NSA and the Supercomputer Industry

(b)(3)-P.L. 86-36

[redacted]

(b)(1)
(b)(3)-50 USC 403
(b)(3)-P.L. 86-36

NSA'S RELIANCE ON HIGH-PERFORMANCE COMPUTING TECHNOLOGY

~~(TSC)~~ NSA is heavily dependent upon high-performance computing (HPC) technology, heretofore known as supercomputing technology, particularly in the area of cryptanalysis. HPC is used for fundamental cryptanalytic (CA) mathematical research, diagnosis of unknown cryptologics, development of attacks, and daily exploitation for SIGINT production. Often, CA efforts which eventually find their way onto special-purpose devices (SPDs), high-performance desktop computers, or [redacted] had their genesis in breakthroughs made on NSA supercomputers. Additionally, NSA has a smaller portion of its HPC assets devoted to signals processing and to the protection of United States cryptographic systems.

~~(S-CCO)~~ As a consequence, NSA has constructed the largest single-site, single-mission supercomputer complex in the world. This has allowed the Agency to gain unique HPC experience as well as leverage with HPC vendors. A symbiotic relationship has evolved over time. NSA shares problems and requirements with HPC vendors; in exchange it gets systems that not only meet its needs better but also create improved products for the entire range of HPC customers. Indeed, NSA can cite multiple examples of supercomputer technology spanning several decades which was designed as a direct result of our requests.

~~(TSC)~~ Of equal importance is technology related to the construction of SPDs. Without an extremely close relationship with HPC vendors, many of NSA's current (and future) SPDs would not have been possible. [redacted]

(b)(1)
(b)(3)-P.L. 86-36

HIGH-PERFORMANCE COMPUTING TECHNOLOGY TRENDS

~~(C-CCO)~~ Although vector supercomputers have commanded the bulk of the HPC marketplace until now, a paradigm shift is under way to massively parallel processor (MPP) platforms. MPPs theoretically offer more cost-effective hardware than vector systems. Unfortunately they suffer from the lack of large uniform access memory, from high latencies (i.e., distributed processor and memory communication times) and from immature operation systems and software support tools. MPPs can outperform vector machines by a wide margin for some types of problems. However, across the broad range of applications most programmers fall well short of achieving the theoretical peak

The opinions expressed in this article are those of the author(s) and do not represent the official opinion of NSA/CSS.

performance of MPPs. NSA has made extremely effective use of its MPP systems. But it has come at the expense of the very best programmer talent, which is a precious resource. Furthermore, for some CA users, in particular the highly interactive researchers and those working in diagnosis, the characteristics traditional vector supercomputers have to offer are still the best fit for the problems. They provide an environment that fosters ingenuity and creativity while also providing rapid response to the researcher.

~~(C-CCO)~~ The last point on software is multifaceted. Vector supercomputers from Cray Research Incorporated (CRI) have well-developed and powerful operating systems and tools which are superior to those found on competing machines. NSA has the world's largest pool of experience in making use of these platforms. Programmers can spend their time concentrating on their difficult mission of cryptanalysis with minimal attention to the computing infrastructure. Also, the Agency of course has a large collection of legacy software (estimated at close to $3 billion of development effort) which can be easily reused on each succeeding generation of vector machines without the timely and expensive task of porting that code to a new architecture.

(U ~~PROPIN~~) The two largest U.S. vector supercomputer vendors have not neglected the technical and market trends. Both CRI and Convex Computer Corporation have announced they have introduced their last pure vector supercomputers. Both have MPP products available today. Indeed, in CRI's first year they went from 0 percent market share to become the world's leading MPP supplier in terms of revenue.

THE VENDORS

~~(FOUO)~~ Figure 1 shows the market share for vector supercomputers. Vector supercomputers have long been the mainstay of CA processing and, at least at the moment, command the bulk of the HPC market for most other applications as well. Fortunately for the U.S., CRI is the world leader in this area. Convex is in the number-three position; all other vendors with any significant market share are Japanese companies.

(U) One of NSA's major concerns in the HPC area is the declining marketplace at the high end. Systems costing over $10 million have been declining at almost a 50 percent rate per year as shown in figure 2. This is due to a number of factors including erosion of the lower end of the market by high-performance, reduced instruction set computers (RISC), the high research and development costs of HPC in contrast to their small niche in the overall computing marketplace, a paradigm shift between vector processors and MPPs, increasing computing performance over time for constant dollars, and reduced governmental investment in HPC technology. U.S. supercomputer vendors are also under pressure from export controls as well as significantly higher research and development spending by Japanese HPC multiproduct companies.

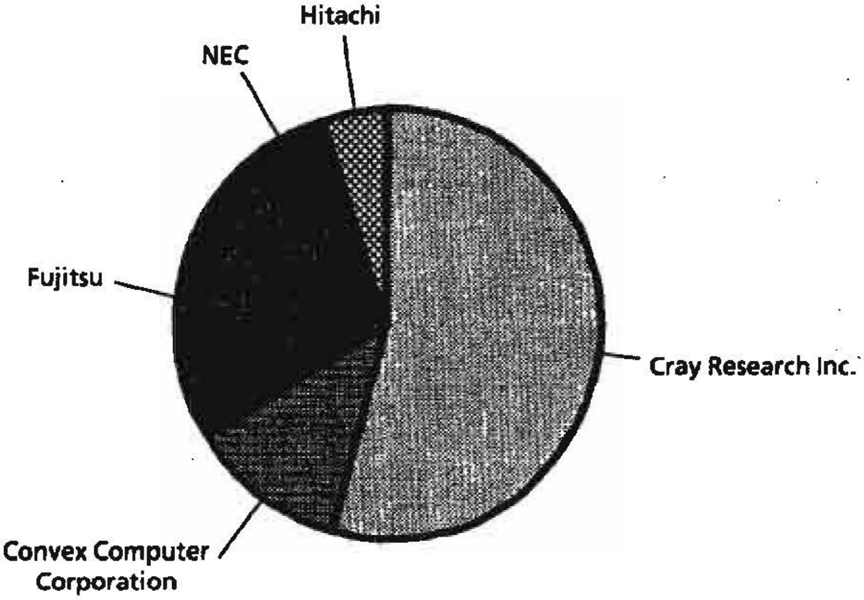

Fig 1. Vector market share by company

(U ~~PROPIN~~) All of the above have taken their toll on U.S. HPC vendors. In the last few years, sixteen domestic HPC vendors either have failed completely or have left the field of HPC.[1] Two more companies are in Chapter 11.[2] Of the remaining companies, many have never shown a profit; e.g., Intel's supercomputer division has been kept afloat through subsidies by the larger chip manufacturing division. Furthermore, some have never even sold a machine; e.g., Tera has been living off research grants and soon off investment capital alone. Only four companies can simultaneously claim economic profitability, significant market share, systems which offer medium to high levels of performance, and reasonable software tools, including support for mass storage and high-performance networking. Those four companies are CRI, Convex, IBM, and Silicon Graphics Incorporated (SGI).

[1] Alliant, BBN, Cray Computer Corp., Control Data Corp., Denelcor, Elxsi, ETA, Floating Point Systems, Goodyear, Multiflow, Myrias, Prime, SCS, Sequent, Supercomputer Systems Inc. (in partnership with IBM), and Symult (aka Ametek).

[2] Kendall Square Research and Thinking Machines Corp.

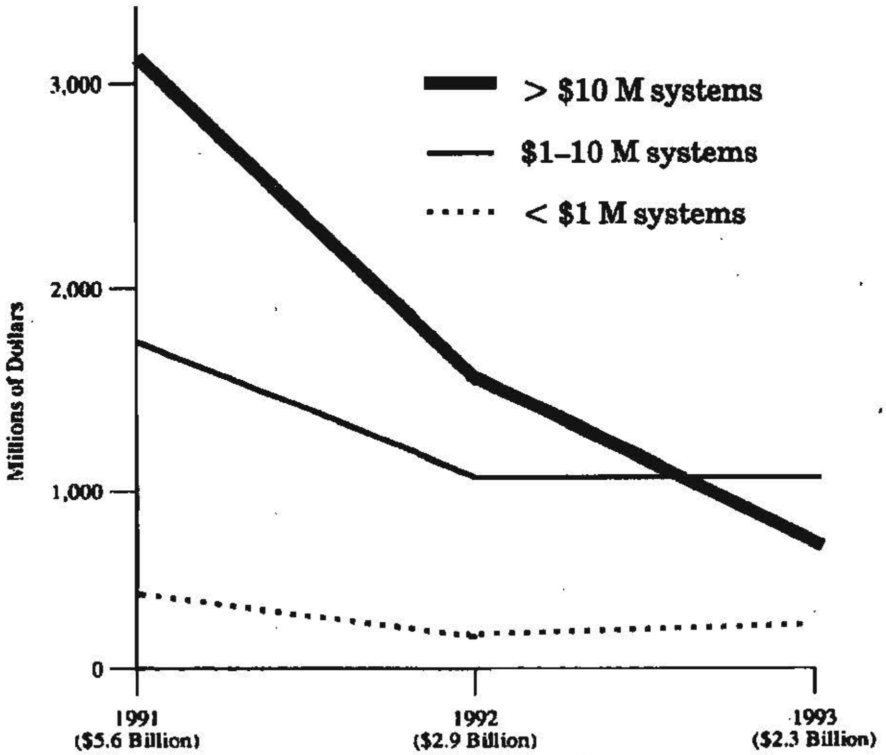

Fig 2. Worldwide HPC revenue by year (both Vectors and MPPs)

~~(C-CCO)~~ Of those four companies only one, CRI, currently offers systems at the very extreme levels of high performance necessary for the demanding task of research, diagnosis, and exploitation of the most challenging target cryptologics. Furthermore, CRI offers a completely balanced system which incorporates not only high performance processors but also large global memory, high memory bandwidth, and very fast I/O. It does little good to design high theoretical computational capacities if the processors are starved for data – a lesson many vendors failed to learn. This is, of course, one of the major reasons why we have been so dependent upon CRI for mainstream CA processing.

(U) Even while companies have been failing rapidly in the last few years and revenue from high-end systems has been declining, CRI's market share has been steadily increasing. The declining revenues have come, by and large, at the expense of CRI's competitors. However, that does not mean that CRI is healthy or that the future of HPC in the U.S. is secure. Indeed, CRI's revenues (approximately $950 million) have been flat for the past few years in an overall declining marketplace.

(U) Of the Japanese vendors, Fujitsu has the largest market share, followed by NEC and then Hitachi. For some (but not all) standard industry benchmark sets, the Fujitsu and Hitachi machines have scored better than CRI, so their hardware can be competitive for certain applications. However, the U.S. has a commanding lead in software development. The Japanese currently seem to be concentrating on the migration to parallel processing systems with plans for both indigenous development as well as the acquisition of U.S. technology when appropriate.

THE FUTURE OF HIGH-PERFORMANCE COMPUTING

~~(FOUO)~~ The future can be divided into three phases: now to 1998, 1998 to 2000, and 2000 forward. From now to 1998 there is little likelihood that anyone will be able to surpass CRI in performance. NSA should depend on CRI for most of its CA needs during that period.

~~(FOUO PROPIN)~~ Starting in 1998, things become more interesting. CRI will introduce its scalable node architecture (SNARK), a hybrid system which will try to combine the best of both traditional vector processing and low-cost MPPs. At the same time, their plans are to discontinue introducing pure vector or pure MPP products, thereby betting the company on the success of SNARK. SGI, hampered today by a bus-based architecture which prevents scaling up to a larger number of processors, will be ready with a product they currently call LEGO, which will use a true interconnection network. SGI has the further advantage of customers in the growing multimedia and entertainment business, which provides them with a volume market. Convex has just announced an agreement to be purchased by Hewlett-Packard (HP). HP is further teamed with Intel. Convex will introduce a sixty-four-bit MPP system based upon the merged HP and Intel RISC chips. This alliance will be difficult to ignore as it will offer architecturally consistent hardware platforms from the desktop to the supercomputer, all fully software compatible. IBM is always difficult to count out; however, their SP systems are relatively high in latency, and we have no indication that they plan to change this anytime soon. IBM systems seem best suited for true embarrassingly parallel problems, which are the most simple parallel processing problems, and form only a small fraction of the total problem set at NSA.

(U) It is also possible that some of the other smaller players (e.g., MasPar, NCube, Tera, a restructured Thinking Machines, or a new start-up company) might emerge as contenders in a few more years. But unless they can break out of a small market niche, it will be difficult for them to challenge the larger, more well-established, and more experienced HPC vendors. Funding may be another very tough obstacle for them to overcome as well in a very expensive business.

(U) So from 1998 to 2000 it is likely that CRI will face serious competition from other vendors. There will probably be a shakeout of the industry, with not all players surviving at the high end. Furthermore, there will probably be no room for two vendors the current size of CRI.

(U) From the year 2000 and beyond, things become more difficult to predict in an industry changing as rapidly as computers. However, it is possible that as RISC chips continue their phenomenal performance growth rate and as networking technology becomes better and less expensive, we may begin to see yet another paradigm shift. Just as vector processors are in the process of yielding to MPPs, development of MPP systems may yield to the implementation of clusters of distributed workstations in the business world. Clustered RISC workstations have the same advantages and disadvantages as MPPs – only more so. They are theoretically cheaper. However, they are currently tuned with the desktop applications in mind, which means very high communications latencies, poor memory bandwidth, small memories and caches, slow I/O, and immature software support. Until business applications demand that these constraints be removed, cluster computing using workstations will not be a viable alternative for our research and diagnostic problems. However, an additional force driving the industry in this direction is the ever-spiraling cost of fabrication facilities. Without the huge volumes of desktop and embedded systems only RISC processors will have, the capital investments necessary to sustain research and production will simply not exist. Furthermore, as a result of a consolidation in the software industry (also brought about by the volume represented by desktop computing), there may be further market pressures to standardize on a fewer number of higher volume hardware platforms.

~~(FOUO)~~ If CRI fails in the next few years, NSA and the U.S. government would simply have no choice but to rely upon a less productive mix: Convex for vector systems, and Convex, IBM, and SGI for MPP systems. [redacted] NSA would also be even more dependent than ever on its SPD and chip manufacturing technologies.

(b)(3)-P.L. 86-36

~~(FOUO)~~ Were that to happen, it is important to keep several things in mind. First, the Agency would not be caught off guard; NSA works very closely with CRI and would have sufficient advance notice. Second, it would probably be able to arrange for ongoing maintenance of its installed base of CRI hardware for the duration of its useful lifetime (either through contracted maintenance or perhaps in-house support). Third, before CRI completely failed, NSA would of course begin migrating its installed base of software to alternate systems in order to support mission requirements to the maximum extent possible.

(U) NSA will continue to closely monitor the volatile HPC technology and marketplace. Planning should begin before 2000 to ensure that the U.S. will continue to dominate HPC technology as we move into the next millennium.

RECOMMENDED COURSES OF ACTION

~~(FOUO)~~ The emphasis on courses of action should not be based on the premise that CRI will be out of business within the next five years. It is highly unlikely that they will. NSA should target its emphasis toward two other goals. First, what can the Agency do to keep companies like CRI sufficiently profitable enough to be able to invest in the development of HPC hardware and software to meet NSA's needs? Second, what can NSA do to encourage those throughout the HPC community, not just the vendors, to develop state-of-the-art software needed to take best advantage of emerging HPC technology? NSA recommends the following courses of action be taken.

- **Treat the supercomputing industry as an asset critical to national defense.**

~~(FOUO)~~ The commercial viability of CRI and the rest of the supercomputing industry is critical not only to NSA but also to the entire Western world cryptanalytic community. NSA relies on balanced, high-performance computing products which currently are uniquely available from CRI and the Agency's major investment is in legacy applications software on CRI computers.

(U) NSA alone cannot sustain the health of CRI or any of the other HPC vendors. It needs the major investment assistance of DoD and other governmental agencies at a minimum. It is important to continue sustained annual DoD and other U.S. government investment in supercomputers. Establishing a consortium of potential DoD users who would agree to earmark a specific sum of money for CRI or other vendors' purchases after consultation with the company could provide additional investment for the vendor and should be considered. At this moment, the bulk of this investment would be made with the only viable company at the very high end of performance: CRI.

(U) In order to counter the influence that the commercial market has on vendors' business strategies, the U.S. government – not just the NSA nor the DoD – should present a united front to the industry stating the importance of HPC to national security. The interagency High-Performance Computing and Communications Program should place emphasis on projects that benefit companies interested in scientific supercomputing.

- **Reduce export controls on supercomputer technology.**

(U) The difficulties associated with the export of high-end machines materially affect the viability of the supercomputer industry. Increased market opportunities for CRI and others can produce substantial gains in their financial health.

(U) Effective changes should be implemented in disclosure requirements within trade agreements in order to allow American companies to compete with other foreign vendors on a balanced and fair playing field relative to export sales.

- **Support expanded research funding in HPC technologies.**

(U) The U.S. government, commercial, and academic organizations are critical to support fundamental research and development of HPC technologies. Each has a role to play in ensuring that vendors such as CRI continue to lead the world in HPC.

~~(FOUO)~~ Research funding should continue for the Institute for Defense Analysis (IDA) Center for Computing Sciences and the Center for Communications Research. These IDA centers should be able to acquire large research and development supercomputers for their vital work. Also, the interagency High-Performance Computing and Communications Program, and joint research efforts (such as Cooperative Research and Development Agreements and the SPD development projects) should encourage continued progress in HPC technology via public/private partnerships.

(U) Research funding should be supported so that small systems from other likely HPC companies can be brought in for evaluation and feedback. Ongoing revenue even at a small level is important for product development.

(U) Increase staffing and training to organizations which have the expressed charter to develop or port all levels of software to MPP technology. This includes traditional NSA organizations as well as Federally Funded Research and Development Centers.

(U) Since industrial research tends to be short-term, long-term computer science research being done in academia is important for future technology transfer to the marketplace.

(U) In those cases in which there is an insufficient commercial marketplace for HPC technologies, the federal government may need to intervene. One example of this is the limited demand for high capacity fast SRAM (static random access memory) memories. Although it is technically feasible to build these memories today, the bulk of the production is going to slower DRAM (dynamic random access memory) memories used on desktop PCs, which command a larger share of the commercial marketplace.

(b)(3)-P.L. 86-36

[redacted]

~~(FOUO)~~ [redacted] is the Assistant Technical Director in J1, the Office of Cryptanalytic Processing. He has been with NSA since 1987, serving as a computer analyst and computer scientist. From 1987 to 1988 he was assigned to T3353, Supercomputing UNIX Operating System Support. From 1988 to 1990 he was a T Systems Research Integree at the Institute for Defense Analysis Supercomputing Research Center, and from 1990 to 1992 he was assigned to T3353. [redacted] [redacted] [redacted] is a member of [redacted] and NSA Science and Engineering Society. He received the [redacted] The author of several articles, he is professionalized as a Computer Systems Analyst and is a Master in the Computer Systems Technical Track.

(b) (6)

(b) (6)

(b) (6)

Derived From: NSA/CSSM 123-2,
Dated 3 September 1991
Declassify On: Source Marked "OADR"
Date of source: 3 Sep 91