A History of Supercomputers

Les Davis, Cray Research Inc.

ABSTRACT: *The design and manufacture of large scientific computers is only successful if scientists, engineers and designers are able to use them as successful tools. After all, it is the computer in the hands of the user that makes remarkable things happen. It was in response to a tremendous appetite for scientific computing power that Engineering Research Associates (ERA), Control Data Corporation (CDC) and Cray Research Incorporated (CRI) invested heavily in the development of ever faster and more powerful computers.*

Introduction

I would like to share with you some of my observations about one area of high performance computing. I will point out some of the enabling technologies and some of the external influences that shaped the development of the most powerful computer systems during the decades of the 1950's, 1960's, 1970's, 1980's and now the 1990's.

Decade of the 1950's

During the late 1940's and early 1950's vacuum tubes replaced relays as the primary computer logic building block. The vacuum tube had a significant impact on computer performance by enabling clock rates to go from milli-sec's to micro-sec's. At the same time, magnetic core material was being used as a memory element allowing for faster and larger main memories (these memories by todays standards would be small register files). The alterable stored program versus the fixed or hard wired program was a significant and important architectural development that helped advance the development of computers. Vacuum tubes, magnetic cores and the stored program provided computer designers with the technology to build high performance computers. IBM and Remington Rand Univac built several very large and powerful computers during the 1950's which were used primarily by U.S. government agencies.

ERA (Engineering Research Associates) was a small company started in 1946 by a group of U.S. Navy reserve officers who had worked together during WWII in a code breaking unit. A member of this group was Bill Norris. ERA was located in a former glider factory in St. Paul, Minnesota and attracted many talented engineers who were eager to work on the design and manufacture of computers. One of the engineers was Seymour Cray. The ERA computers were scientific computers most often built under government contract which was a thriving business. ERA had very aggressive development and manufacturing plans which required significant capital to fund. To obtain the necessary capital ERA turned to Remington Rand (latter to be Sperry Rand Corporation and then Unisys) and became, in 1952, a subsidiary of Remington Rand's growing computer business. Remington Rand had acquired, in 1950, Eckert-Mauchly Computer Corp. developers of Univac's (Universal Automatic Computers) which were intended for commercial computing. The Remington Rand - ERA relationship soon caused a great deal of difficulty as Remington Rand and ERA management clashed over R&D expenditures and product direction. In 1957 Bill Norris, Seymour Cray and several other key people left Remington Rand to form Control Data Corporation.

Summary of the 1950's

- Speed was of primary concern even at the risk of non-compatibility.
- Clock rates improved dramatically. Vacuum tubes provided significant speed improvement over relays.
- Capital expenditures were significant for small companies.
- Business was primarily government.
- Primary supplier of parts: U.S. companies.
- Competitors: Remington Rand Univac, IBM.

Decade of the 1960's

The first commercial transistor computers in the U.S. had been delivered in early 1960, among them the 1604 built by Control Data Corporation (CDC) then only a 3 year old company. In 1960 the demand for significant increases in computer power was very great. The 1604 computer was in great demand by government laboratories because it offered significant improvement in performance over then installed computers. However, there was great pressure for even greater levels of computer performance. New computer architectures and new ideas were felt essential to meet the increased demand for computer power. At Control Data, just as the first 1604's were being delivered, a new design project was initiated. The goal - -- to significantly exceed the performance of the fastest computers currently in production or development. The Control Data project started in 1960 and led by Seymour Cray was completed in September 1964 with the delivery of the first 6600 to Lawrence Livermore Laboratory. The 6600 set a new standard for the most powerful scientific computers.

The 6600 included some bold advances in computer design: an architecture which extended the use of parallel processing; an I/O scheme using multi-processors to minimize CPU interference; a unique and dense packaging scheme which minimized wire and interconnect path lengths; and a very fast and, for that time, large central memory.

During the 1960's , the development of the transistor was one of the most significant enabling technologies that allowed the design and construction of fast and powerful computers. In order for the transistor technology to be successful, packaging technology and interconnect technology had to make significant improvements or the speed improvements of the transistor would not be realized. Those companies which were able to quickly master the packaging complexities were successful. In addition to packaging technology, significant increases in memory size were essential. The development of very fast magnetic core materials which could switch in nano-seconds instead of micro-seconds provided memory speed and memory density. Again, packaging expertise was an absolute essential.

Architectural innovation was also important and even though scalar calculation was the primary mode of calculation, parallel operation was being experimented with. The Control Data 6600 and 7600 systems embodied all of the above technologies: transistor technology, magnetic core technology, packaging and architectural technology, to produce very successful products that set the standard of performance for many years. During the 1960's, the investment in capital equipment for the computer manufacturer, although thought to be expensive at the time, was minimal. The dimension of components although significantly smaller than vacuum tubes allowed for assembly with the aid of rather inexpensive microscopes. The design tool requirements also were quite small and most design checking and verification was done manually.

Software concerns were minimal. During the 1960's ,the primary user of the large scientific computers were government laboratories who had large staffs of programmers dedicated to writing operating systems, compilers and application codes for dedicated problems. The time from delivery of a system to production use often ranged from six months to one year. Hardware designers placed software compatibility low on the priority list during this period. System performance increased dramatically during this period primarily due to the significant increase in clock speed.

Business conditions were very favorable during the early and mid-1960's due to large budgets and government support for the purchase of large scientific computers. Toward the end of the 60's demand slowed as budgets began to tighten and the appetite for computing slowed. The result was a significant slow down in the rate of R&D expenditures for the development of the fastest computers. This trend continued into the early 1970's and was a primary factor in the departure of Seymour Cray from CDC and the start of Cray Research.

Summary of the 1960's

- Speed was of primary concern even at the risk of non- compatibility.
- Clock rates improved dramatically.
- Capital expenditures were minimal.
- Business was primarily government, but very good.
- Primary supplier of parts: U.S. companies.
- Competitors: Control Data, IBM.

Decade of the 1970's

In the early 1970's development of faster transistors and the development of packaging technologies to allow for faster clock rates did not progress well enough to allow for significant improvement in high performance computing. In an attempt to build a faster system despite these limitations, the 8600 project at Control data was to be a four scalar processor system utilizing faster, higher powered discrete transistors. The number of mechanical connections increased dramatically as did the power density of the computer modules. The difficulties in solving the cooling and connection problems led to project delays at a time when budget cuts were being considered. The lack of Control Data support for the project caused Seymour Cray to leave Control Data and start Cray Research. The lack of progress in discrete transistor development and packaging techniques to assemble and cool large numbers of individual transistors was offset by the significant improvement in integrated circuit technology. By 1972 register chips, memory chips and small but high speed logic circuits were available. The integrated circuit technology which substantially reduced the number of mechanical connections and significantly reduced the power consumption of the computer module proved to be a key enabling technology for the building of a high performance computer during the 1970's. With a few exceptions, assembly techniques were simplified and yet the physical size of the system actually became smaller allowing for faster clock rates.

The design of the Cray-1 computer capitalized on these technologies and when coupled with the Cray-1 vector register implementation provided a computer system with un-matched performance. The integrated circuit technology used in the memory area allowed the system memory size to be significantly increased over previous high performance systems which was an added bonus for the early users of the Cray-1 systems.

The investment in capital equipment for the computer manufacturer although still moderate began to increase as the need for better IC test equipment became apparent. The need for newer design tools was also important as the expense of IC design increased and to avoid the costly delays incurred by a faulty IC design.

Gate counts (the number of logic circuits) increased but package counts decreased significantly. Moderate improvements in clock rates did occur during the 1970's but the rate of improvement slowed significantly. Software concerns began to be more important than in the 1960's. Budget restraints at government laboratories and the increasing complexity of the newer computers required the computer manufacturer to take a larger role in the development of operating systems and compilers. R&D dollars for software development started to increase.

During the 1970's, system performance increased primarily due to new architectural implementations - vector processing - and to advances in software compiler technology. This allowed significant improvements in performance compared to just scalar processing.

Business conditions during the 1970's were generally poor compared to the business climate of the 1960's. Tighter budgets resulted in many delayed or cancelled computer procurements.

However, these delayed decisions resulted in a built up demand for computer power and by the end of the 1970's government laboratories were rushing to update their computing facilities with the latest technology which included the Cray Research computer systems. There was also an increased demand from non U.S. computer users to have access to the latest computer technology.

Summary of the 1970's

- Speed was still important. However, concerns for compatibility and reliability were becoming as important.
- Clock rates continued to improve somewhat. Integrated circuits provide significant size and power reductions.
- Capital expenditures starting to increase.
- Business was generally slow.
- Primary supplier of parts: U.S. companies.
- Competitors: Control Data, Cray Research, IBM, Burroughs, TI.

Decade of the 1980's

The success of the Cray-1 computer with the introduction of vector processing in the late 1970's presented a challenge to computer designers and manufacturers in the 1980's. Scientists and engineers in government labs and in industry now had available computer power capable of solving mathematical problems that once were impractical to try solving. Automobile crash simulation, aircraft design, oil exploration, weather prediction were some of the areas that flourished. However, it seemed that the users now became dependent on their new computing resource and the demand for even more powerful and more cost effective computer systems was made.

The enabling technologies to allow the design of more powerful computers were not readily available in the early 1980's. IC technology was increasing the density of the IC circuits but was not significantly increasing the speed of the circuits. Packaging technologies required significant work to provide cooling and interconnection of the IC's. At Cray Research, in order to meet the challenges, two design teams were funded one for the XMP project and one for the Cray 2 project. Both projects designed computer systems using multiple vector processors (four) to provide the increase in system performance as well as designing packaging and cooling for the systems. By 1985 both the XMP system and the Cray 2 system were completed and in production and the YMP and Cray 3 follow on projects were initiated. The YMP project was completed in 1988 while the Cray 3 project was actually spun off into a separate company in 1989. At Cray Research, the C-90 project was begun in 1988.

Significant changes occurred in the 1980's in the design and manufacture of the most powerful computers. Because the speed of the IC circuits were not increasing at a fast enough rate to meet the needs of computer users, Cray system designers included more parallel structure in the design of their systems. Vector processing and multi- processing were essential to improve system performance. In addition to the increased processor speeds, significant improvements were required in the Input/Output areas of the computer system to handle the large amounts of data. The designers in the 1980's had to consider very carefully the impact new designs would have on the software development. The expanded use of supercomputers beyond government labs had generated many new application codes which in many cases could not be easily changed to accommodate a new computer design. If not designed to run the application codes well, a new system could be non competitive in the market place. The rate of software development expenditures increased rapidly during the 1980's. Operating system development, compiler development and networking software development required expenditures equal to the hardware development. The need for large memories was also important as the system speeds increased resulting in considerable development effort to build large, fast memories.

At Cray Research the investment in capital equipment increased at a very rapid rate in the 1980's as integrated circuit facilities and printed circuit facilities were constructed to allow for the rapid design and production of the new generation supercomputers.

Summary of the 1980's

- Speed was still important. However, price/performance became a measure of a systems real value.
- Application programs became a key to the successful use of a high performance computer system and to it's competitiveness in the market.
- Memory sizes increase substantially and greater degrees of parallelism appear in systems.
- Clock rates increase marginally.
- Software expenditures increase significantly.

- Capital expenditures start to increase dramatically.
- Business was very good.
- Primary supplier of parts: U.S. companies(logic) and Japanese companies (logic and memories).
- Competitors: Control Data, Cray Research, IBM, Fujitsu, Hitachi, Nec, Thinking Machines

Decade of the 1990's

The promise of MPP (Massively Parallel Processor) for high performance and the cost effectiveness of workstations presented a serious challenge to the designers of high performance computers.

To meet the challenge of the 1990's, Cray Research's development projects were ambitious. Significant increase in computer power were essential to meet the demands of engineers scientists and designers. To continue to be the leader in the design and manufacture of supercomputers required significant development in all areas of computer design: architecture, components, packaging, and software.

The C-90 and Triton computer systems continue to provide Cray Research with excellent high performance parallel vector systems. The enabling technologies that Cray Research pursued to provide for the design and manufacture of the C-90 and Triton were very different than those of the decade of the 1960's or 1970's. Very large scale integration of very fast silicon bipolar circuits required an investment in an internal IC facility. The need for very complex and fast printed circuit substrates again required the investment in internal printed circuit facilities. Manufacturing and assembly of the small and delicate components required significant investments in capital equipment.

Despite the performance gains provided by the C-90 and Triton systems, the demand for greater performance and better price performance required the investment in new architectural development. High performance work stations competed for the low end of the high performance work load and provided an attractive price performance for many tasks. The desire for teraflop performance levels versus gigaflop performance spurred a great deal of interest in MPP systems for the high performance tasks.

In 1993, Cray research brought to completion a significant project - the T3-D MPP system. The T3-D MPP system provided significant performance improvements for certain applications which have high degrees of parallelism. The need for increased software development effort was essential in order to take advantage of the design characteristics of this new system.

IC component development, during the 1990's, shifted away from speed enhancements to density improvements. It was now possible to put an entire CPU on a single chip. Memory improvements were even more dramatic. DRAM memories provided great density improvements allowing for very cost effective large memories at the expense of speed improvements. The development of faster logic and memory circuits traditionally used in the design of the very high performance computers began to lag substantially the development of the slower but more dense integrated circuits used in PC's and workstations presenting a real dilemma for designers.

Summary of the first half of the 1990's

- Speed was still important. However, price/performance continues to be a measure of a systems real value.
- Application programs are also key to the successful use of a high performance computer system and to it's competitiveness in the market.
- Memory sizes continue to increase and greater degrees of parallelism appear in systems.
- Clock rates increase marginally.
- Software expenditures continue to increase.
- Capital expenditures continue to rise during this period.
- Business slows dramatically for the high performance systems as defense spending drops and competition from MPP and workstations grows.
- Primary supplier of parts: U.S.(logic) and Japanese(memories) companies.
- Competitors: Cray Research, IBM, Fujitsu, Nec, Thinking Machines, Intel and other MPP companies at the high end. Many workstation vendors at the low end.

Conclusion

The future of supercomputing is challenging, we have seen almost five decades of spectacular growth. There is a challenging period through the remainder of this decade. The challenge of changing technology, the challenge of a changing economic environment, the challenge to retain and attract talented people, the challenge as we continue to move from defense spending to peace time spending and the challenge to the industry to remain competitive. The development of hardware and software technology today is driven by the low end of the market which is dominated by PC's and workstations. This is in contrast to the period from the 1950's thru the 1980's when the development of hardware and software technology was driven by the high performance market. The high performance systems of the future need to provide continued memory and I/O bandwidth and latency improvements in order to deliver the processor performance improvements to the user. Memory and I/O bandwidth and low latency designs have been strengths of Cray Research systems. The high performance system of the future may not be determined by one architecture. Parallel vector and MPP architecture will be important elements of future systems. New circuit technologies, new packaging technologies, architectural concepts and software design will have to be used together to provide the highest level of performance at a cost effective price.