Supercomputers: The Amazing Race

(A History of Supercomputing, 1960-2020)

Gordon Bell
November 2014

Technical Report

MSR-TR-2015-2

Gordon Bell, Researcher Emeritus

Microsoft Research, Microsoft Corporation

555 California, 94104

San Francisco, CA

Supercomputers: The Amazing Race

Timeline (The top 20 significant events in the History of Supercomputers. Constrained for Draft IEEE STARS Article)

1. *1957 Fortran introduced for scientific and technical computing*
2. 1960 **Univac LARC, IBM Stretch, and Manchester Atlas** finish 1956 race to build largest "conceivable" computers
3. 1964 Beginning of Cray Era with **CDC 6600** (.48 MFlops) functional parallel units. "No more small computers" –S R Cray. "First super"-G. A. Michael
4. *1964 IBM System/360 announcement. One architecture for commercial & technical use.*
5. *1965 Amdahl's Law defines the difficulty of increasing parallel processing performance based on the fraction of a program that has to be run sequentially.*
6. 1976 **Cray 1** Vector Processor (26 MF) Vector data. Sid Karin: "1st Super was the Cray 1"
7. 1982 **Caltech Cosmic Cube** (4 node, 64 node in 1983) Cray 1 cost performance x 50.
8. *1983-93 Billion dollar SCI--Strategic Computing Initiative of DARPA IPTO response to Japanese Fifth Gen. 1990 redirected to supercomputing after failure to achieve AI goals*
9. 1982 **Cray XMP** (1 GF) Cray shared memory vector multiprocessor
10. *1984 NSF Establishes Office of Scientific Computing in response to scientists demand and to counteract the use of VAXen as personal supercomputers*
11. 1987 **nCUBE** (1K computers) achieves 400-600 speedup, Sandia winning first Bell Prize, stimulated Gustafson's Law of Scalable Speed-Up, Amdahl's Law Corollary
12. 1993 Beginning of Multicomputer era with **Thinking Machines CM5** (60 Gf Linpack, 1024 computers) proof of multicomputer efficacy
13. *1993 Top500 established with LINPACK Benchmark*
14. *1994 MPI-1 Standard 1993 (first draft)*
15. *1994 Beowulf standard package of tools*
16. *1995 Accelerated Strategic Computing Initiative (ASCI) > 2002 Advanced Simulation and Computing (ASC) Program for programming at DOE labs*
17. *1996 Seymour Cray killed in a car accident*
18. 1997 **ASCI Red** (1 TF, 9.1K) at Sandia
19. 2008 **IBM BlueGene** (1.1/1.5 Pf, 122K) breaks Pf barrier using >100K ism
20. 2012 **Cray Titan** (17.6; 560K) GPU and CUDA demonstrates use of GPU

Abstract

The "ideal supercomputer" has an infinitely fast clock, executes a single instruction stream program operating on data stored in an infinitely large, and fast single-memory. Backus established the von Neumann programming model with FORTRAN. Supercomputers have evolved in steps: increasing processor speed, processing vectors, adding processors for a

program held in a single memory monocomputer; and interconnecting multiple computers over which a distributed program runs in parallel. Thus, supercomputing has evolved from a hardware engineering design challenge of the Cray Era(1960-1995) of the monocomputer to the challenging of creating programs that operate on distributed (mono)computers of the Multicomputer Era (1985- present).

Introduction

Supercomputers are the highest performing computers of the day that extend the limits of size, technology, power, and buyer budgets to solve technical and scientific "grand challenge" problems. In 1956 the $10M IBM Stretch was aimed at providing more than 100x the speed of existing computers to operate at near one million operations per second. Supercomputer progress is marked by fifty five years of amazing races to build "the world's fastest", *and usually the most expensive,* computer running programs written in FORTRAN that had fortuitously been introduced in 1957. By 2015, hundreds of companies, spending tens of billions of dollars, many of whom lost their lives, have increased performance by over a billion in the race to build the next "supercomputer".

Perhaps the first race started in 1956 resulting in the Univac LARC (1960), IBM Stretch (1961), and University of Manchester-Ferranti Atlas (1962) computers. The origin and creation of the supercomputer class and market is shared by pioneer user demand for the LARC and Stretch at Lawrence National Labs at Livermore and Los Alamos, respectively. Supercomputer cost and size, has limited the market to a few worldwide institutions supporting scientists and engineers working in aerodynamics, chemistry, climate, defense, geology, finance, fusion energy, image processing, life sciences, materials sciences, etc. By 2010 applications have been developed for nearly every engineering and scientific discipline. Without supercomputing and validated models, the scientific community would still be wondering whether climate change is a man-induced phenomenon.

Thus, government funding has played "the" critical role in development as the customer by enabling component research, with the scientific community establishing decade-by-decade "grand challenges" to justify research and development in a perpetual cycle. Beginning in 1997

with ASCI Red at Sandia, the first machine to reach a Teraflops, all of the highest performing computers have been supported by Chinese, Japanese, and US government funding.

"Supercomputer" appeared in print in 1970 and by 1980 was in more common use. It first appeared in the minutes of an NRDC meeting in January 1957 long before any such computer. "Supercomputer" wasn't used to identify or define the first or for that matter any computer as a supercomputer until the 1990s, however as a visitor to the Lawrence Livermore National Laboratory and to the University of Manchester in the early 1960s, I knew a supercomputer when I saw one. In June 1993 the Linpack benchmark was used to establish a semi-annual supercomputers ranking of the worldwide Top500 computers, providing a solid measure. In 2010, Graph500 for performance accessing non structured data and Green500 measuring watts/floating-point operation were added to create a "triple crown". In the twenty-first century, scalable computers exist at virtually all size levels from personal use to national facilities enabling widening use. While programmability and measures of time to solution are discussed as the important metrics, the top three design criteria for supercomputers are speed, speed, and speed.

Two Eras: Cray Monocomputer & Scalable Multicomputer supercomputers

In 1964 with the Control Data Corporation, CDC 6600 delivery to the Lawrence Livermore National Laboratory, Seymour Cray and Cray derivative architectures defined supercomputing for thirty years. Parallel units, pipelining, vector processing by a single processor, followed by multiple processors sharing a single memory characterize the Cray era shown in Figure 1. By 1995, the 32 vector processor multiprocessor, 60 Gflops, Cray T90 was about 64,000 faster than the CDC 6600--*yet the last of the Cray era*. The era was dominated by Cray Research competing with CDC and a few IBM System/360 machines. In the middle of the era, encouraged by their government, Fujitsu, Hitachi, and NEC entered the race using Cray's multiple, vector processor architecture template. The limited scalability of a single memory monocomputer accessed by one or more processors constrained performance thereby requiring an architecture for unconstrained size computers.

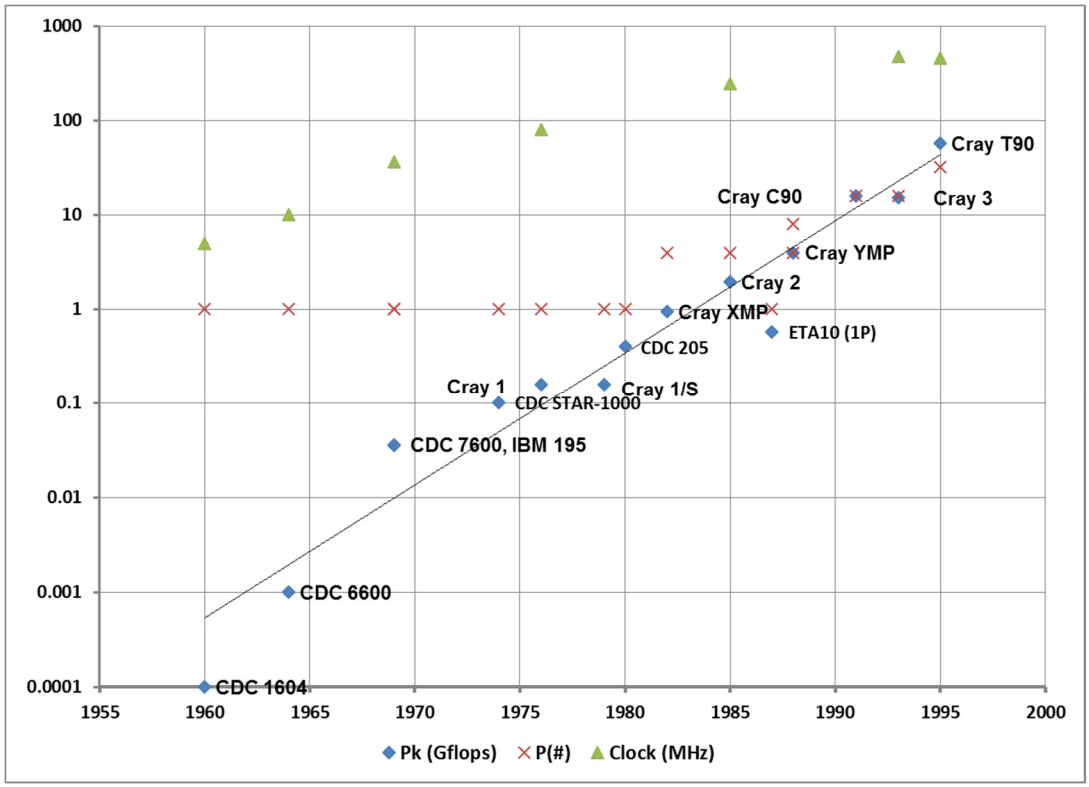

Figure 1. CDC, Cray Research, Cray Computer, and IBM computer clock speed in MHz, number of processors, and performance in billions of floating point operations per second versus first shipment dates.

In 1993 Thinking Machines' 1024 computer CM5 using SPARC microprocessors ran the newly established Linpack benchmark at 60 Gflops to exploit interconnected "killer micros" and mark the beginning of the scalable, multicomputer era. In the same year the 16 processor, shared memory Cray C90 peaked at 16 Gflops. Measuring progress beginning with the 1983 64-node Caltech Cosmic Cube multicomputer, by 2013 the number of computing nodes for the largest machines has increased to over 3 million or 50,000 times for a 43%/year gain in parallelism. Microprocessor clock speed has increased 1000 times from 5 MHz (Intel processor used in the Cosmic Cube) for a annual gain of 26% or a factor of 1.7 per year. The Top500 Linpack benchmarked computers show that the highest performing computers have increased almost 1,000 per decade since 1993 for annual doubling.

In 1994 the US Dept. of Energy Laboratories (DOE) established the ten year Accelerated Strategic Computing Initiative to address the problem of programming distributed memory computers or multicomputers. ASCI was a critical and successful program that included procurement through programming, including attracting IBM to return to high performance

computing. In the future with the multicomputers recipe, performance would come mainly by increasing the number of nodes and fast inter-connecting networks together with microprocessor node advances—and the challenge for speed rests wholly with the programmer. In 2015, the topmost 25 machines are various kinds of multicomputers: Clusters, Constellations (multiprocessors per node clusters), vector processors, or Massively Parallel Processors (clusters with special networks). Lesser performing computers on the Top500 list include shared memory single and multi-vector monocomputers. SIMD computers disappeared. The most recent advance has been the increase in parallelism within a processing node of a cluster with multi-threading and multi-core multiprocessors and then finally attaching Graphics Processing Units (GPUs) with over 100 processing elements.

Table 1 contrasts the machines, challenges and innovations of the single memory, Monocomputer Cray era with the Multicomputer era of distributed interconnected computers.

Table 1. The Cray Monocomputer and Scalable Multicomputer Eras of Supercomputers.

	Cray Era (Monocomputer)	Multicomputer Era
Characterization	One, shared memory, accessed by one or more processors	Multiple independent computers that are interconnected via one or more high speed networks
Date	1964-1993	1983-present
Defining Systems	CDC 6600, CDC 7600, CDC Star, Cray 1 … C90, NEC SX	Caltech Cosmic Cube, Thinking Machines CM5, ASCI Red, …Tianh2
Challenge	Hardware and architecture. Semiconductors. Processor parallelism, memory latency & bandwidth. Multiprocessing and multithreading	Programming a multicomputer. Fast processors and processing elements that form each node. Fast and low latency network interconnections.
Standards	FORTRAN (single memory)	MPI; Beowulf
Gain: Linpack Gain: Clock Gain: Explicitly controlled P's	 10 MHz-500 MHz 13% 64 P	 5 MHz-5 GHz 25%/yr 64-3 million P&PE 43%/yr

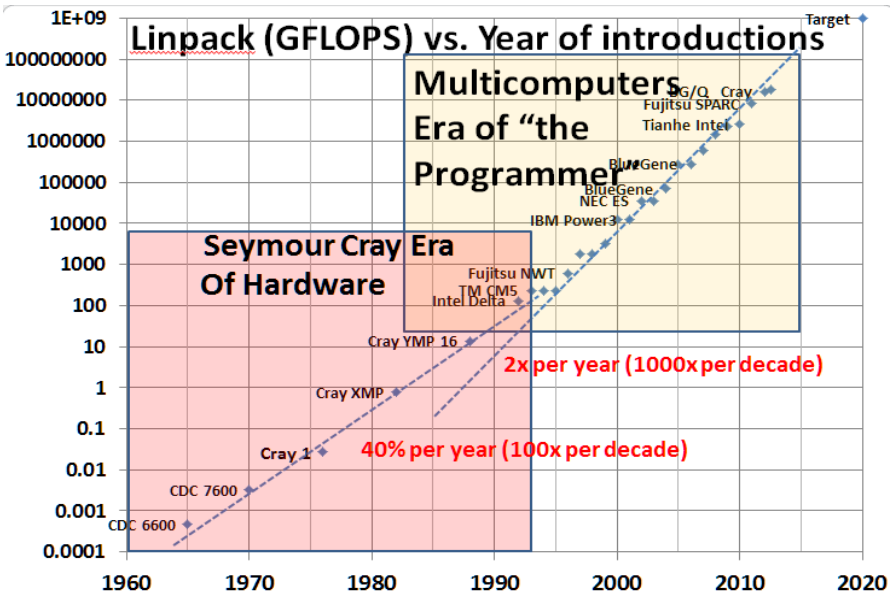

Figure 2 Two supercomputing eras: the Cray, single memory (mono-memory computer), was followed by a decade of overlap beginning in 1983 with the first successful Caltech Cosmic Cube demonstration. In 1993 the multicomputer era was established when the TMC CM-5 operated at 60 GFlops, exceeding the fastest Cray shared memory multiprocessor

Figure 2 shows the extraordinary gain in performance that has occurred over the six decade, two eras—roughly 40% per year or 100x per decade for the Cray single memory era versus a yearly doubling for 1000X per decade for scalable multicomputers. In 1983, a decade-long "ECL-to-CMOS microprocessor" transition was marked by Seitz's Caltech 64 node hypercube connected multicomputer with Intel microcomputers. In the transitional period, nearly 50 companies started up and perished in the search for more effective architectures or the next supercomputer using CMOS microprocessors and stimulated by DARPA's 1983 Strategic Computing Initiative (SCI) program. While searching for speed and/or scalability only a few advances were made until microprocessors became fast enough to challenge the vector processor—typically 8-10 scalar microprocessors are required to equal a vector processor architecture. In 1994 the first MPI (Message Passing Interface) standard *proposal* established a programming model; Sterling and Becker provided a recipe for building Beowulf clusters from standard commodity computers and local area networks.

Since 1993 and for the foreseeable future i.e. 2020 when an exaflops computer is expected, the programming of computers with increasing numbers of computer nodes combined with processing elements dominates the race challenge. Giving up the single memory established by the first FORTRAN in favor of multicomputers clusters using the MPI programing model required over a decade of training, new algorithms and completely new software. Having a common recipe and standard programming model has enabled a Chinese lab at the National

University of Defense Technology, Cray, IBM, Japanese manufacturers, and several more international efforts to compete in the amazing supercomputer race using AMD, IBM, Fujitsu, Intel, and most recently nVidia graphics processing units, all under the operational control of a UNIX dialect operating system.

The Cray, Monocomputer Era (1964-1993)

The early part of the Cray Era of CDC, Cray Research, and IBM computers, people, and company story is told by Elzen and McKenzie (1993). In the early 1980s NEC, Fujitsu, and Hitachi entered the market with competitive computers using the Cray vector architecture recipe. A timeline of the era is given in Figure 3.

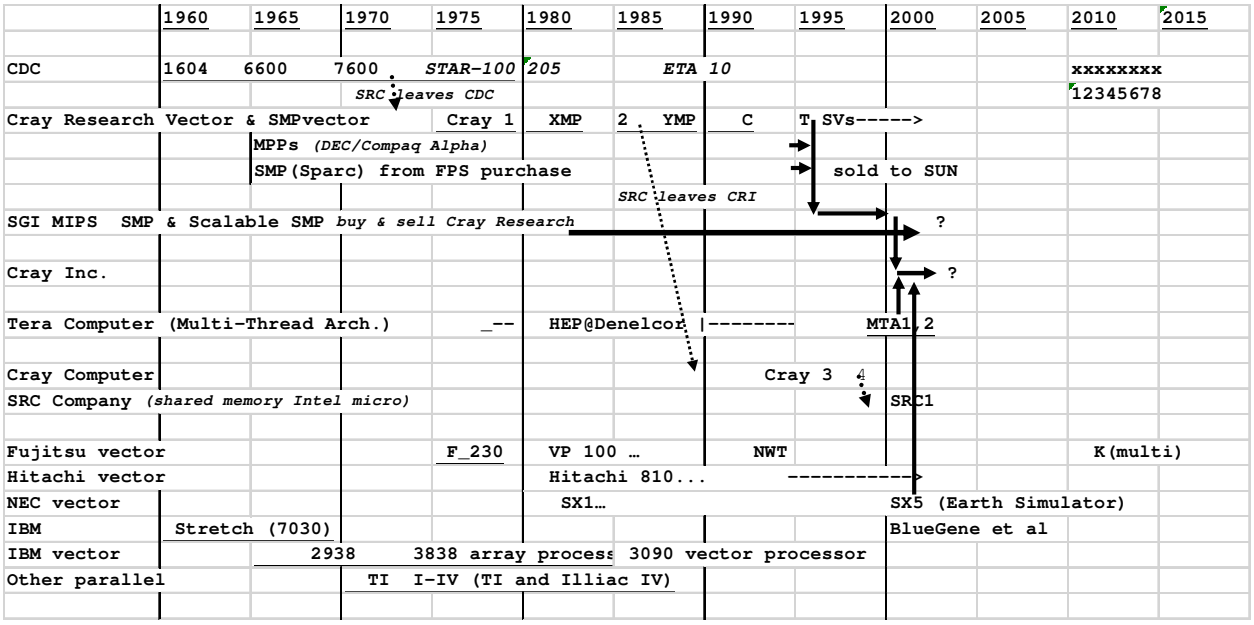

Figure 3. The Cray Era Timeline began with Control Data (CDC 6600) and continuing with Cray Research (Cray 1, XMP, Y, C, and T vector multiprocessors).and Fujitsu, Hitachi, IBM, and NEC vector supercomputers. SGI and Tera had merged with Cray Research, and eventually divested with Tera ultimately acquiring Cray and taking the name Cray Inc.

Conspicuously absent from the story is Remington Rand's UNIVAC Philadelphia division headed by J. Presper Eckert, LARC's designer and the Engineering Research Associates division in St. Paul with their 1107 series. In 1957, Bill Norris, Frank Mullaney, and Seymour Cray left UNIVAC to form the Control Data Corporation in Minneapolis—just across the river from St. Paul. The company's first product, the CDC 1604 with its 160 peripheral minicomputer designed to compete with the IBM 7090, Univac 1107 and a several other large scientific computers was immediately successful.

In 1960, with revenue to build on, Seymour Cray was able to start the 6600 “no more small machines” project by moving to a new lab in his home town of Chippewa Falls, WI 90 miles from CDC headquarters in order to avoid the distractions of a growing company. From the introduction, the 6600 stood as the dominant technical computer, establishing Cray as the world’s foremost computer designer, defining supercomputer architectures for three decades until his untimely accidental death in 1996 at age 71. In an August 1963 memo following the 6600 announcement, IBM’s president T.J. Watson wrote:

> “I understand that in the laboratory developing this system there are only 34 people, "including the janitor." … Contrasting this modest effort with our own vast development activities, I fail to understand why we have lost our industry leadership position by letting someone else offer the world's most powerful computer.”

IBM’s April 1964 introduction of the System 360 line lacked a high performance competitive offering. IBM responded with a 1964 announcement and the System 360 model 91 was delivered in 19x67, but the premature announcement allowed CDC to win the “consent decree” suit against IBM for anti-competitive behavior.

Figure 4 CDC 6600 from LLNL with console at Computer History Museum..

The 6600 (Figure 4) like all of the Cray designs was simply elegant as well as physically imposing, built in the shape of a plus sign. The need for speed required short wire lengths and dense packaging that conflicted with low density packaging that could be air cooled. Each of the 4 bays held four swing out logic chassis with 756 modules holding 600,000 transistors and a refrigeration compressor that cooled the modules by circulating Freon through cooling plates. The cordwood constructed modules were 2.5” square and 1” thick (Figure 5).

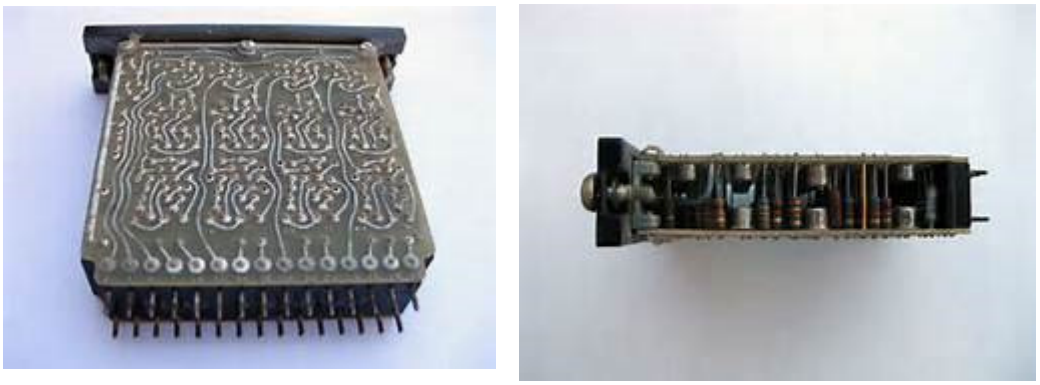

Figure 5. CDC 6600 Cordwood modules with approximately xxx transistors per module of the yy modules,

In contrast to a massive array of lights (one for every flip-flop in the system) and switches used for maintenance and operator's console of Stretch, the 6600 had a separate console with two CRTs that displayed the computer's state.

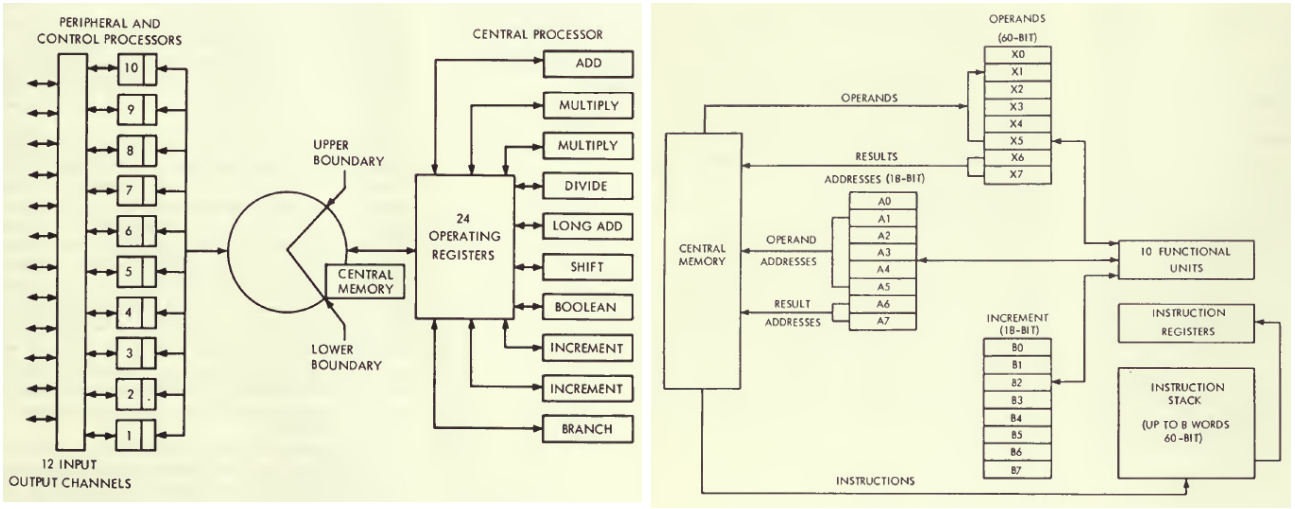

Figure 6. Block diagrams of the CDC 6600 computer with 10 peripheral processsors and the main central processso with registers and functional units

The 60-bit word architecture (Figure 6) was equally innovative and impressive centered around a memory of 128 K words composed of 4K x 12 bit, 1 usec core modules providing high bandwidth. The memory had no parity, as Cray was reported to have claimed "parity is for farmers". Although the clock was 10 Mhz, it had four 25 ns. phases. Ten peripheral processing units or 12 bit computers roughly equivalent to the CDC 160 peripheral minicomputer were interpreted by a single physical hardware unit in a multi-threaded fashion every 100 ns. The main CPU was highly "parallel" with 10 independent function units, yet it had only 3 sets of 8 registers for memory access, indexing (18 bits) and arithmetic operands (60 bits). Data access

and retrieval between memory and operand were implicit by writing the access registers. The function units that operated in parallel included: two floating point multipliers, a divider, adder, long adder, two incrementers for address arithmetic, shift, Boolean logic, and branch. The synchronization of all the function units was handled out by the Scoreboard that assured that instructions were executed and retired in the proper sequence.

In 1969, the 7600 was introduced, providing over a factor of five in performance over the 6600 that came from a faster, 36 MHz clock and the addition of pipelining. Both machines had an Extended Core Storage whereby 8 word blocks from up to 2 million words could be directly transferred into the working registers. Up to four computers could share this large, peripheral memory.

Figure 7 Cray 1 with Seymour Cray. 1

Jim Thornton, who had designed the 6600 Scoreboard, went then to design the CDC STAR. Cray began work on the 8600, a shared memory quad processor computer, extending the 7600 with the ability to run the processors in a coupled mode.

In 1972, CDC's funding of two large projects, and with the 8600 reaching an impasse to achieve Cray's goals, resulted in Cray leaving CDC to found Cray Research. STAR, first delivered to Livermore in 1974, four years later than expected, was perhaps the world's first vector computer. The STAR had a large instruction repertoire that CDC suggested was the hardware equivalent of IBM's APL language. Instructions took vector data from memory, performed an operation, and returned the result to memory in a three address fashion much like early computers. STAR was the forerunner of the CDC Cyber 205 (1980) that was able to gain a small share of the small supercomputer market that was now defined by Cray Research. In 1983, CDC started ETA, a wholly owned company to focus on the small supercomputer market. Their ETA10, a 10 gigaFlops, eight processor based on the STAR

architecture was announced to be up to 10 times faster than the current Cray computers. The ETA10 was built using custom CMOS and immersed in liquid nitrogen, providing a doubled clock speed. Unfortunately the company could not gain traction in light of CDC's own financial problems at the time. As a result, ETA was closed down in 1989 and CDC was eventually sold.

While it might be argued that the ETA was a technical success because the ETA10 did operate, it is equally arguable that the STAR was a fatally flawed architecture from the beginning that was found guilty of violating Amdahl's Law with computation being determined by the slowest sequential parts. Its ability to operate on scalars was also slower than the earlier 6600 and 7600. However, the 1976 delivery of the superior Cray 1 to Los Alamos was the competitive reason for the demise of CDC's supercomputers. On scalar problems, the newly introduced Cray 1 outperformed the 7600, adhering to the rule that to replace a previous generation computer, there can be no corner cases whereby the new computer is slower.

Steps leading to the vector architecture

A 1963 memo on the Solomon architecture by John Cocke of IBM and Daniel Slotnick outlining an array of 1024 interconnected processing elements, each with 128 32-bit memories and a Livermore RFP stimulated Slotnick's U of Illinois ILLIAC IV and the Texas Instruments Advanced Scientific (or Seismic) Computer (ASC). Slotnick, proposed a 4 x 256 processor array, each with its own 2K, 64-bit work memory and attached 10 MByte fixed head disk storage. Burroughs delivered one quadrant of the contracted machine in 1972 to NASA's Ames Research facility and the machine became operational in 1975 (concurrent with the Cray 1) and was decommissioned in 1982. ILLIAC IV was the first of the SIMD (single instruction, multiple data) execution architectures. The machine was used for specialized calculations including a prime number sieve. Twenty years later, Thinking Machines using custom CMOS processing elements, built their Connection Machine, SIMD CM-1 and 2 as part of the DARPA Strategic Computing Initiative (SCI), but abandoned the architecture due to lack of generality i.e. a small, limited set of applications not constrained by Amdahl's Law. At least two other smaller SIMD computers, MasPar and Wavetracer were built in the 1980s as part of the "CMOS Gold Rush". While the SIMD architecture works well on specific problems it fails both flexibility and generality. As a result, it is only applicable to specific applications requiring performance or cost/performance.

However, in 2014, over 10,000 Parallela SIMD chip computers operating at 100 GFlops, consuming only 5 watts, costing about $100 each have been delivered. The company was crowd funded and has an enviable Green Record in terms of flops/watt.

TI began an ASC project in 1964 and the computer became operational in 1971. As a memory-to-memory vector architecture similar to the CDC STAR, a single pipeline for arithmetic was fed by four memory fetch-store processors. Six computers were built for the Seismic market and the company left the business when the Cray-1 was announced.

In 1968 IBM introduced the 2938 array processor for Geological processing that operated at 10 MFlops following it with the 3838 in 1974. While IBM Research built several computers during the 1980s, vector processing wasn't introduced into their 3090 mainline until 1985.

Multithreading

In 1974, Burton Smith started down a lifelong path to explore the multi-threading architecture (MTA) whereby multiple independent program threads are executed in quasi parallelism by a single physical processor—operating in the same fashion as the barrel for the CDC 6600 Peripheral Processing Units. With many threads, the idea is to recover time lost waiting for memory and blocked processes with a long pipeline from memory though execution and back to memory. By tightly coupling the threads, programs can effectively work on arrays. The only fault of such an architecture is that on the non-parallelizable parts, the scalar instructions take a long time to execute—hence like the SIMDs, is plagued by Amdahl's Law. The first Denelcor HEP became operational in 1979, with six HEPs being constructed when the company folded in 1985.

In 1987, Burton and James Rottsolk opened Tera Computer in Seattle to continue with the multi-threading architecture providing an ideal multiprocessor with constant access time, independent of the number of processors. The company struggled to demonstrate the efficacy of their multi-threading architecture implementation using Gallium Arsenide (GaAs) logic. In 1998, an 8 processor MTA was delivered. In 2000, Tera's breakthrough was to buy the Cray Research division from Silicon Graphics that the supercomputer company had merged with in 1996. Tera wisely changed the name of the company to Cray Inc. to build on the Cray brand.

Cray Research established the vector supercomputer recipe for two decades

With the Cray-1 introduction in 1975 it was clear that Seymour Cray had found the recipe for high performance computing using vector register memories. Three newly introduced ECL integrated logic circuits—a two NOR circuit chip with 5 & 4 inputs, a 6 ns. 4x16 register memory to hold vectors and a 1 x 1,024 bit fast *main* memory from Fairchild and Motorola enabled the Cray 1. The fast memory with extraordinary bandwidth, fast scalar speed and instructions operating on vector registers established the design and established differentiated it from STAR. The Cray-1 clock was 80 MHz, and the peak speed was over 160 Mflops. According to the Wikipedia site, 80 machines were sold at a price of $5-$8 million. Like its predecessors, the Cray-1 was elegantly plumbed, powered, and packaged in a C-shape set of twenty-four 28" high Freon cooled racks holding 72 densely packaged module boards. The final stage power supplies of 114 Kwatts were placed at the base of the computer that Cray described as the world's most expensive "love seat".

Convergence on the Cray vector architecture

Figure 8 shows the magic of the Cray vector processor architecture and principles that were used for the various machines. In the earliest Cray computers, the memory-processor connection provided bandwidth for at least two operand loads and one operand store. Thus, for a peak of 80 Mflops, 240 million words per second are required. The 50 ns. memories could deliver 20 Million, 64-bit word accesses per second, and the 16 banks provided 320 million words per second (2.56 GBytes/sec). The Cray 1 had 2-4 high speed I/O channels fed by 16 bit computers operating at 80 MHz. While the scalar performance was several times the 7600, the real performance gain came from the vector register architecture that allowed vectors to be loaded and stored in parallel, together with operations on data in the vector registers--in contrast to the CDC STAR memory to memory vector architecture that required accessing main memory access for every operation.

The Cray-1 vector register architecture implemented with bipolar LSI chips held for two decades until challenged by CMOS microprocessor multicomputers. The timeline (Figure 3) shows that other companies and research had been competing to build large numeric machines for over a decade.

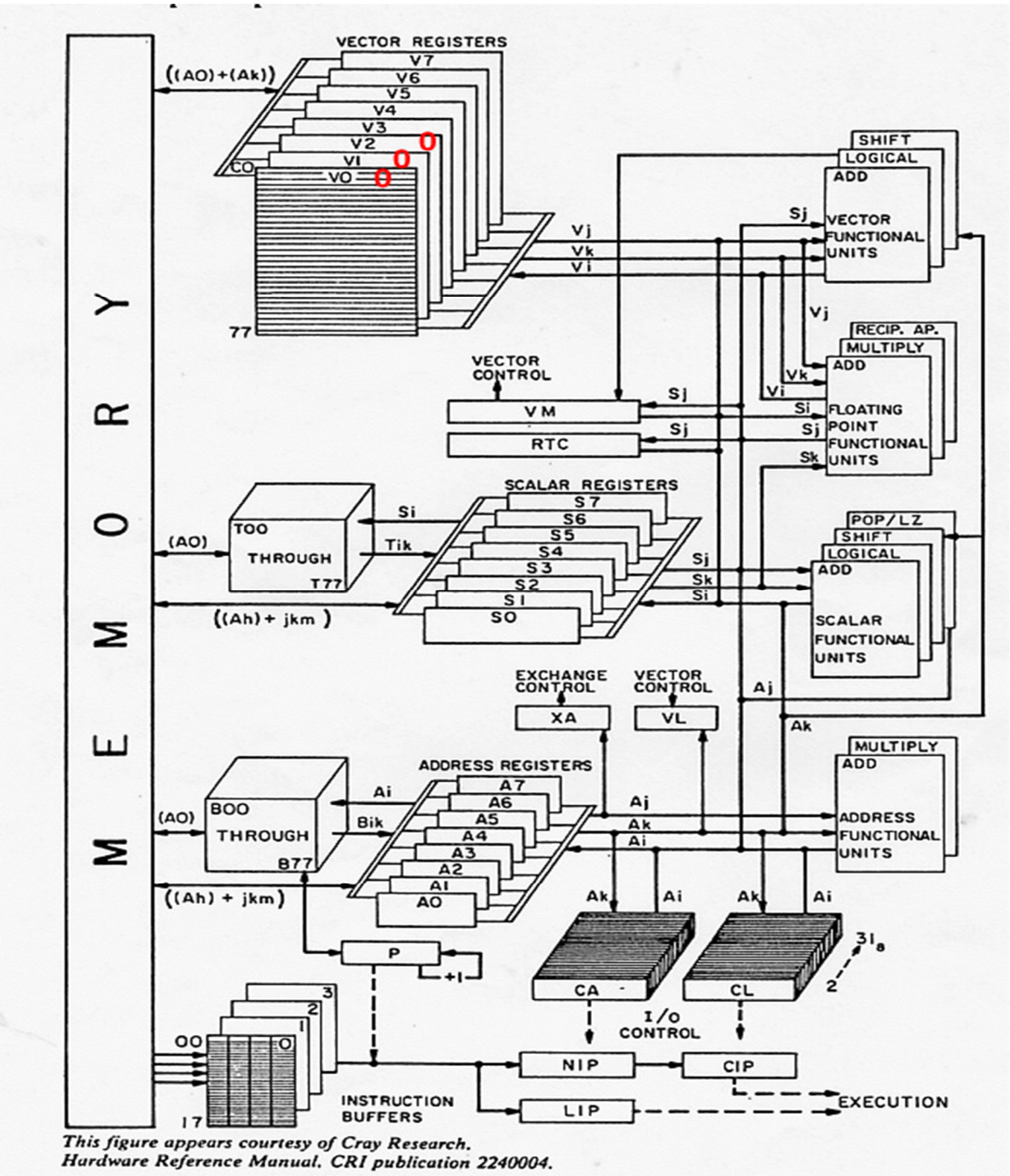

Figure 8. Cray 1 vector processor architecture diagram shows the vector register "magic" of the computer: the processor is fed by a high bandwidth memory of 16 independent banks. A single instruction load or store specifies a vector of up to 64 words and an instruction operates on a vector held in the registers. Operations occur in parallel using the pipelined arithmetic and logic units. Arithmetic operations can also by "chained" to avoid temporary storage and retrieval.

In 1982 the Cray XMP 4 processor, shared memory, designed by Steve Chen was introduced, operating at a peak of 800 Mflops with a 105 MHz clock. The YMP was followed by: the 8 processor YMP operating at a peak of 4 Gflops in 1988; the 16 processor C90 in 1991; and 32 processor T90 in 1995 with a peak of 64 Gflops clocked at 450 MHz and costing almost $40 million to end the era of the single memory, multiple vector processor.

By the mid 1980s. Fujitsu, Hitachi, and NEC all entered the supercomputer market with vector supercomputers. Fujitsu working with the Japanese Aerospace Laboratory delivered two FACOM-230s vector processor in 1976. Hitachi introduced their S-810 in 1982 operating at 800 Mflops while NEC delivered their first vector supercomputer, the SX-1 in 1983. The *highest performance Japanese computer aka THE* supercomputer was delivered by Fujitsu to the National Wind Tunnel (1993-5), Hitachi (1996) followed by NEC's Earth Simulator (2002-4), and Fujitsu SPARC in 2011.

Cray Computer and SRC

After the Cray 1, Seymour began work on the incompatible Cray 2 (Figure 9) that featured shared memory with 4 processors and delivered in 1985, four years after the Cray 4 processor XMP. The Cray 2 was cooled by Fluorinert circulating through the modules in what was called the most expensive aquarium.

Figure 9 Cray 2 with Fluorinert cooling

In 1989, repeating history, Cray Research was unable to fund two major developments and Seymour formed Cray Computer in Colorado Springs to finish the Cray 3. It had a 2 ns. clock, constructed with GaAs logic on cordwood modules immersed in liquid. A Cray 4 with 1 ns. clock was also constructed. The SRC company was formed to build a large shared memory multi-microprocessor using the Intel Itanium when Cray was killed in a 1996 automobile accident.

The CMOS Goal Rush: The transition to multicomputers (1983-1993)

Beginning in 1982, five events began to cause the transition from proprietary shared memory ECL vector multi-processors to low cost, powerful distributed CMOS multicomputers. Numerous startups were exploiting the newly introduced 32-bit CMOS microprocessors by building computers of all shapes and sizes to compete with minicomputers through mainframes. Mead and Conway custom VLSI design methodology was being established in universities. The Japanese Fifth Generation research program was initiated to build parallel machines for AI. A year later, DARPA responded with the Strategic Computing Initiative (SCI) research initiative and funding aimed at scalability and parallelism for AI.

In 1982, the Caltech Cosmic Cube, with eight followed by 64 independent hypercube interconnected Intel 386 board computers, demonstrated the programmability, utility and overall cost-performance effectiveness of the distributed multicomputer architecture. Inspired by the Cosmic Cube, Ametek and nCUBE started up to build similar scalable architectures. Intel as a supplier of microprocessors, adopted the idea, introducing its iPSC line of scalable multicomputers in 1985 and continued development and limited marketing to the government into early 2000. In addition to the three Caltech inspired companies, in 1984 Inmos (UK) introduced the Transputer as a scalable microprocessor with serial message passing links that could connect to other Transputers to form multicomputers. These were used extensively in process control and embedded applications but was also adopted by Meiko Scientific (UK), Parsytec (German), and Floating Point Systems (US) for parallel processing.

Cul de sac in the scalability search: Fifty mini-supers, super wanna-be's, super-minis, and other architectures 1985-1995

Alliant, American Supercomputer, Convex, and Supertek (becoming part of Cray) adopted the Cray vector mP architecture introducing lower priced computers for the technical market in a fashion similar to the use of minicomputer to substitute for mainframes. By the early 90s, DEC added a Vector facility to their VAXen computers. Ardent and Stellar followed in the late 80s by introducing even lower priced personal supercomputers. For example the Ardent 4 processor Titan with MIPS microprocessor, Weitek floating point, and several custom VLSI chips operated at over 100 Mlfops or roughly the speed of the decade old Cray 1. Two thousand computers

were sold that delivered more floating point operations per month, than Cray Research computers.

Besides the three Japanese supercomputer and the IBM 3090 Vector Facility introductions there were three abortive efforts included Supercomputer Systems Inc. by Steve Chen with IBM backing, Evans and Sutherland, and Scientific Supercomputer Systems. The French Advance Computer Research Institute (ACRI) and German Supernum national efforts came to naught.

Other startups used either multiple CMOS microprocessors or custom CMOS to address the super minicomputer market. “Multis” or multiple microprocessors included BBN, Encore, Flexible, Floating Point Systems (sold to Sun, creating a web server business), Myrias, Sequent, Stratus, and Synapse. During this period Cydrome, Elexsi, and Multiflow introduced very long instruction word (VLIW) architectures.

A half dozen university efforts based on parallel microprocessors besides the Cosmic Cube helped train a number of researchers. These included Berkeley’s Network of workstations; Carnegie Mellon’s Cm* computer modules and C.mmp multiprocessor; Cedar at the University of Illinois; DASH at Stanford that eventually fed the SGI architecture; and several machines at IBM Research. Columbia Universities Dado and non-Von multicomputers headed by David E. Shaw is noteworthy because David subsequently used the multicomputer structure he had researched, in high speed and quantitative trading that certainly captured more money, than the entire profit derived from all of the sales of supercomputers!

The important lesson of this transitional period is all companies failed to become independent and perished and at most two lived to be acquired. Only Cray, Fujitsu, Hitachi, IBM, and NEC were large enough to afford supercomputer development. From my architect’s perspective: SIMDs are inherently limited, multi-threading is intolerant of slow scalar speed, VLIW is ineffective compared to the vector architecture, and in general, slow computers aren’t marketable. While the issue of programmability and measures of time to solution are touted, speed, speed, and speed are the design criteria!

Measuring Progress: Goals, Benchmarks, and Rewards

Progress in computer performance demonstrates "you get what you measure" and the "power of prizes". In December, 1985 Alan Karp made the following challenge:

> I will pay $100 to the first person to demonstrate a speed-up of at least 200 on a general purpose, MIMD computer used for scientific computing. This offer will be withdrawn at 11:59 PM on 31 December 1995.

In 1987, I raised the ante to $1,000 (now $10,000), creating the annual Gordon Bell Prize that continues indefinitely, administered by the ACM and judged by a committee that Alan chaired for the first five years.

In 1986 Frank McMahon established the Livermore Fortran kernels (LFK) or Livermore loops as a benchmark for parallel computers with 24 kernels coming from various codes with varying degree of parallelism, which typified the lab's workload. Both the average and harmonic mean are given with the harmonic a brutal measure that extracts heavy penalties for just one, slow loop. Ironically, the ETA10 was rejected by its first customer based on the harmonic mean acceptance criteria that its design was impossible to meet.

In 1987 a team at Sandia won the first Gordon Bell Prize for parallelism using a 1024 computer nCUBE demonstrating that when problems could be scaled enough, Amdahl's Law could be thwarted, establishing Gustafson's Law (1988). The announcement of the event helped convince scientists of the multicomputer approach. By 1991, with almost a decade of results that pointed "the way", Eugene Brooks of Livermore proclaimed the "attack of the killer micros".

In 1992, Linpack semi-annual listing of the world's Top500 computers established the standard metric and ranking. The ranking resolved some performance ambiguity of the chaotic market when a manufacture's advertised speed became known to users as the "guaranteed not to exceed speed". In 2010, the Graph500 benchmark was added that was likely to be more typical of jobs with more random memory access especially machine learning. Effectiveness measured in ops/watt got added as the Green500.

The other 499 of the Top500: Operations per day (*c*apacity) versus Operations per second (capability)

There are 499 other computers in addition to the top one (or two) that is our subject. Table 2 shows numbers of computers, cores, and aggregate petaflops delivered to the 500 sites by the top 7 supercomputer vendors.

Table 2 Top500 Market share of top 7 vendors, November 2013, with the number of millions of cores or processors and the aggregate peak petaflops coming from each vendor. For example NUDT with 4 computers delivers more aggregate petaflops than 46 computers from SGI, Bull, Fujitsu and Dell or about the same as all the computers from HP and Cray.

Company/Org.	Number	Share	Cores (Millions)	Peak Petaflops
HP	196	39.2	3.79	67.9
IBM	164	32.8	7.15	105.4
Cray	48	9.6	2.4	57.8
SGI	17	3.4	0.55	10.4
Bull	14	2.8	0.45	7.7
Fujitsu	8	1.6	0.92	14.9
Dell	7	1.4	0.55	9.97
NUDT (China)	4	0.8	3.37	61.3

With such extraordinary concentration of power in a few very large systems, effectiveness is the operational challenge simply because nearly every program reaches a point of diminishing return whereby additional processors can decrease performance. Thus while supercomputers are justified on the *capability* to solve a single problem in the shortest time, operationally, each computer is measured by *capacity,* the total amount of work it does on all jobs. For example, if a problem only scales to utilize a maximum of 10,000 processors, then a 100,000 processor computer has the capacity to run 10 such jobs simultaneously. All supercomputers are run in a batch fashion with relatively static partitions for different sized jobs. Often, running ten limited parallelism jobs in parallel for parameter cases or Monte Carlo experiments allows a user to get work done 10 times faster.

The Multicomputer Era (1993-present)

By 1991, with almost a decade of results that pointed “the way”, Eugene Brooks of Livermore proclaimed the “attack of the killer micros“. In June 1993 the first report of the world’s fastest Top500 computers (measured by the Linpack benchmark) was issued and results marked the

clear transition to multicomputers, signaling the end of the Cray shared single memory architecture (Figure 1, Table 1). A Thinking Machines CM-5, 1024 SPARC microprocessor multicomputer at Los Alamos ran at a peak of 160 Gflops, the nearest competitors were a 4 processor NEC SX-3 at 26 Gflops; while the 1991, 16 processor Cray YMP (C90) operated at 15.2 Gflops. Then in November 1993, Fujitsu delivered a 140 node vector processor multicomputer with very tightly interconnected nodes and system with global addressing to the National Wind Tunnel with a peak of 260 Gflops and held the record until June 1996 when a Hitachi 1024 node vector processor multicomputer at Tokyo U won first place. By 1995, the 32 processors Cray T90 multiprocessor could almost achieve 60 Gflops at a cost of $39 Million. One was delivered.

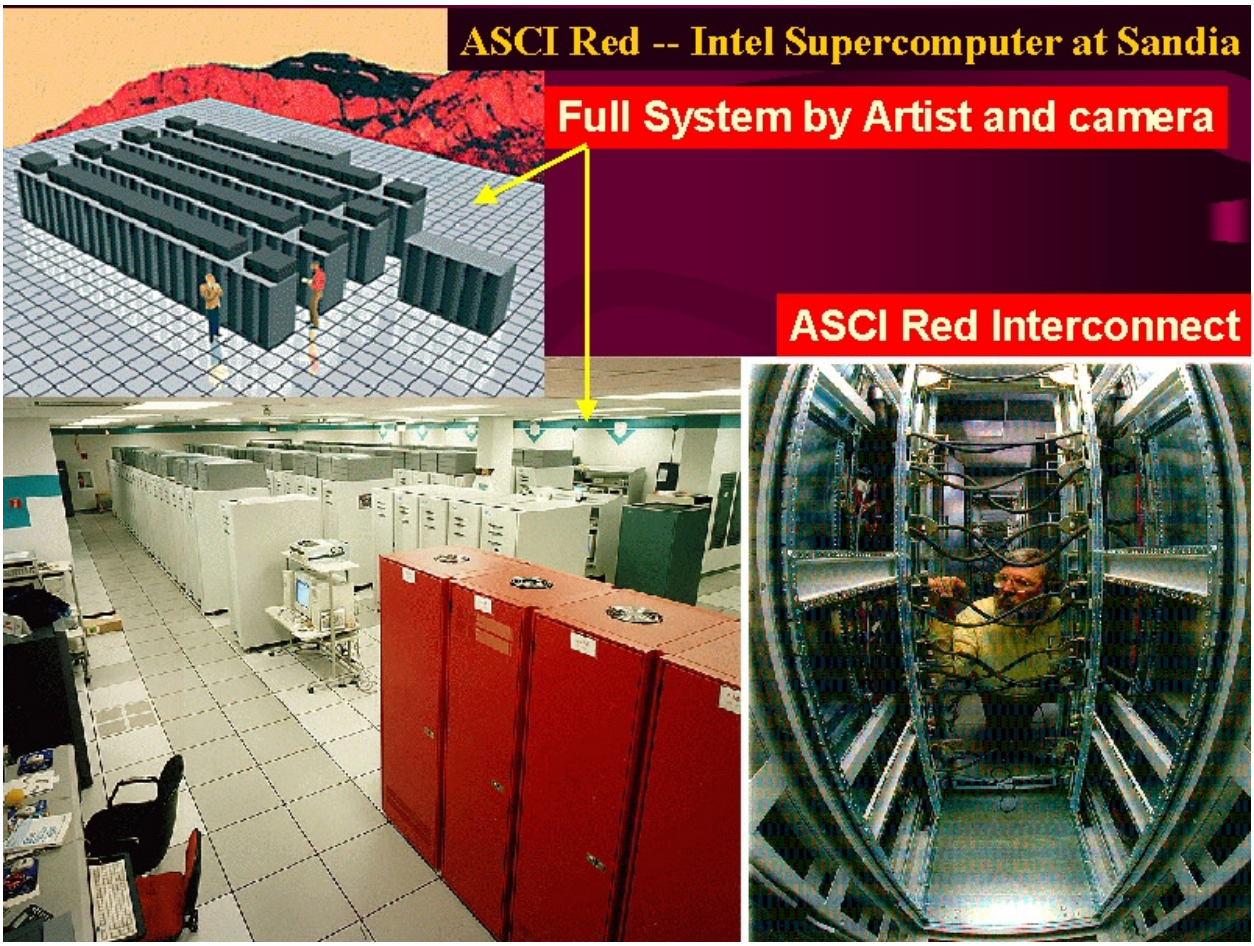

Figure 10 ASCI Red, first computer to reach a Teraflops.

It was the situation at the Dept. of Energy and the intelligence community, seeing no U.S. supplier on the horizon that prompted ASCI-- Accelerated Strategic Computing Initiative that resulted in the switch to CMOS multicomputers. The win for killer micros came with the June

1997 Top500 list when Intel's ASCI Red (figure 10) with 7,264 computers benchmarked at just over one teraflops and maintained the position until November 2000 when, with upgrades, it had peaked at 3.1 teraflops using 9,632 Intel micros. Red was housed in 74 cabinets with 64 dual processor computer nodes occupying 1600 sq. ft. and consuming 850 KWatts. A mesh of 2 x 34 x 38 interconnects the computers (figure 11). Red was decommissioned in 2006.

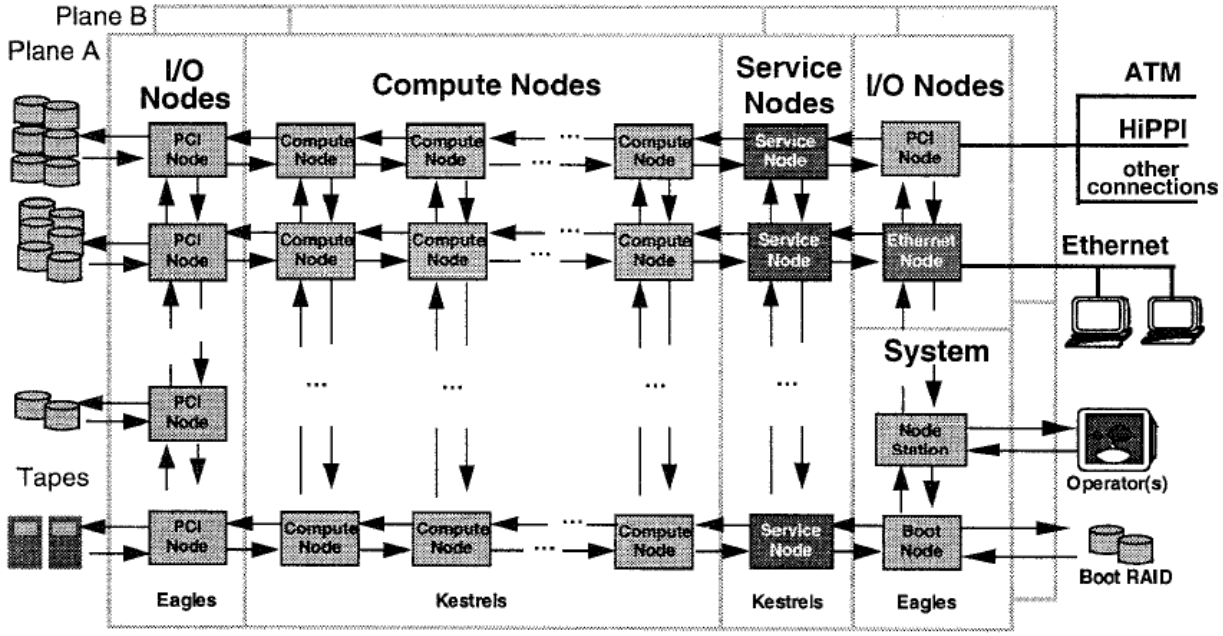

Figure 2: ASCI TFLOP System Block Diagram. This system uses a split-plane mesh topology and has 4 partitions: System, Service, I/O and Compute.

Figure 11 ASCI TFLP system block diagram with 4 partitions. System, Service, I/O and Computer

The Multicomputer Era's formative events

Like the previous decade that opened up scalable computing, three standardization events contributed to the productive decade to establish it. Linpack combined with the Top500 established measures and goals.

In 1994 the first message passing interface (MPI 1.0) standard that had begun in 1991 was established. MPI specified how communication and synchronization was carried out among multicomputers. Libraries were available in C, C++, and FORTRAN with additional language adoptions following.

Sterling and Becker's Beowulf cluster software became available to enable laboratories to interconnect commodity personal computers with Ethernet switches to form a parallel supercomputer. The Beowulf software with Linux, included both the MPI and parallel virtual machine (PVM) libraries. The software enabled wide scale training and use in much the same fashion as Mead-Conway enabled CMOS design a decade earlier. Various texts have been written and many laboratories run Beowulf for scientific computing.

The ASCI program was the crystallizing event that made the transition to multicomputers happen by providing a clear market, linking design with use and a strong focus on software.

Table 3. Multicomputer era with decade transition to multicomputers beginning in 1982 at Caltech, followed in 1987 by 1024 nCUBE at Sandia, and 1024 computer Thinking Machines at Los Alamos that provided performance proof of the CMOS microprocessor multicomputer approach. Since 1992, only Fujitsu, NEC, and Hitachi built multicomputers using vector processors with 10x the power of microprocessors.

Yr.	R_{max}(Gflops)	R_{peak}(Gflops)	Cores	Kwatt	Computer
1982	**1**	**1**	**4**		**Cray XMP mPv**
1983		?	4>64		Cosmic Cube @Caltech
1987		?	1024		nCUBE @Sandia Nat. Lab.
1988	14	15.2	16		Cray YMP 16 (C90) mPvect
1993-4		20	512		Intel Delta/Paragon @Caltech
1995		64	32		Cray T90 32 mPvect
	Top500 Listing of computers started June 1993				
1993.6	60	131	1024		TM CM5 @LANL
1993.9	124	236	140		Fujitsu mCvect @NWT
1994	124	236	140		
1995	124	236	140		
1996	368	614	1024		Hitachi mCvect @Tokyo U.
1997	**1,338**	**1,830**	**9,632**		**ASCI Red Intel @Sandia**
1998	1,338	1,830	9,632		
1999	2,400	3,200	9,632	850	(upgraded)
2000	4,938	12,288	8,192		ASCI White IBM SP3 @LLNL
2001	4,938	12,288	8,192		IBM Power3
2002	35,860	40,960	5,120	3,200	NEC Earth Simulator mCvect.
2003	35,680	40,960	5,120	3,200	(June 2002-June 2004)
2004	70,720	91,750	32,768		IBM BlueGene/L @LLNL
2005	281,000	367,000	131,072	1,433	
2006	281,000	367,000	131,072	1,433	
2007	478,200	596,000	212,992	2,329	
2008	**1,105,000**	**1,456,000**	**129,600**	**2,483**	**IBM Roadrunner Blade @LANL**
2009	1,759,000	2,331,000	224,000	6,950	Cray Jaguar 6 core Intel @ORNL
2010	2,567,000	4,701,000	186,386	4,040	Tianhe Intel @China NUDT
2011	10,510,000	11,280,000	705,024	12,659	Fujitsu SPARC
2012	16,325,000	20,132,000	560,640	8,209	Cray XK7 Intel & GPU @ORNL
2013	33,860,000	54,902,000	3,120,000	17,808	Tianhe2 Intel @China NUDT
…		…			…

2020	**1,000,000,000**	**Target**

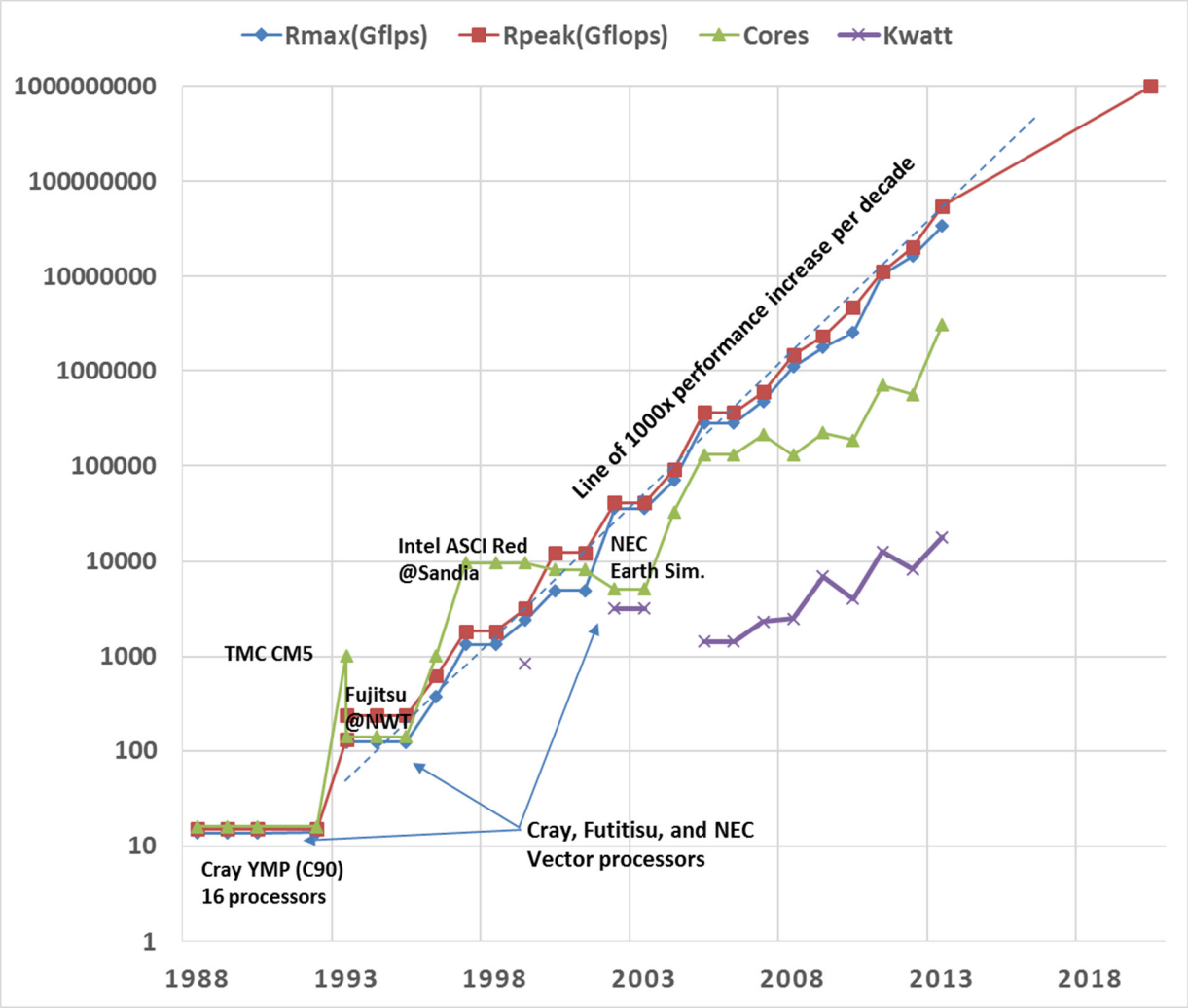

Figure 12 Linpack performance, peak performance, and number of cores of computers 1988-2020 showing the transition of monocomputer, vector supercomputers to the multicomputer era 1992 to present. Peak speed increased from 16 Gflops Cray supercomputer to the petaflops in 2008 and projected exaflops in 2018.

ASCI- Accelerated Strategic Computing Initiative

Given the situation in November 1993, a need for a domestic supercomputer supply, and the underlying need for Science Based Stockpile Stewardship (SBSS), undersecretary for defense, Vic Reis mobilized the effort for the next decade with a comprehensive plan with clear goals and cooperative operational principles across national labs, vendors, and other contributors by covering the future generations and especially the programming environments, tools, and

applications with a focus on materials modeling. Dona Crawford's commissioned report (LARZELERE, A. R., 2009) describes this important and productive decade.

In September 1994 ASCI was kicked off with a workshop of potential suppliers met (Cambridge Parallel Processing, Convex Computer Corp., Cray Computer Corp., Cray Research Inc., IBM, Intel SSD, Meiko Scientific Inc., nCUBE, and Thinking Machines Corp). A year later Intel had been selected for the Red platform that demonstrated teraflops performance in December 1996 and June 1997 delivery to Sandia. ASCI Red held the performance record until November 2000, when IBM's ASCI White at Livermore won and held the record for two years. It was followed by IBM Blue Gene/L that held an unusually long four year record from 2004-2007 that achieved a half petaflops with 200,000 cores. NEC's Earth Simulator, a multicomputer with just 5120 vector processors but operating at almost 40 Gflops was off the curve and won the title for two years June 2002-2004 that again caused national concern even though the Japanese plans had been public before 2000. With new supercomputers, schedule uncertainty is almost certain.

In May 2008, one petaflops was achieved with an IBM Roadrunner supercomputer at Los Alamos with 129 thousand cores composed of a combination Intel and 8 processing element cells. Roadrunner was decommissioned in 2013. The computer was most 3-4x larger than Red: 300 vs 74 cabinets; 6,000 vs 1600 sq. ft.; 2.35 vs. 0.85 MWatts; and 129,600 processing elements in 6,480 dual core microprocessors and 12,960 IBM power cells with 8 processing elements or 19,400 chips vs 9160 microprocessors. It might be argued that Roadrunner's memory of 100 TByte of RAM was small, since Red had 1.2 TByte and was a factor of 1,000 slower. Infiniband was used as the central switch.

Since 2005, the winners are almost decided by budget and power consumption: Cray (using Intel multicore and nVidia graphics processing units) at ORNL, IBM with a multicore PowerPC, Fujitsu with Sparc vector processors.

The asymmetry of machine nodes composed of a microprocessor and co-processing elements such as a GPU adds another programming challenge. The Livermore's IBM Sequoia, June 2012

top500, was subsequently displaced by more asymmetrical computers, all with over a million elements.

Figure 13 Livermore Sequoia

Sequoia compute nodes are straight forward multiprocessor computers with a 16-core PowerPC and 16 GBytes of memory (Figure xx, with author) Thus the system contains in total 1,572,864 processor cores (96 cabinets * 1024 processing nodes/cabinet * 16 cores and 16 GBytes of memory per processing node) for a total of 1.5 PByte memory covering 3000 square feet. The computer nodes are interconnected in a 5-dimensional torus topology via Infiniband switch. The net result is that while being displaced in the Linpack500, Sequoia remains the fastest Graph500 computer by a wide margin.

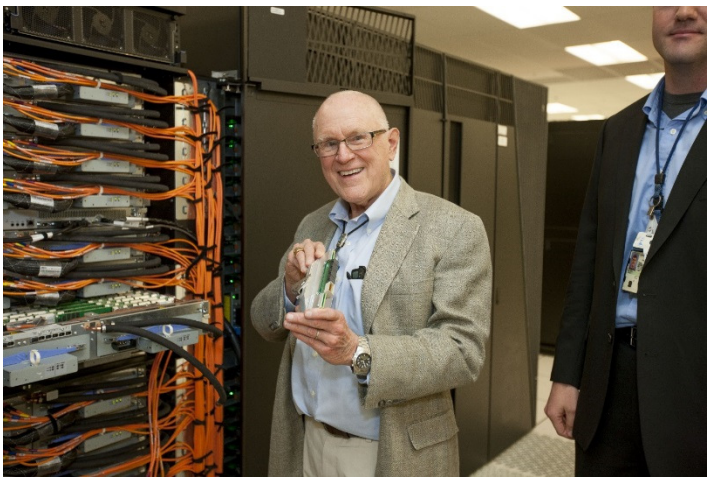

Figure 14 Author holding computer module at Livermore Sequoia in May 2013.

In 2013 Tianhe 2 at the Chinese National University of Defense Technology operates at a peak speed of 50 petaflops using over 3.12 million cores and 12.4 PBytes of memory consuming 17

Mwatts making it number one of the Top500. Tianhe's 16,000 computation nodes are composed of two Intel Xeon E5 and three Xeon Phi (Many Integrated Core chips with 88 GByte memories. Each 8 GB memory computational Phi has 50+ cores and operates at over a teraflops. Nodes are connected with a fat tree topology switch. Tianhe's Graph500 ranks 6.

Cloud computing: storage, processing, and network bandwidth

Cloud computing is the most rapid growing segment of computing with many companies supplying various levels of service coming from: Amazon, Google, IBM, Microsoft, Rackspace as well as Facebook, LinkedIn and Twitter that operate social networking sites. Cloud computing has been defined by web server evolution from a few multiprocessors to host web sites in the early 1990s to computer clusters with thousands of computers with disks and solid state storage consuming up to 10- 50 MWatts. Clouds host a wide range of sites: for searching, email, and office services e.g. Google, Microsoft; commerce stores e.g. Amazon; social networking with over a billion users e.g. Facebook, twitter; or thousands of web properties e.g. Amazon, Microsoft.

In 2006 Amazon introduced Amazon Web Services with the Elastic Computer Cloud EC and Simple Storage System. S3. In 2014 AWS consists of 10 centers with what is estimated to be about a half million computers. Building from their commerce platform the company offers several service levels- XaaS: IaaS--Infrastructure (collections of servers) and PaaS—Platform (programming environment services e.g. SQL, NoSQL, Windows servers). Salesforce.com operates a higher service level with applications specific software that defines: SaaS Software as a Service for what was traditional software for managing sales organizations.

By 2013, cloud services began taking a position in High Performance Computer centers in order to store, manage and interpret the large data sets that come from simulation and real time data that typify HPC use. The data sources provide the scientific component for the "big data" and "visualization" efforts. From a distance, cloud and HPC clusters look fairly similar: HPC requires the highest possible inter-computer bandwidth and minimal local storage whereas commercial clouds nodes have large, attached stores.

Onward to exaflops

By extrapolation of multicomputers reaching a teraflops and then a petaflops in 10 or 11 years, the exaflops would be reached in 2018-2020. China and Japan both have national plans to deliver exaflops machine by 2018-2020. Japan's record of planning and execution is unmatched, and designers are laying out plans e.g. Fujitsu has described their chip architecture for a 2015, 100 petaflops computer. No doubt one or more petaflops machines will exist by 2020. They are likely to consume 50-100 Mwatts of power and have over 10-100 million processing elements each operating at 10-100 Gflops.

References of Historical Significance

Amdahl, G.M. "Validity of the single-processor approach to achieving large scale computing capabilities". In AFIPS Conference Proceedings vol. 30 (Atlantic City, N.J., Apr. 18–20). AFIPS Press, Reston, Va., 1967, pp. 483–485.

Dongarra, Jack J.; Luszczek, Piotr; Petitet, Antoine (2003), "The LINPACK Benchmark: past, present and future", Concurrency and Computation: Practice and Experience (John Wiley & Sons, Ltd.): 803–820

Elzen, B. and Donald MacKenzie, D. "From Megaflops to Total Solutions: The Changing Dynamics of Competitiveness in Supercomputing," in Technological Competitiveness: Contemporary and Historical Perspectives on Electrical, Electronics, and Computer Industries (Piscataway, NJ: IEEE Press, 1993), 119-151.

Gustafson, JL Gary Montry, and Robert Benner "Development of PARALLEL METHODS for a 1024 Processor Hypbercube", SIAM, Vol 9 no. 4 July 1988.

Larzelere, A. R. "The History of the Accelerated Strategic Computing Initiative (ASCI), Lawrence Livermore National Laboratory, UCRL-TR-231286. 2009 Foreword by Dona Crawford, Assoc. Dir. For Computation. Chronicled 5.2B program, specified in Accelerated Strategic Computing Initiative: Strategic Planning Document, July 1994;

Sterling, Thomas, Donald J. Becker, Daniel Savarese, John E. Dorband, Udaya Ranawake, and Charles V. Packer. "Beowulf: A Parallel Workstation for Scientific Computation," in P. Banerjee, ed., Proceedings of the 1995 International Conference on Parallel Processing (ICPP) (August 14-18, 1995), Vol. I, p. 11-4.

References for Further Reading

Office of Science and Technology Policy (OSTP), "A Research and Development Strategy for High Performance Computing", Executive Office of the President, November 20, 1987

Sterling, T., Messina, P. and Smith, P. "Enabling Technologies for Petaflops Computing" 1994 workshop, MIT Press 1995

Trew, A. Wilson, G. "Past, Present, Parallel: A Survey of Available Parallel Computer Systems", Springer, 1991

Vettter, J. S. "Contemporary High Performance Computing: From Petascale toward Exascale" (Chapman & Hall/CRC Computational Science. 2013.

Candidates for Further Reading

Bell, G., "The Future of High Performance Computers in Science and Engineering", Communications of the ACM, Vol. 32, No. 9, September 1989, pp 1091-1101

Bell, G., "Ultracomputers: A Teraflop Before Its Time", Communications of the ACM, Vol. 35, No. 8, August 1992, pp 27-45.

Bell, C. Gordon. "Bell's Law for the Birth and Death of Computer Classes," Communications of the ACM 51, no. 1 (January 2008), p. 86-94.

Buchholz, W (ed) "Planning a Computer System - Project Stretch", McGraw-Hill Book Company. 1962

CDC 8600 Manual, preliminary 8600 reference manual, Control Data Corporation, and CDC 8600 Patent No. 3833889, dated September 3, 1974.

Seymour Cray Oral History. by David Allison, National Museum of American History, Smithsonian Institution 1995

Fox, G.C. , Williams, R.D., and Messina, P.C. "Parallel Computing Works!" Morgan Kaufman, 1994

Hord, R. M. "The Illiac IV: The First Supercomputer". Springer-Verlag (May 1982)

Hwang K., Jack Dongarra, J., and Fox G. C. "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things" Morgan Kaufman

Karin, Sid, and Norris Parker Smith. The Supercomputer Era. Orlando, FL: Harcourt Brace Jovanovich, 1987

Timothy G. Mattson, David Scott, Stephen Wheat, A TeraFLOP Supercomputer in 1996: '' the ASCI TFLOP System" Proceedings of IPPS '96, p84-93.

Office of Science and Technology Policy (OSTP). Grand challenges 1993: High performance computing and communications, A Report by the Committee on Physical, Mathematical, and

ENGINEERING SCIENCES OF THE FEDERAL COORDINATING COUNCIL FOR SCIENCE, ENGINEERING, AND TECHNOLOGY IN THE OFFICE OF SCIENCE AND TECHNOLOGY POLICY, NATIONAL SCIENCE FOUNDATION, 1992.

PAGELKOPF, DON; ET AL. (1975). "REMINISCENCES OF COMPUTER ARCHITECTURE AND COMPUTER DESIGN AT CONTROL DATA CORPORATION". UNIVERSITY OF MINNESOTA DIGITAL CONSERVANCY. CHARLES BABBAGE INSTITUTE. DISCUSSION TOPICS INCLUDE CONTROL DATA CORPORATION, CDC 1604, CDC 6600, CDC 7600, CDC 8600, CDC STAR-100 AND SEYMOUR CRAY

THORNTON, J. E. "THE DESIGN OF A COMPUTER: THE CONTROL DATA 6600", SCOTT, FORESMAN AND COMPANY, 1970