a result of such efforts, over 200 major application packages were available on Cray systems by 1985.[45]

In order not to alienate the existing customer base, Cray Research is compelled to continue building on the existing technology, that is, both hardware and software. This clearly limits its freedom to maneuver; only Seymour Cray himself, by virtue of his high prestige, can afford to "start with a clean sheet of paper," and even his sheet is no longer as clean as it used to be. But while the sociotechnical network surrounding supercomputing limits innovation, it is also a formidable barrier to competition from outsiders.

Network Relations as a Shield against Competitors

Cray Research currently enjoys a near monopoly in terms of actual sales of supercomputers. This is not because there is no competition, but, as we just said, because it has built a network of relationships with a substantial range of customers that is based on more than just a fast computer. This network encompasses advanced software as well as advanced hardware, and includes Cray Research as the manufacturer of supercomputers, a wide variety of users, and a large number of software vendors.

In the early 1980s the only serious competitor to Cray Research was CDC. For certain applications the performance of CDC's CYBER-205 compared favorably to that of the CRAY-1, which enabled CDC to seize about 20 percent of the supercomputer market. But after Cray Research offered the X-MP and the CRAY-2, not many 205s were sold. In September 1983, CDC founded a separate company, ETA Systems Inc., to develop CDC's next-generation supercomputer. ETA announced the target of 10 gigaflops for their ETA^{10} machine, that is, an order of magnitude faster than the CRAY-2 and X-MP, then the fastest machines. The architecture of the 205 was taken as a point of departure. Since this, in its turn, was based on that of the STAR-100, program compatibility would be maintained with the software developed for the STAR and CYBER-205.

The increase in speed of the ETA^{10} over the CYBER-205 was to come from using faster circuits and incorporating eight processors. This goal was achieved by using complementary metal oxide semiconductor (CMOS) chip technology rather than the bipolar technology traditional in supercomputing. The gap in speed between CMOS and the faster bipolar technology had been narrowing and it also appeared that CMOS could be speeded up by a factor of about 2 by cooling it down to the temperature of liquid nitrogen, or about −200°C. The ETA^{10} thus became the first supercomputer to use cryogenic cooling.

In a technological sense, the ETA^{10} was a considerable success. Late in 1986, just over three years after the incorporation of ETA Systems, the first machine was sold, albeit in the rawest of states, to Florida State University. Independent analysts expected that "(w)ith this interesting development one can be assured that the CYBER 205 architecture will . . . be of lasting interest."[46] They were wrong. The immediate problem was the financial difficulties of ETA's parent company, CDC. CDC therefore started to sell or liquidate divisions or subsidiaries that were losing money, one of which was ETA. In April 1989 ETA was closed brutally and painfully with total losses of $490 million.[47]

As a result, Cray Research had lost one potential competitor. During the 1980s, however, a number of other competitors had appeared on the scene. First was Japan. In October 1981, at a conference in Tokyo, the Japanese Ministry of Trade and Industry presented a plan to leapfrog current technology and secure the world lead in advanced computing by the late 1990s. One part of the plan, the National Superspeed Computer Project, involved spending an estimated $200 million over eight years to develop a supercomputer capable of executing 10 gigaflops, or 65 times faster than the CRAY-1, the fastest machine at the time.[48] In 1985, the Nippon Electric Company (NEC) introduced the SX-2, with an announced peak speed of 1.3 gigaflops, which took direct aim at the CRAY-2, which also appeared that year. In 1985, NEC became the first foreign supercomputer maker to install a machine on U.S. soil. The SX-2's buyer was the Houston Area Research Centre (HARC), a consortium of four Texas universities. But further sales of NEC, Hitachi, and Fujitsu supercomputers outside Japan remained very limited. Although there has been strong political resistance in the United States to the purchase of Japanese supercomputers, the Japanese companies also have not been able to supply the range of software that is available for the Cray machines. Gigaflops were no longer all that counted. As Cray president John Rollwagen put it: "(A) supercomputer's peak speed is comparable to the speed of the spinning wheels of a car on blocks: The motion is meaningless unless useful work is being done."[49] Japanese companies are trying to fill the software gap and also continue developing faster machines. In April 1990, NEC announced that it would shortly introduce a computer with a top speed of 22 gigaflops.[50]

A second source of new competition has been IBM. The growth of the supercomputing market (a growth, as we have seen, largely stimulated by deliberate actions of Cray Research) persuaded IBM to reconsider its strategy of leaving this "niche" to Cray Research. Like Hitachi and Fujitsu, IBM achieved supercomputer performance by adding vector processors to a basic mainframe design, the IBM 3090. The IBM philosophy is, however, somewhat different. The Japanese manufacturers achieve high theoretical performance by emphasizing very high levels of vectorization, while IBM has focused upon more modest levels of vectorization, accompanied by high scalar speed. Cycle time for the initial 3090 was 18.5 nanoseconds.[51]

This means that, in scalar mode, this IBM machine is a factor of 2 to 3 slower than the Cray supercomputers, and therefore IBM cannot compete with Cray Research for customers that want the greatest performance per se. But there is a gray area of customers that are interested in "almost the fastest computer in the world" with all the advantages of IBM compatibility. For them, the 0.5-gigaflop performance of a six-processor IBM-3090 with vector facilities[52] may be very attractive. Like Cray Research, IBM has not concentrated on hardware alone. It, too, has looked at specific groups of users, analyzed their problems, and subsequently developed software dedicated to their applications. As a result, in terms of sheer numbers, IBM has probably sold the most vector computers.[53] This policy, together with the fast growth of much cheaper "minisupercomputers" from a variety of vendors, may well limit Cray Research's possibilities for growth into the "near supercomputer" market.

At the same time, by teaming up with Steve Chen after he left Cray Research and formed Supercomputer Systems, Inc., IBM has also left itself the option of trying to get a stake in the market for the fastest machines. IBM partly funds development at Chen's firm in exchange for access to its technology. Little has been disclosed about the machine Chen is working on, but it is believed to be a radically parallel (but probably not massively parallel) design. The performance goal is set very aggressively at the teraflops range (1000 gigaflops), but the machine is not expected to be available until 1995. There are reports that Chen's machine will be significantly more expensive than today's most costly supercomputers, costing perhaps as much as $60 million.[54]

It is striking how most attempts to attack Cray Research's position are not all-out attempts to develop the fastest machine in the world (like Seymour Cray attempted with the 6600, the CRAY-1, and the CRAY-2), but are built on previous work and an established customer base. CDC took the 205 as a point of departure, IBM built further on its 3090, and the Japanese companies generally sought compatibility with IBM. Even Seymour Cray will now make the CRAY-3 compatible with the CRAY-2. Only Steve Chen is reportedly working on a "revolutionary" design. Consequently, no user can simply migrate from a Cray machine to one of the competitors without investing a considerable amount of time and money in problems of conversion.

As a result, Cray Research still dominates the high end of the market. ETA was dead before it had a chance to establish itself; IBM has basically only succeeded in "stretching" its existing market; and the Japanese companies have sold only in modest numbers outside Japan, and then mainly to scientific institutions to which they sell at a large discount. By contrast, Cray has gained a strong foothold in the Japanese private sector, with sales to automotive, electronics, chemical, and aerospace firms.

Aside from the entry of Japan, the most discussed form of competition to the Cray machines is "massively parallel processing." Instead of using a small number of very fast processors, as in a "classic supercomputer," massive parallelism uses a very large number (up to tens of thousands) of relatively slow microprocessors. The difficulty, of course, is that to make such numbers of processors work harmoniously together on a single job poses a tremendous software problem. Some massively parallel computers have achieved top speeds on particular problems in excess of the speed of classic supercomputers, but parallelizing compilers to permit ready generalization of this performance are still far from being realized. Some people therefore believe that the future of massive parallelism lies not in "general-purpose" computing, but rather as dedicated "add-ons" to more general-purpose machines. Certainly, massive parallelism shows that sheer speed is now far from sufficient for supercomputing success.

The worldwide market for massively parallel machines is estimated to be about $200 million in 1990, and is still heavily subsidized by government bodies. This is, however, about one-fourth of Cray Research's annual turnover. Cray Research has its own massive parallelism project, but rather than aiming to supplant mainstream supercomputing, Cray Research sees massive parallelism as an enhancement of general-purpose supercomputing and pursues integration of the two different architectural approaches. The firm hopes that users will do their general-purpose supercomputing as they always have, while improving the performance of their highly parallel, specialized jobs on a massively parallel system designed and manufactured by Cray Research.[55]

National Interests

Competition in the supercomputer arena is not only an intercompany issue, it also has international dimensions. In the 1980s, the use and production of supercomputers became a national issue in the United States. Supercomputer use was portrayed as being of vital importance to the country's national security and to the economic competition with Europe and Japan. It was argued that U.S. scientists and engineers had insufficient access to supercomputing facilities. Successive panels, composed of scientists from many fields as well as computer specialists, examined the situation and issued reports deploring it. One theme was stressed repeatedly: the United States could lose its leadership in technology because its scientists lacked adequate access to this fundamental tool of advanced research. Admonitory fingers were pointed toward Japan and Europe, where supercomputers were being installed and used at a growing rate. The picture was sketched of an imminent and perhaps irreversible loss of America's international competitiveness.

Petitioners from universities looked primarily to Congress and a variety of sources of federal funds for academic research. Requests for support were directed to the National Institutes of Health, the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), the Department of Energy, the Defense Advanced Research Projects Agency (DARPA), other parts of the Department of Defense, and smaller federal agencies. During the mid-1980s, an increasing number of these projects found favor in Congress and among the budget makers of these agencies. Ears that did not heed earlier appeals emphasizing scientific need became responsive to arguments about preserving America's prestige, American jobs, and American technological leadership. Legislators and administrators frustrated by the difficulty of finding solutions to problems of industrial policy, protectionism, exchange rates, and the cultural differences between Japanese and American factories discovered a talisman. It wasn't easy to fix all those other things, but it was possible to do something about the supercomputer shortage.[56]

The NSF decided to devote $200 million for the support of the supercomputing initiative over a five-year period, beginning in fiscal year 1985. It started by buying time at existing facilities and moved swiftly to select and fund five new supercomputer centers. All over North America, states and universities that did not secure formal federal support—or preferred to go their own way—also launched supercomputer programs. Around 1987 more than one hundred North American institutions participated in supercomputing.[57]

Another initiative was taken by the Office of Science and Technology (OSTP) when it launched the High-Performance Computing Initiative. This was intended to maintain the U.S. edge by focusing the research advantage in high-performance computing toward applications with high value to the economy and national security. The OSTP identified a wide range of scientific and engineering problems that it called "Grand Challenges." A Grand Challenge would be a fundamental problem with potentially broad economic, political, or scientific impact, that could be solved by applying supercomputer resources. While Grand Challenges usually were already being addressed by existing resources, it was argued that progress was limited because of the state of both hardware and software technology.[58]

Training in supercomputing also expanded rapidly during the period. The NSF-supported centers and many other facilities offered one-day introductory workshops for potential users, more thorough seminars, summer institutes for graduate students and other researchers, and additional opportunities. The Department of Energy offered summer institutes in supercomputing for promising high school students; other courses and summer internships were available at supercomputing centers in the national laboratories. Supercomputing facilities at universities offered similar opportunities.[59]

Besides access to supercomputers by scientists and engineers, another national issue was the competitiveness of the U.S. supercomputer industry. First to cause alarm was the Japanese government-funded program to leapfrog state-of-the-art technology. These fears were aggravated when Japanese companies started to offer supercomputers on the U.S. market in the mid-1980s. Articles in the *New York Times* and the *Washington Post* suggested that the U.S. supercomputer industry was on the brink of decline.

IEEE's Scientific Supercomputer Committee, chaired by Sidney Fernbach, former head of Livermore's computing division, recommended in a 1983 report that the government should commit itself to buying by 1987 a dozen supercomputers specified to have significantly increased performance. George Michael of Los Alamos National Laboratory doubted that this would help U.S. industry, since "Cray and CDC aren't positioned well for hotshot competition because they're not vertically integrated like the Japanese companies."[60]

The supercomputer was only one of several commercial products on which the United States felt it was losing ground to the Japanese. The U.S. government and various industry groups claimed that Japan was using unfair trade practices in a number of areas, ranging from semiconductors to construction contracts. These allegations spread to include supercomputing when NEC sold an SX-2 to the Houston Consortium. This U.S.–Japanese "trade war" cumulated in 1987 in the imposition of (now lifted) trade sanctions on Japan.

Interestingly, Cray Research did not welcome these sanctions. It had to compete with the Japanese companies everywhere in the world, not just in Japan and the United States. In that competition U.S. export licensing procedures were a handicap to Cray Research. Furthermore, Cray Research bought components from Japan and also sold computers to Japan (considerably more than the Japanese sold in the United States). The paradox is that it was not until Fujitsu and Hitachi entered the supercomputer market that Cray Research started to sell supercomputers in Japan. The firm was convinced that it could create additional business by having more alternatives, more competitors, more activity.[61]

From Megaflops to Total Solutions

In summarizing the historical sketch presented in this chapter we can identify three important changes in the dynamics of the supercomputer world. These are:

- It is no longer sheer performance that encourages the development of next machines, but rather performance combined with compatibility.
- Software has become increasingly important, also leading to the involvement of third-party vendors.
- The user base has expanded considerably and, moreover, has changed in character, with users from a larger variety of sectors and different levels of in-house computing expertise.

J. Neil Weintraut, a computer industry analyst, characterized the situation of the supercomputer market in the late 1980s as follows: "Cray Research will be facing increasing competition in the 1990s from Japanese firms, IBM, and Cray Computer at a time when the supercomputer market's annual growth has slowed from about 50 percent [in the early 1980s] to about 10 percent [in 1988]."[62] Other analyses doubted whether there would be any place for the United States in the world market in the 1990s because of the bad state that U.S. supercomputing was in. ETA had died in 1988 and Cray Research's revenues were leveling off. The spinning off of Cray Computer was not seen as a good sign either.

To assess these arguments, let us take a closer look at what happened. We have already described a number of attempts to build supercomputers faster than those of Cray Research. The end products of several of those attempts have claimed success over the past decade, but in the early 1990s Cray Research still dominates the supercomputing market. In the 1980s, too, U.S. supercomputing was perceived to be at the brink of collapse, which everyone feared would be caused by Japanese competition. Yet the 1990s see the Japanese companies being far from dominant, and the trade sanctions against Japan (never welcomed by Cray Research) have been dropped. Finally, in the 1980s it was widely predicted that massive parallelism would rapidly replace conventional supercomputing. This, too, has not happened.

All these failed predictions underestimated the sources of stability in high performance. They focused on either "the social" (e.g., Japanese competition) or "the technical" (the theoretical advantages of massive parallelism), without grasping the interrelations of the social and the technical, interrelations that give the field its stability.

One source of stability is that companies with an established customer base first of all seek not to alienate their current users and to build follow-on machines that are compatible with their predecessors. Cray Research, for example, has developed an evolutionary product line starting with the CRAY-1 that, via the 1S, 1M, X-MP, X-MP EA, and Y-MP, was continued up till its most recent undertaking, the C-90. The commitment of the company to this line became most evident when, in the second half of the 1980s, strains increased on research and development funds. As a result, Chen's MP project was killed and, when that did not solve the problem, Seymour Cray's CRAY-3 was spun off.

The CYBER-205 and ETA engineers did not change the basic architecture of their machines either. When they argued, in connection with the ETA, that a new operating system would be needed, CDC management forced them to port the old CYBER operating system to the ETA^{10}. They were allowed, however, to start a low-level effort to port UNIX to the new machine.[63] The designers of the vector box at IBM could get no architectural relief from the 3090's basic architecture, while the Japanese took their existing mainframes as a starting point. Even Seymour Cray modified his "clean sheet of paper" approach and has designed the CRAY-3 to be compatible with the CRAY-2.

This implies that when users change vendors they have to modify their programs while, when they buy follow-ons from the same vendor, they can run them again immediately (although some conversion effort may lead to additional speed-up by making use of novel features on the new machine). Therefore, when performance and price of different machines are comparable, users are not likely to turn to a different vendor.

The quoted performance figures for supercomputers currently available (or near availability) are all in the same ball park. Although quoted peak performances may vary, such values have limited meaning in practice. Performance often depends on the type of code run and how well the code is optimized to the various architectures. So users, when they initially buy a supercomputer, choose the vendor and architecture that offers the best solution to the user's specific requirements. With vendors offering compatible follow-on machines, and no vendor offering a completely decisive general speed advantage, users have few incentives to change vendors.

The phrase "user specific requirements" is key in the preceding paragraph, since its meaning has changed considerably over time. Initially, with only a small number of highly sophisticated users (particularly nuclear weapons laboratories), it predominantly meant more megaflops. Through the 1980s, however, with the increase in commercial customers, it came to refer to "total solutions," to imply fast hardware, software, and support. Concerning hardware, it not only referred to the number of megaflops, but also to input/output devices, high-speed channels, and high-speed graphics. Concerning software, users expected connectivity, interactivity, ease of use, upgrades, and so forth. In both hard- and software they wanted reliability and support.

In the 1970s Cray Research was able to take over CDC's leading position in the supercomputer arena. This seems to contradict the argument made earlier about the lack of incentives for users to change vendors. That argument is, however, valid only if all major vendors play in the same ball park as regards performance, which they do now, but did not in, say, 1976. Then, looking only at the megaflops, Cray Research and (belatedly) CDC were the only players in the top of the league. What Cray Research then did was to go after "hero-problems" in industry; that is, offering solutions to problems that could not be solved before. When the users got interested, Cray Research worked closely with them and third parties to develop the total solution these users sought. In this respect, Cray Research had the advantage that its products, especially the combination of its architecture and compiler, were more general-purpose than those of CDC and the Japanese supercomputers that came to the market in the 1980s.

When the name of the game was "more megaflops," users were prepared to change vendors on the basis of who offered most of those and had the sophistication and resources to overcome the accompanying problems. That was the basis of Seymour Cray's great successes, the CDC-6600 and CRAY-1. What Cray Research has subsequently done, however, is to change the rules of the game from more megaflops to "total solutions." So catching up with Cray Research in megaflops, or even surpassing it, is no longer enough. IBM and the Japanese companies know this, and offer "total solutions" of their own. But without a decisive speed advantage, there is little reason for users to migrate from one "total solution" to another. For individual companies, the growth will mainly have to come from finding new customers that are not tied too much to one of the competitors. This is a key rationale for Cray Research offering "low-end" products compatible with its top-of-the-line models.

Thus there is little reason to expect sudden dramatic realignments within mainstream supercomputing. But what about change from outside, such as from massive parallelism?

In creating a machine with a parallel architecture, designers must address several critical questions. First, how many processors are necessary? Depending on the type of problem a machine is intended to solve, the number may vary from a few dozen powerful units (the so-called large-grain approach) to thousands or millions of less powerful units (the small-grain approach). For instance, the small-grain approach may be well-suited to problems that are easily broken into many parts, such as artificial intelligence programs requiring the machine to retrieve information quickly from enormous data bases. The second question concerns the fact that the more processors the machine contains, the more critical is the scheme by which they are linked to each other, and the way they are synchronized. Therefore, should each processor be able to communicate with every other one, or only with its nearest neighbors? Similar questions arise with respect to memory: Should each processor be connected to a small amount of distributed memory, or should all the processors share a single large memory? Then there is the issue of memory bandwidth, especially when memory is shared among many processors. Finally, designers and software engineers must decide who should be responsible for dividing a problem into parts to run on the parallel machine. Programming is simplified if the computer is equipped with software that can perform this chore automatically, but chances are the resulting program will be inefficient. Leaving the job to the human programmer might make for greater efficiency, but requires programs to be expensively rewritten for parallel machines.[64]

There is thus a dilemma to be faced, which is well spelt out by former ETA chief Carl Ledbetter: "Parallelism is an extremely difficult technology, the software and management, the overhead management systems, the operating systems and management systems are very complex. Bandwidth issues for instance, where to put the memory in and how to move data; very complex. Nevertheless it is all there is left to hope for [if we want to achieve] another factor of [a] thousand or ten thousand in performance."[65]

This latter claim provides the key to much of the activity in massive parallelism. Until the 1970s, supercomputer developers could simply ride the bandwagon of the semiconductor industry that offered faster and denser chips "for free." The increased strain on packaging and the divergence between the demands of supercomputing and the hugely expanding microprocessor business meant that supercomputer manufacturers started to have to try to drive technological development rather than sit back and wait. This was true, for instance, for ETA, Cray Research, and the Cray Computer Corporation.

The increased strain on packaging was a result of design choices in which, for all of the individual supercomputer vendors, the basic architecture of the processor had become relatively fixed. This implied that further speed-up had to come from using faster technology and using more processors per system. Because of signal propagation delays, using faster clocks implied that the machine had to be shrunk physically. The same was true for the requirement that to increase speed the number of processors needed to be increased. Thus, to a certain extent, on the technological level the name of the game has become "packaging and cooling." Many people, however, see that game reaching its limits. To put it euphemistically, even Seymour Cray has had some trouble incorporating a 2-nanosecond clock and 16 processors into a Cray-3. Many people therefore believe that large increases in speed can only come from massive parallelism.

As a result, although there is now widespread awareness of the difficulties of massive parallelism, most supercomputer centers are experimenting with such architectures, and there has been very considerable state support for their development, particularly from DARPA. Several firms—Thinking Machines, Intel, NCube, and Meiko, for example—have begun to enjoy commercial success with their massively parallel machines (although, as noted before, their combined sales are still dwarfed by those of Cray Research).

It may well be wrong to see an "either/or" future in which either conventional supercomputing or massive parallelism must triumph. Various forms of hybrid systems are likely to develop. The first signs of such systems have already appeared, with a classic supercomputer and a Connection Machine becoming part of the same network and being looked at as a "total system." The specific architecture of the massively parallel component of such a system might make a specific "metacomputer" more suited to some jobs rather than others, and might lead to the development of different "classes of architectures" optimized for different "classes of codes." On the basis of their own requirements, users would then have the options of choosing which parallel architecture(s) to incorporate in the system. There might even be some problems that make it worthwhile to develop a wholly dedicated machine. Some commercial applications might also be so economically valuable that it makes sense to consider spending tens of million of dollars to develop a machine that does nothing but those jobs.

Does this mean that general-purpose supercomputing, to quote Carl Ledbetter, has become an oxymoron? In many ways it does, particularly in connection with production codes that have stabilized. Architectures that are (partly) dedicated may then provide a much more cost-effective way to solve the problems at hand. However, when the user is doing new research and does not know what the applications really are or what they look like, a special-purpose machine is of no use. In that environment, the most appropriate solution is to have the fastest general-purpose machine possible, that is, a computer that does everything "reasonably well." Unless massively parallel machines become much more general-purpose, there will thus be a major role for large vector machines and for moderately parallel machines, that is, a place for the CRAY Y-MP and its successor, for the IBM 3090/VF and its successors, etc. They may not be the fastest machines in the world, but could be expected to be very important machines when used for generic research.[65]

Indeed, the way the supercomputer world has developed implies that the concept of "fastest computer in the world" has lost much of its meaning. Perhaps classic supercomputers will become faster by one, two, or even more orders of magnitude, but here too the bulk of the speed-up is expected to come from increasing the number of processors. Therefore, software development for these machines will increasingly have to address multiprocessing issues akin to those of massive parallelism. Faster hardware for them now seems increasingly to mean exotic technologies, and that means increased cost. The gallium arsenide, ultradensely packed CRAY-3 is quoted as likely to cost on the order of $30 million, while Steve Chen's S-1 might even cost double that. A combination of a less powerful classic design with a massively parallel architecture may then provide a much more cost-effective solution to many problems. In that case, the question of the fastest machine in the world becomes the matter of the *fastest code in the world*, and loses much of the prestige that is now often attributed to it.

Don't Judge Tools by Their Gloss

Nevertheless, the megaflops figure especially prominently in the public debates. In an earlier epoch, IBM's Thomas Watson refused to be second best. In the 1980s, the Japanese were seen as threatening "our number one position." CDC entered the 205 in the *Guinness Book of World Records.*[66] Glitter and glamour also play an important role, and even the physical appearance of supercomputers matters. The CRAY-1's "love seat design" appealed to many. The same is also true for the CRAY-2's very visible liquid immersion cooling system[67] and the ETA[10]'s not so visible cryogenic cooling system. There has been much attention paid to the use of gallium arsenide for the CRAY-3, which has attracted much more publicity than the C-90, although both aim for similar speeds.

Thus, supercomputers appeal to the general public as well as to policymakers. It is easy to talk about supercomputers in a rhetorical way and to connect them to national security and economic competitiveness, and that rhetoric certainly played a key role in the United States during the 1980s. Yet it is difficult to gauge just how important supercomputer development really is for security and competitiveness. An increase in speed of three- or fourfold, as is typical for Cray Research's products, implies that in three-dimensional models (the type of codes most frequently run) the linear dimensions of cells in a model can be reduced by a factor of 1.5. Helpful, yes, but surely not decisive.

In any case, the direction in which supercomputing is developing implies that narrow nationalism will be self-defeating. Future supercomputing environments can be expected to be much more heterogeneous than current ones, with a variety of hard- and software all linked together. In this context, it is surely advantageous for users to be able to choose from a variety of products coming from different vendors. A combination of U.S., European, and Japanese products may often appear to be called for, and any country taking a protectionist line would harm its own industry and research institutes by outlawing any one of them.

Supercomputers started being developed in an esoteric world, for computer élites at Los Alamos, Livermore, the National Security Agency, and some other key users. Over time, however, they have become embedded in a wider environment, where they are used as tools to solve practical problems—no more, no less. In public debates, however, they are still considered to be the highly prestigious flagship of high-tech, the yardstick of "where we stand in the world." It makes more sense, however, to see them in the context of all of their interrelationships in the current world. Such a view could prevent panic reactions on the basis of hurt pride that only have counterproductive effects.

Acknowledgments

The research reported on here is partly based on interviews with a large number of people who are or have been engaged in the development or use of supercomputers. We appreciate the time they spent with us and the information they provided. Organizing these interviews was greatly facilitated by the hospitality and resources of the Charles Babbage Institute at the University of Minnesota. The research was supported by the U.K. Economic and Social Research Council, under the Programme on Information and Communication Technologies (PICT). Some sections of this article were published earlier in the Dutch *Yearbook for Corporate History and History of Technology.* We thank the editors of the Yearbook for their permission to reprint those sections here.

Notes

1. A previous article by one of the authors examined the influence of one particular category of user—the nuclear weapons laboratories—on supercomputing: Donald MacKenzie, "The influence of the Los Alamos and Livermore National Laboratories on the development of supercomputing," *Ann. Hist. Comput.*, Vol. 13, 1991, pp. 179–201. In another article we focused on Seymour Cray's role in the development of supercomputing: Boelie Elzen and Donald MacKenzie, "The charismatic engineer—Seymour Cray and the development of supercomputing," in *Jaarboek voor de Geschiedenis van Bedriff en Techniek (Yearbook for Corporate History and History of Technology)*, Vol. 8 (Amsterdam: NEHA, 1991), pp. 248–277.
2. We use the term "network" in the sense that the term has been used in recent sociology of science and technology. See, for example, Boelie Elzen, Bert Enserink, and Wim A. Smit, "Weapon innovation—Networks and guiding principles," *Sci. Public Policy*, June

1990, pp. 171–193; or Bruno Latour, *Science in Action*, ed. Milton Keynes (London: Open Univ. Press, 1987). In these approaches, a network does not consist of artifacts alone (as in more technical uses of the word), nor of human beings alone (as in traditional sociology), but of both. There are important issues in the newer sociological usage concerning whether artifacts are to be considered as agents equivalent to human actors, but these are not central to our discussion here.

3. There is ambiguity in this definition as there is no universally accepted way of establishing the speed of a computer. Due to differences in architecture and software a computer that is the fastest on a specific code need not be the fastest on another code. The pragmatic definition given here, however, suits our needs because everybody in the computer world would agree that the machines we discuss should be classified as supercomputers.
4. The following discussion of the 1604 and 6600 is based on James E. Thornton, "The CDC 6600 Project," *Ann. Hist. Comput.*, Vol. 2, No. 4, October 1980, pp. 338–348; and on an interview with James E. Thornton, Minneapolis, Minn., April 3, 1990. Thornton helped design these computers.
5. Biographical notes on Cray can be found in Chapter 18 of Robert Slater, *Portraits in Silicon* (Cambridge, Mass.: M.I.T. Press, 1987).
6. For the history of Engineering Research Associates, see Erwin Tomash and Arnold A. Cohen, "The birth of an ERA: Engineering Research Associates, Inc., 1946–1955," *Ann. Hist. Comput.*, Vol. 1, No. 2, October 1979, pp. 83–97.
7. David E. Lundstrom, *A Few Good Men from Univac* (Cambridge, Mass.: M.I.T. Press, 1987), p. 136.
8. Slater, *Portraits*, op. cit., suggests that Cray told Norris he was leaving.
9. On "committee design," see Lundstrom, *Univac*, op. cit.
10. Charles J. Bashe et al., *IBM's Early Computers* (Cambridge, Mass.: M.I.T. Press, 1986), p. 446.
11. Thornton, *Ann. Hist. Comput.*, op. cit., p. 343.
12. Lundstrom, *Univac*, op. cit., p. 214; MacKenzie, *Ann. Hist. Comput.*, op. cit.
13. Kenneth Flamm, *Creating the Computer: Government, Industry and High Technology* (Washington, D.C.: Brookings, 1988), pp. 152–155. See Jacques Jublin and Jean-Michel Quatrepoint, *French Ordinateurs—de l'Affair Bull à l'Assassinat du Plan Calcul* (Paris: Alain Moreau, 1976).
14. Lundstrom, *Univac*, op. cit., p. 138, quotes the entirety of one report: "Activity is progressing satisfactorily as outlined under the June plan. There have been no significant changes or deviations from the outlined June plan." This contrasted with 20- or 30-page reports from the leaders of comparable parts of the corporation.
15. Quoted in, for example, Russ Mitchell, "The genius—Meet Seymour Cray, father of the supercomputer," *Bus. Week*, April 30, 1990, pp. 81–88.
16. Interview with Les Davis, Chippewa Falls, Wis., May 3, 1990.
17. Interviews with Neil Lincoln, April 3, 1990; Chuck Purcell, April 4, 1990; and James Thornton, op. cit., all in Minneapolis, Minn. Purcell worked as a salesman of the STAR and also took part in its design.
18. *Chippewa Falls Herald-Telegram*, March 15, 1972, quoted in *Speed and Power—Understanding Computers* (Alexandria, Va.: Time-Life Books, 1987), p. 32.
19. John M. Lavine, "New firm here blueprinting most powerful computer in the world—Cray Research Inc. putting together 'think tank', production facilities," *Chippewa Herald–Telegram*, May 17, 1972, pp. 1, 4.

20. See, for example, the discussion of the work of guidance engineer Charles Stark Draper in D. MacKenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge, Mass.: M.I.T. Press, 1990).
21. Cray acknowledged the debt of the CRAY-1 to the STAR-100 in a lecture at the annual supercomputing meeting in Florida on November 15, 1988. This lecture has been recorded on video: Seymour Cray, "What's all this about gallium arsenide?"
22. Interview with Les Davis, op. cit.
23. A more thorough technical description of the CRAY-1 can be found Richard M. Russell, "The CRAY-1 computer system," *Commun. ACM*, Vol. 21, No. 1, January 1978, pp. 63–72.
24. John Rollwagen, quoted in "Gambling on the new frontier," *Interface*, Vol. 9, No. 8, September 1986, pp. 6–7.
25. Construction of Serial 2 was halted when Serial 1 displayed a high level of memory errors when installed at Los Alamos, perhaps because of the high incidence of cosmic rays at the nuclear laboratory, high in the mountains above the New Mexico desert. Error detection and correction facilities were added to later models, starting with Serial 3.
26. Bradford W. Ketchum, "From Start-up to $60 million in Nanoseconds," *INC.*, November 1980, p. 56.
27. Cray Research, Inc., *Annual Report 1978*, Minneapolis, Minn., 1979, p. 2.
28. "The seismic processing problem: Cray Research's response," *Cray Channels*, Vol. 5, No. 2, 1983, pp. 6–9; also interview with Dave Sadler, Minneapolis, Minn., May 2, 1990. Sadler worked on the software for handlng the tapes.
29. Cray Research, Inc., *Annual Report 1984*, Minneapolis, Minn., 1985, p. 2.
30. Interview with Margaret Loftus, Minneapolis, Minn., May 4, 1990. Loftus was Cray Research's vice-president of software at the time.
31. Cray Research, Inc., *Annual Report 1982*, Minneapolis, Minn., 1983; also interviews with Les Davis, op. cit., and John Rollwagen, Minneapolis, Minn., April 2, 1990.
32. "Cray Research Fact Sheet" giving marketing data as of December 31, 1989.
33. Steve Gross, "Quicker computer? Cray leads the way," *Minneapolis Star*, December 22, 1981.
34. Cray video, op. cit.
35. Interview with Les Davis, op. cit.
36. Interview with John Rollwagen, op. cit.
37. "What comes after Y?" *Interface*, Vol. 12, No. 6, Summer 1989, p. 8.
38. "Cray lures users with entry-level price supermini," *Computing*, October 31, 1991, p. 8.
39. The following description of the CRAY-3 is largely based on the Cray video, op. cit.
40. Marc H. Brodsky, "Progress in gallium arsenide semiconductors," *Sci. Am.*, February 1990, pp. 56–63.
41. Letter from Neil Davenport, president of Cray Computer Corporation, to the authors, December 27, 1990.
42. Video tape "Cray Private," May 15, 1989, on which John Rollwagen explains the reasons for splitting off Cray Computer Corporation from Cray Research. See also Steve Gross, "Will Cray suffer without its founder?" *Star Tribune*, May 21, 1989.
43. Rollwagen quoted in Carla Lazzarechi, "Cray left his company to prevent internal conflict," *Los Angeles Times*, June 4, 1989, part IV, p. 4, and in Russell Mitchell, "Now Cray faces life without Cray," *Bus. Week*, May 29, 1989, p. 27.

44. Cray, quoted in Mitchell, ibid.

45. Cray Research, Inc., *1984 Annual Report*, Minneapolis, Minn., 1985, p. 2.

46. R. W. Hockney and C. R. Jesshope, *Parallel Computers 2—Architecture, Programming and Algorithms* (Bristol/Philadelphia: Adam Hilger, 1988), p. 117.

47. Interviews with Neil Lincoln, op. cit.; Tony Vacca, Chippewa Falls, Wis., May 3, 1990; and Carl Ledbetter, St. Paul, Minn., May 17, 1990. Lincoln and Vacca worked on various CDC and ETA machines, while Ledbetter was president of ETA.

48. *Speed and Power*, op. cit., p. 89.

49. Ibid, 91.

50. Thomas Levenson, "Wetenschap op topsnelheid" ("Science at topspeed"), *Intermediair*, Vol. 26, No. 20A, May 18, 1990, p. 29.

51. Sidney Karin and Norris Parker Smith, *The Supercomputer Era* (Boston/New York: Harcourt Brace Jovanovich, 1987), p. 252.

52. Argonne National Laboratory; quoted in John Markoff, "A computer star's new advance," *New York Times*, February 17, 1990, pp. 2 and 21.

53. This is a rough estimate because, as a matter of policy, IBM does not release figures about numbers of machines installed; Karin and Smith, *Supercomputer Era*, op. cit., p. 291.

54. Markoff, *New York Times*, op. cit., p. 21; Amanda Mitchell, "Renegade supercomputer maker pushing frontiers of technology," *Parallellogram*, October 1990, p. 9; Mitchell, *Bus. Week*, op. cit., p. 85.

55. Jerry Sanders and Amanda Mitchell, "Dateline 1995!" *Parallellogram*, October 1990, pp. 8–9.

56. Karin and Smith, *Supercomputer Era*, op. cit., pp. 105–106.

57. Ibid., pp. 107–108, 110–114.

58. The Superperformance Computing Service, *Superperformance End-User Requirements and Expectations*, Mountain View, Calif., 1990, p. II-3.

59. Karin and Smith, *Supercomputer Era*, op. cit. pp. 109–110.

60. Willie Schatz, "A call to arms," *Datamation*, April 1, 1984, p. 34.

61. Cray Research, Inc., "Keeping up with current events—An interview with John Rollwagen," *Interface*, Vol. 10, No. 5, May 1987, pp. 6–7.

62. Quoted in Gross, *Star Tribune*, op. cit.

63. Interviews with Carl Ledbetter and Neil Lincoln, op. cit.

64. *Speed and Power*, op. cit., p. 97.

65. Interview with Carl Ledbetter, op. cit.

66. Davis Stamps, "The Cray style—John Rollwagen has a company to run and a reclusive genius to satisfy. The trick is to do both at the same time," *Corp. Rep.*, December 1982, pp. 68–72.

67. The CRAY-2 is sometimes referred to as "the world's most expensive aquarium."