A History of Supercomputing

Burton Smith
Microsoft

Outline

- Definitions
- Early days: Colossus and Eniac
- Mainframes: Univac LARC to CDC 7600
- SIMD systems
  - Processor arrays: Illiac IV, PEPE, and BSP
  - One-bit arrays: Staran, DAP, and the Goodyear MPP
  - Vector pipelining: TI-ASC, CDC Star-100, and Cray-1
  - Wide instructions: FPS, Multiflow, and Cydrome
- MIMD systems
  - Shared memory: Cray X-MP to SGI Altix
  - Distributed memory: Cosmic Cube to Blue Gene
- Japan-US competition
- What next?

Definitions

- Supercomputer: the world's fastest computer at the time
- SIMD: Single Instruction Stream, Multiple Data Stream
  - One instruction at a time, each operating on an array of data
- MIMD: Multiple Instruction Stream, Multiple Data Stream
  - Multiple processors asynchronously issuing instructions
- Shared Memory: MIMD computer in which a common memory is accessible by all processors
  - UMA: Uniform memory access by all processors
  - NUMA: Non-uniform memory access, based on placement
- Distributed Memory: MIMD computer in which the memory is partitioned into processor-private regions
- RISC: Reduced Instruction Set Computer
  - Uses fast and simple instructions to do complex things
- Pipelining: Processing in assembly-line style

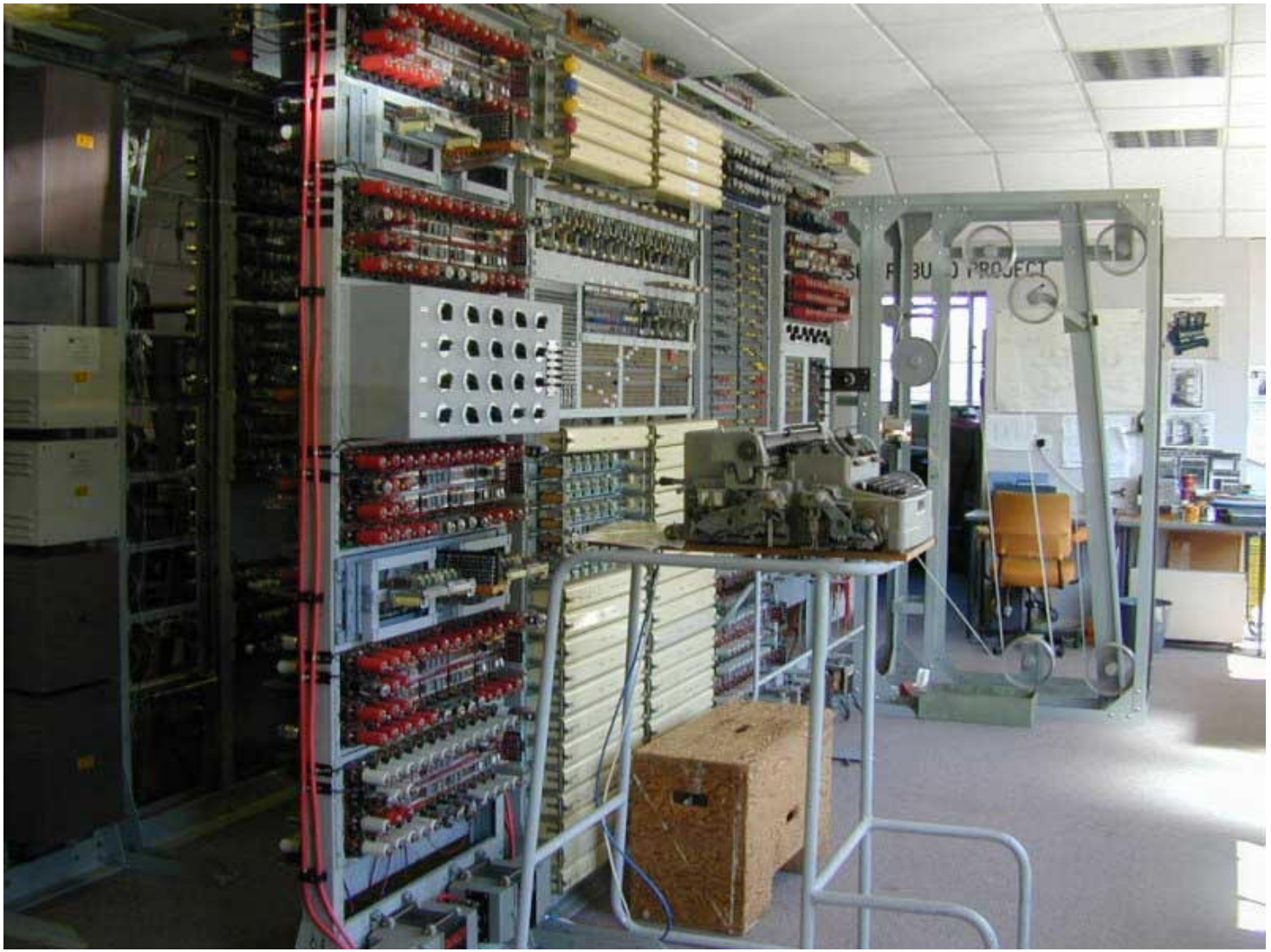
Names and Places

- BMD/ATC: Ballistic Missile Defense Advanced Technology Center, Huntsville, AL
- BRL: Ballistics (now Army} Research Laboratory, Aberdeen, MD
- DARPA: Defense Advanced Research Projects Agency, VA
- NASA-Ames: Ames Research Center, Mountain View, CA
- NASA-Goddard: Goddard Space Flight Center, Greenbelt, MD
- Livermore: Lawrence Livermore National Laboratory, Livermore, CA
- Los Alamos: Los Alamos National Laboratory, Los Alamos, NM
- NSA: National Security Agency, Fort Meade, MD
- Sandia: Sandia National Laboratories, Albuquerque, NM

Early days: Colossus

- Used at Bletchley Park, England during WW II
  - for cryptanalysis of Lorenz SZ40/42 rotor systems ("Fish")
  - Only recently declassified by Her Majesty's Government
- Concepts by Max Newman *et al.*
- Designed and built by Tommy Flowers
  - at the Post Office Research Station, Dollis Hill
- Features:
  - Paper tape loop data input at 30 mph = 5000 characters/sec
  - 1500 vacuum tubes ("valves") in the Mark I, 2400 in Mark II
  - Programmed by switches
  - Not a general-purpose computer

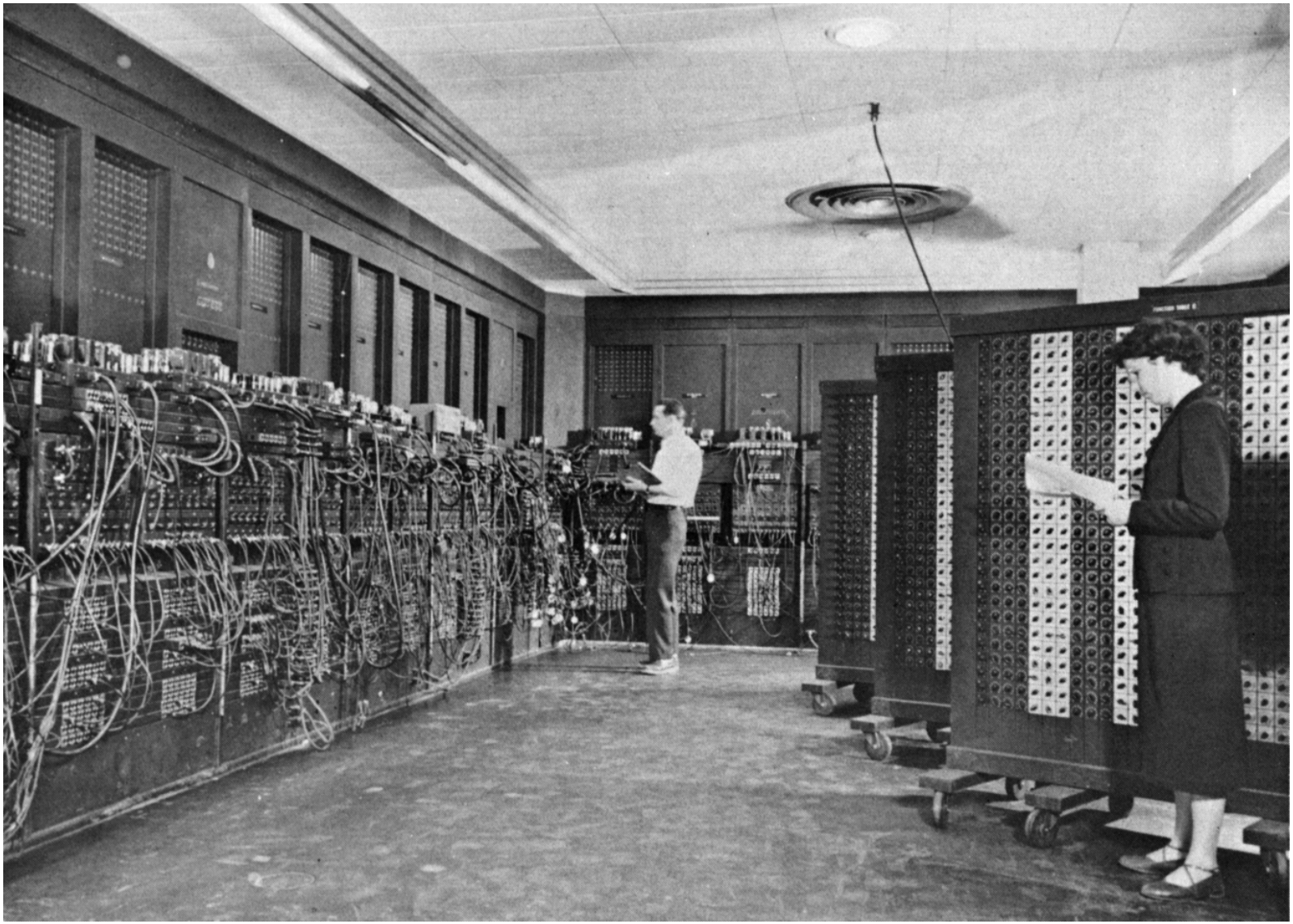
Colossus



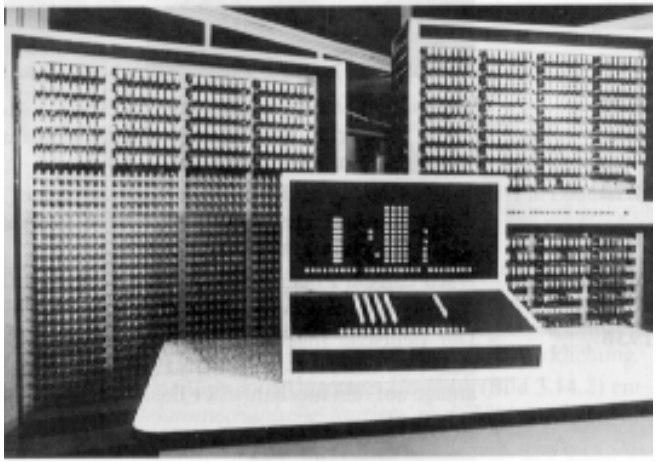
Early days: Eniac

- Built at Penn's Moore School for BRL, Aberdeen, MD
  - The motivating application was artillery firing tables
  - Its first use was evaluating H-bomb feasibility
- Designed and built by J. Presper Eckert and John Mauchly
- Unveiled at the Moore School on February 15, 1946
- It had 17,468 vacuum tubes and consumed 174 KW
  - Reliability was a concern, so it was left on continuously
- It was 80 feet long and weighed 30 tons
- Programming was via switches and patch cables
- Not a general-purpose computer

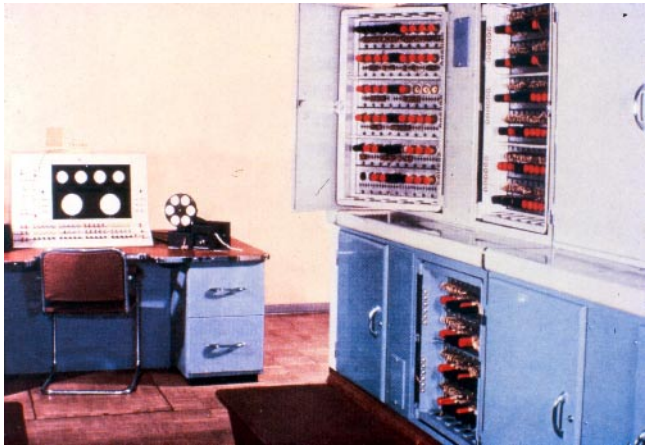
Eniac



Other early supercomputers

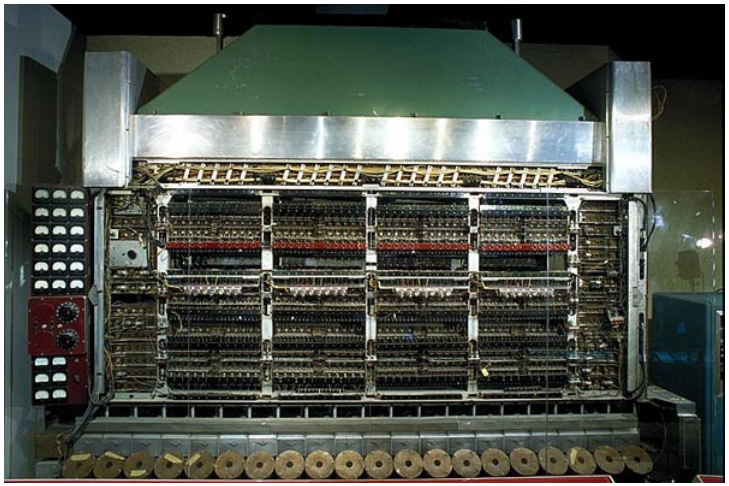


Zuse Z3 (1941)



Manchester/Ferranti Mark I (1951)

Univac 1 (1951)



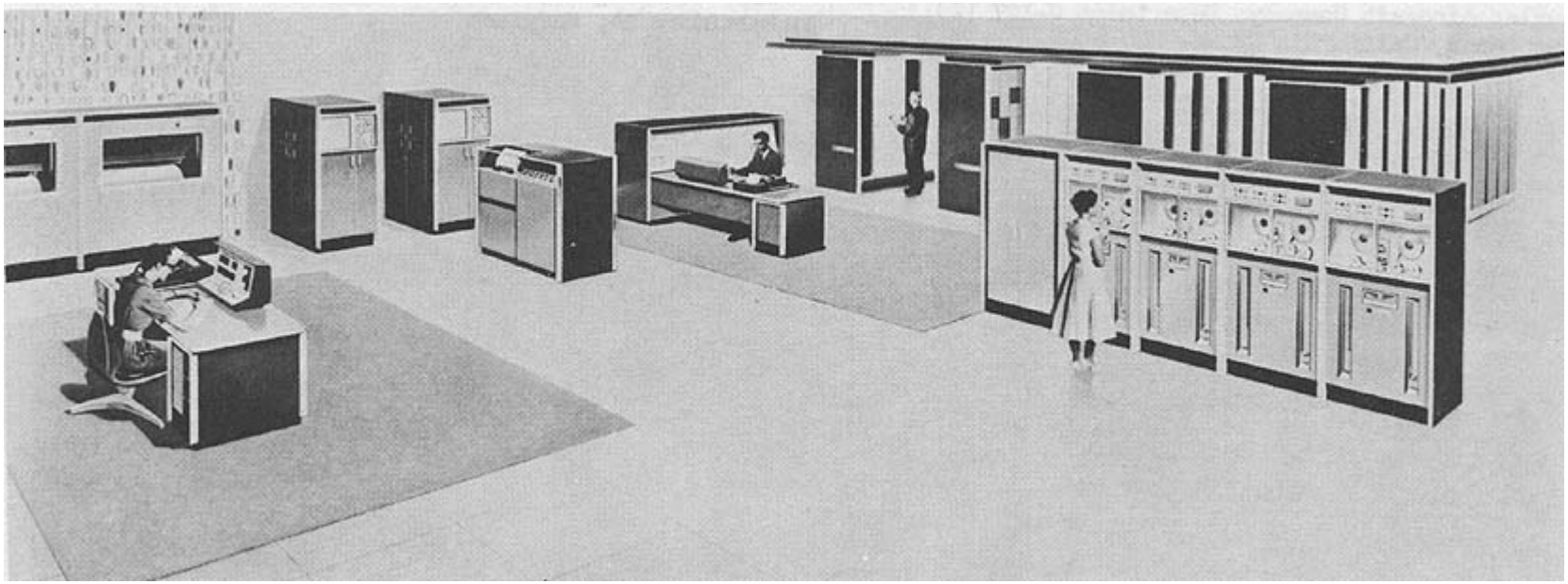
The IAS machines (1952)

Mainframes

- The 50's and 60's saw fast progress in computer technology
- Most of it was first used in "mainframe" supercomputers
  - These systems were used for both business and science
  - Later, supercomputers became much more science-oriented
- Examples of new technology included:
  - Magnetic core memory
  - Transistor logic circuits
  - Floating point hardware
  - Pipelining
  - High level languages and compilers

Mainframes: LARC

- Begun in 1955 for Livermore and delivered in 1960
- Had dual processors and decimal arithmetic
- Employed surface-barrier transistors and core memory



Mainframes: Stretch and Harvest

- IBM 7030 (STRETCH)
- Delivered to Los Alamos 4/61
- Pioneered in both architecture and implementation at IBM

- IBM 7950 (HARVEST)
- Delivered to NSA 2/62
- Was STRETCH + 4 boxes
  - IBM 7951 Stream unit
  - IBM 7952 Core storage
  - IBM 7955 Tape unit
  - IBM 7959 I/O Exchange

Mainframes: CDC 6600

- Seymour Cray and Bill Norris left Univac to start CDC
- Cray's CDC 1604 was the first successful transistor system
- He wanted to build a scientific supercomputer
- The 6600, built with silicon transistors, shipped Sept. 1964
  - The chief architects were Seymour Cray and Jim Thornton
- The 6600 had several notable features
  - A very simple instruction set, lending itself to more speed
    - No decimal arithmetic needed or wanted
  - 10 arithmetic function units able to process in parallel
    - This overlapped execution style is now called *superscalar*
    - It was coordinated by a famous circuit called the *scoreboard*
  - Parallel I/O processors using a new idea called *multithreading*
    - Some say RISC means "Really Invented by Seymour Cray"
  - A futuristic operator's console was provided at no extra charge

CDC 6600 Console

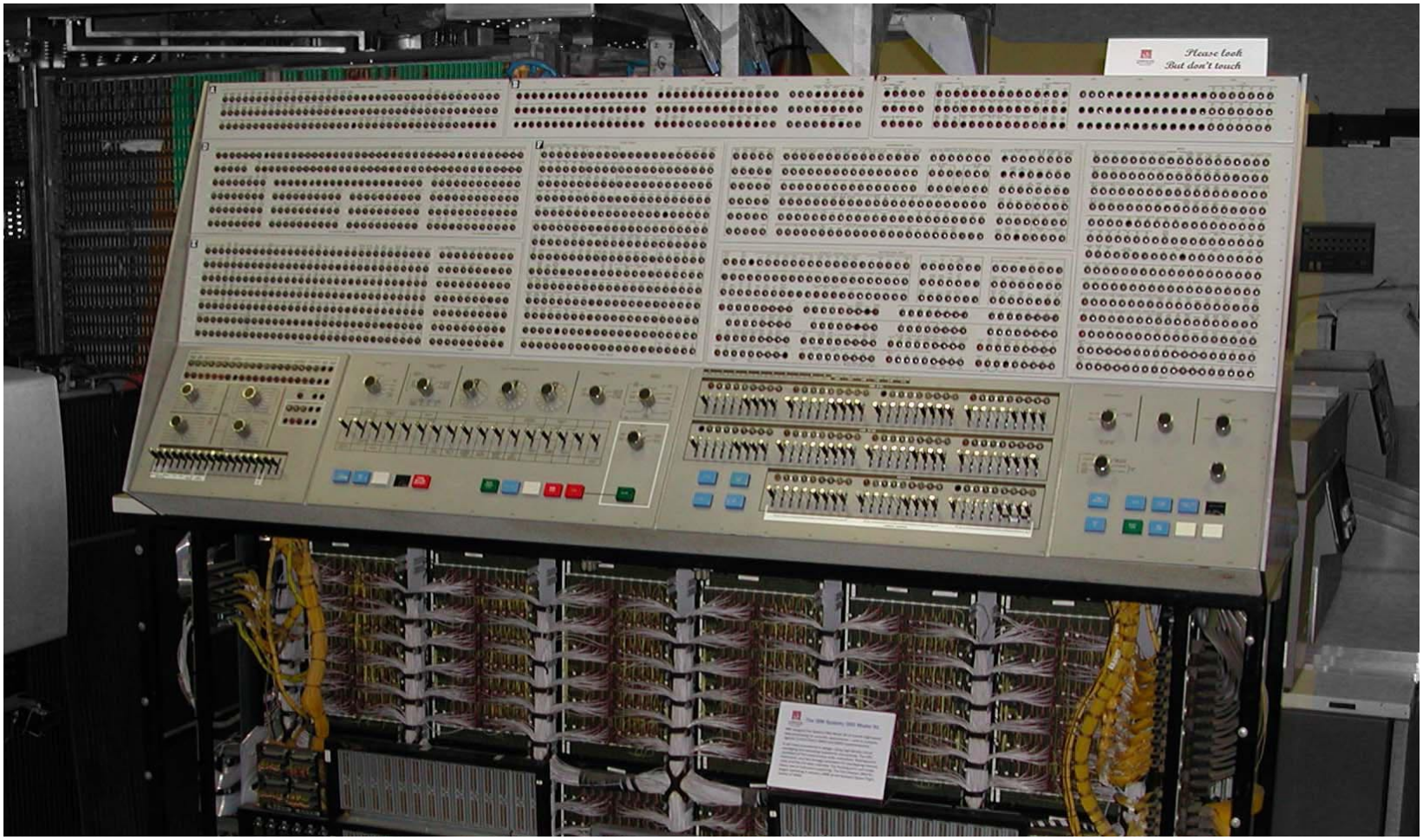
Mainframes: IBM 360/91

- IBM's Tom Watson was angry:

  "Last week, Control Data ...announced the 6600 system. I understand that in the laboratory...there are only 34 people including the janitor. Of these, 14 are engineers and 4 are programmers. Contrasting this modest effort with our 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."

- The 360/91 (AKA 370/195) was IBM's answer
- It was delivered to NASA-Goddard in October 1967
- Killer App: Passenger Airline Reservation Systems (1964)
  - Jointly developed by IBM, American, and United Airlines
  - Originally written in IBM assembly (*i.e.* machine) language
- It was very close to the CDC 7600 in performance
- The 360/91 also had a pretty impressive console…

IBM 360/91 Console



Mainframes: CDC 7600

- The 7600 was delivered to Livermore in early 1969
- It was highly compatible with the 6600
- Besides being faster, its arithmetic was pipelined
  - This meant the arithmetic units were never busy
- It was so compact it had to be cooled with freon
- It was also one of the worlds most beautiful computers

Seymour Cray

Two CDC 7600s

SIMD arrays: Illiac IV

- By the late 60's, it was clear mainframes weren't enough
- To improve performance, SIMD array machines were built or proposed with many arithmetic processing units
  - Solomon was an early Westinghouse SIMD array prototype
- The Illiac IV was a University of Illinois/Burroughs project
  - Funded by DARPA from 1964 onward, it was usable in 1975
  - The chief architect, Dan Slotnick, came from Westinghouse
- It was to have 256 arithmetic units, later cut back to 64
- The thin-film memory system was a major headache
- After student demonstrations at Illinois in May 1970, the project was moved to NASA-Ames
- Languages like IVtran (*not* pronounced *four-tran*) aimed to use parallel loops to express the necessary parallelism
  - Much compiler expertise was developed working on Illiac IV

ILLIAC IV

SIMD Arrays: PEPE and BSP

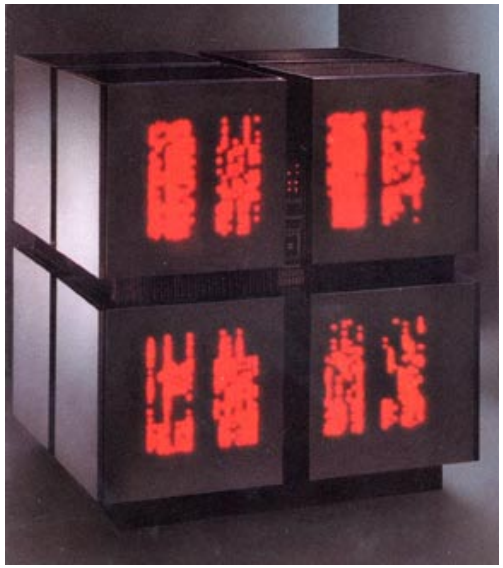
- Both were Burroughs designs derived from Illiac IV
- PEPE had 288 processing elements and was designed to track ballistic missiles
  - It was delivered to BMD/ATC in Huntsville in 1976
- BSP had 16 processing elements and was aimed at commercial use
  - Richard Stokes was chief architect
  - The first customer was Marathon Oil Co. in Denver, CO
  - The project was cancelled in 1980, before its first delivery

One-bit Arrays: STARAN, DAP, MPP

- These were arrays of processors that were only one bit wide
  - They were inexpensive to implement, especially on one chip
  - Arithmetic was serial (like grade-school, only binary)
    - It could use whatever precision made sense for the problem
- Goodyear Aerospace developed STARAN in 1972
  - It was a 256 by 256 array with lots of data permuting power
  - Ken Batcher, an Illinois graduate, was chief architect
  - It was great at image processing, though not intended for it
  - The ASPRO version was used on Navy AWACS aircraft
- The Goodyear MPP was a single-chip STARAN successor
  - The array size was reduced to 128 by 128
  - Delivered to NASA-Goddard in 1984
- ICL (UK) developed the Distributed Array Processor
  - Stuart Reddaway was the architect

One-bit Arrays: Connection Machine

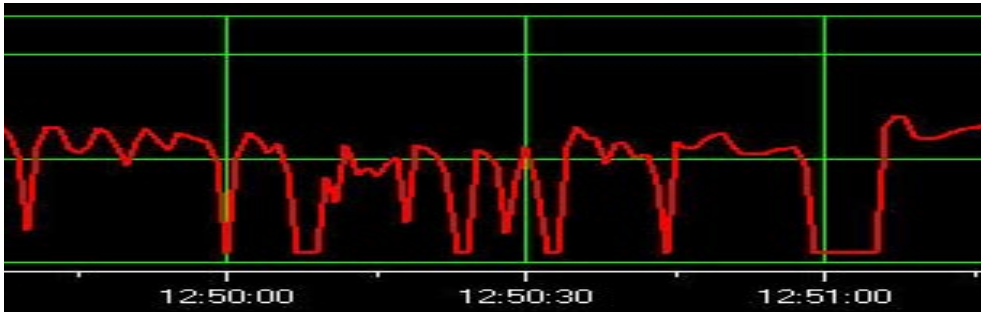
- Danny Hillis and his colleagues initially based the design on a "connectionist" approach to artificial intelligence
  - The design came out of Danny's PhD thesis at MIT
- The CM-1 was a hypercube-connected one-bit array
  - High-level language support was excellent
  - The machine proved useful for scientific computation
- The CM-2/CM-200 added some 32-bit floating point hardware
  - Many HPC users like NSA and Los Alamos bought them
- The CM-2 had an array of light-emitting diodes (right)
  - Not one per processor, sadly
- The CM-5 was MIMD, not SIMD



CM-2

Vector pipelining: TI ASC

- The ASC was designed to exploit Fortran loops
  - This motivated the vector pipelining decision
  - The ASC's Fortran compiler was state-of-the-art
- Several systems shipped, beginning in 1974
  - Geophysical Fluid Dynamics Laboratory, Princeton, NJ
  - Ballistic Missile Defense, Huntsville
  - Naval Research Laboratory, Anacostia, MD
- The GFDL machine had an interesting performance tool: a strip-chart recorder perched on top that plotted vector length
  - The recording was returned to the user with the job output



Vector Pipelining: CDC Star 100

- CDC used APL rather than Fortran loops as a vector model
  - APL is a language invented by Ken Iverson of IBM Research
  - Neil Lincoln and Jim Thornton were the architects
  - Sidney Fernbach at Livermore was the target customer
- The Star-100 had poor scalar (non-vector) performance
  - The result was highly limited applicability
  - This effect is called *Amdahl's law*
- Both Livermore and Los Alamos were scheduled to get one
  - Los Alamos backed out in favor of the Cray-1
  - Livermore got both Star-100 systems as a result

Star-100

Amdahl's law

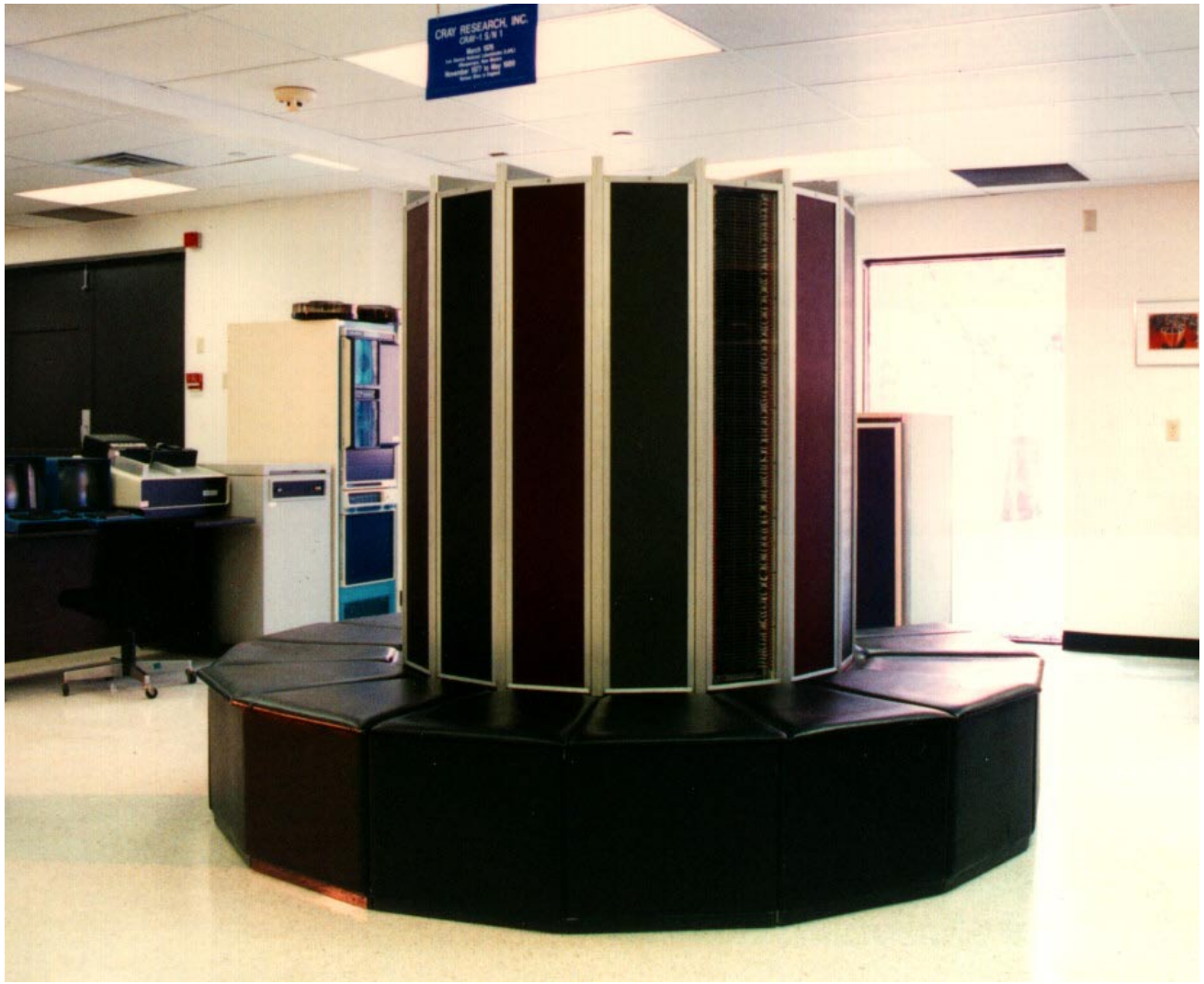
- If w_1 work is done at speed s_1 and w_2 at speed s_2, the average speed s is $(w_1+w_2)/(w_1/s_1 + w_2/s_2)$
  - This is just the total work divided by the total time
- For example, if $w_1 = 9$, $w_2 = 1$, $s_1 = 100$, and $s_2 = 1$ then $s = 10/1.09 \cong 9$
  - This is obviously not the average of s_1 and s_2

Amdahl, Gene M, "Validity of the single processor approach to achieving large scale computing capabilities", Proc. SJCC, AFIPS Press, 1967

Vector Pipelining: Cray-1

- Unlike the CDC Star-100, there was no development contract for the Cray-1
  - Mr. Cray disliked government's looking over his shoulder
- Instead, Cray gave Los Alamos a one-year free trial
- Almost no software was provided by Cray Research
  - Los Alamos developed or adapted existing software
- After the year was up, Los Alamos leased the system
  - The lease was financed by a New Mexico petroleum person
- The Cray-1 definitely did not suffer from Amdahl's law
  - Its scalar performance was twice that of the 7600
  - Once vector software matured, 2x became 8x or more
- When people say "supercomputer", they think Cray-1

Cray-1

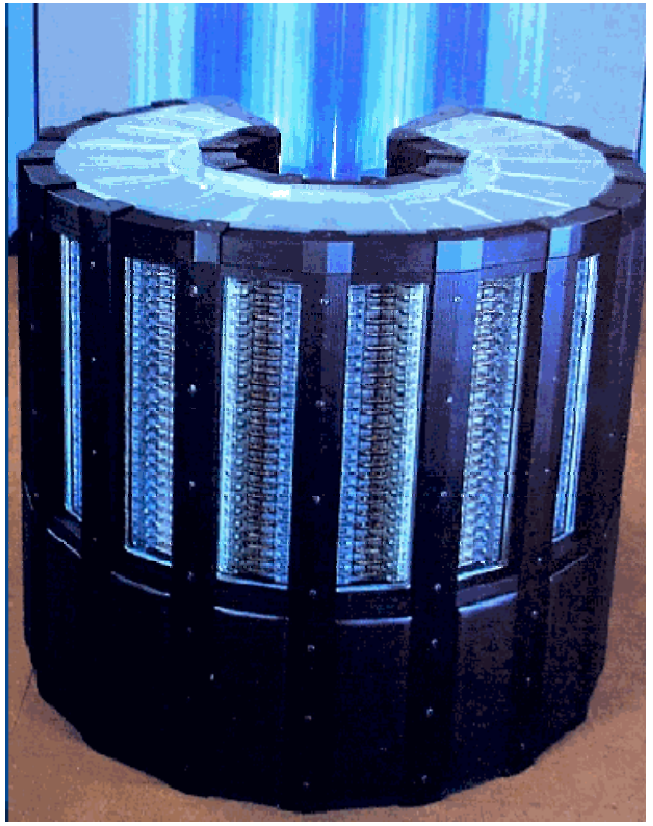


VLIW: FPS, Multiflow, and Cydrome

- VLIW stands for Very Long Instruction Word
  - It is a kind of SIMD system in which each arithmetic unit can be doing something different (thus the need for width)
- Floating Point Systems led in 1976 with the AP-120B
  - Glenn Culler was the chief architect
  - These systems were embedded in GE CAT scanners
  - They were also used widely for seismic data processing
- Multiflow was founded in 1984 by Josh Fisher to exploit some compiler technology he had been pursuing at Yale
  - The first shipments of the Trace 7/200 were in 1987
- Bob Rau's company, Cydrome, was also founded in 1984 and demonstrated the Cydra-5 machine in 1987
- Both systems had instructions with seven fields and 256 bits
- Technically, none of these machines was a supercomputer

Shared Memory: Cray Vector Systems

- Cray Research, by Seymour Cray
  - Cray-1 (1976): 1 processor
  - Cray-2 (1985): up to 4 processors*
- Cray Research, not by Seymour Cray
  - Cray X-MP (1982): up to 4 procs
  - Cray Y-MP (1988): up to 8 procs
  - Cray C90: (1991?): up to 16 procs
  - Cray T90: (1994): up to 32 procs
  - Cray X1: (2003): up to 8192 procs
- Cray Computer, by Seymour Cray
  - Cray-3 (1993): up to 16 procs
  - Cray-4 (unfinished): up to 64 procs
- All are UMA systems except the X1, which is NUMA



Cray-2

*One 8-processor Cray-2 was built

Shared Memory: Multithreading

- These systems pipeline MIMD architecture in the same way vector systems pipeline SIMD architecture
  - Like vector multiprocessors, they can be UMA or NUMA
- Examples include
  - Denelcor HEP (1980): up to 16 procs
  - Tera MTA-1 (1998): up to 16 procs
  - Cray MTA-2 (2002): up to 256 procs
- A variant form of multithreading is Simultaneous Multithreading (SMT)
  - Intel® calls SMT "Hyper-Threading"

Cray MTA-2

Shared Memory: CC-NUMA

- Cache-coherent NUMA systems were researched at Stanford and developed as products by Silicon Graphics
  - Kendall Square Research used a variant called COMA, which stands for Cache-Only Memory Architecture
- These systems automatically move data among *caches* at each processor to minimize processor data access times
- SGI CC-NUMA systems:
  - Origin 2000 (1997): up to 128 procs
  - Origin 3000 (2003): up to 512 procs
  - Altix (2003): up to 512 procs*
- Kendall Square COMA systems:
  - KSR1: (1991): up to 1088 procs
  - KSR2: (1994): up to 5000 procs

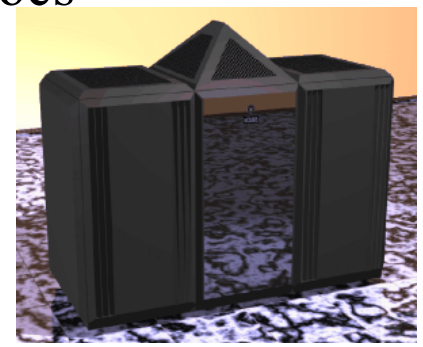
SGI Altix

* "Columbia" at NASA-Ames clusters 20 of these

Distributed Memory: Cosmic Cube

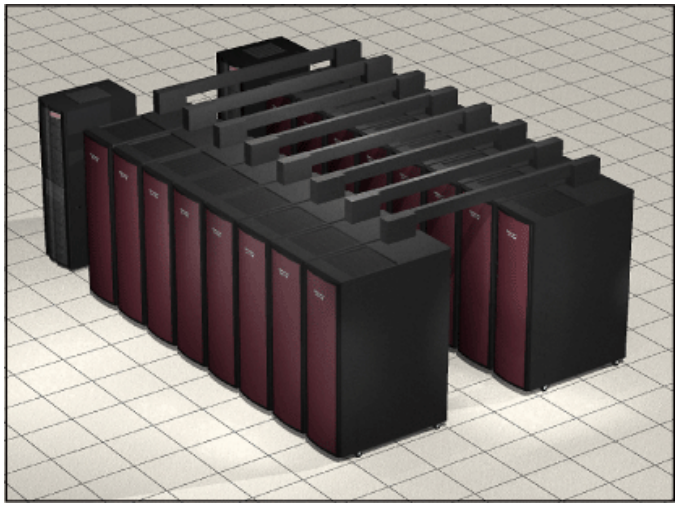
- Chuck Seitz and Geoffrey Fox pioneered the idea at Caltech
  - The first system (1982) was called the *Cosmic Cube*
  - It had 64 Intel® 8086 processors, each with an 8087 attached
  - The processors were connected in a 6-dimensional hypercube
  - They communicated by sending messages to each other
  - This system architecture now dominates supercomputing
- Early commercial systems included:
  - Ametek/S14 (1986): 68020, over 1000 procs
  - Intel® IPSC/1 (1985): 80286, up to 64 procs
  - nCUBE/10 (1985): custom, up to 1024 procs
- Variations since then are legion

nCUBE/10

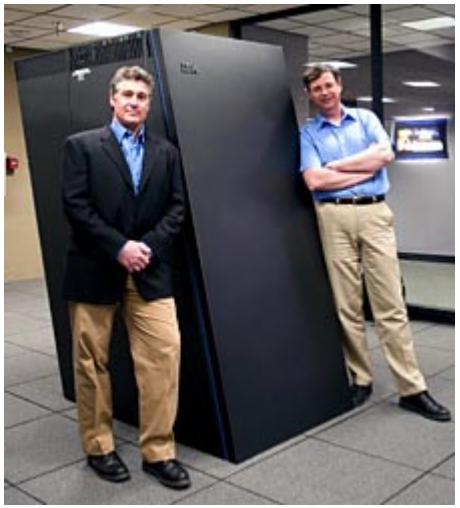


Distributed Memory: Classic MPPs

- A classic MPP is a distributed memory system whose processing elements are individual (micro)processors
- Examples include:
  - Cray T3D (1994), T3E (1996), XT3
  - Intel Touchstone series, Paragon (1992)
  - IBM SP-1 (1992), SP-2, Deep Blue, Blue Gene
  - And many others



Cray XT3

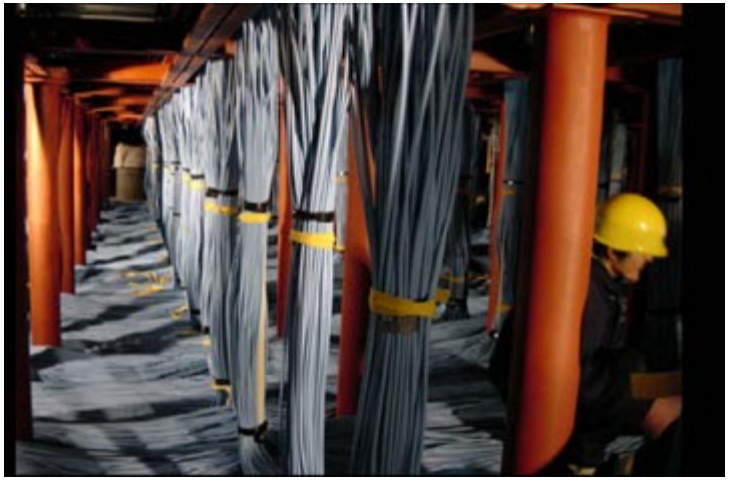


IBM Blue Gene/L

Distributed Memory: Clusters

- A cluster MPP is a distributed memory system whose processing elements are shared memory multiprocessors
- Examples include:
  - IBM eServer p575
  - NEC Earth Simulator (2002), SX-8
  - A vast multitude of home-made “Beowulf” clusters
    - Usually based on Intel or AMD processors
    - Even Government Labs have done this

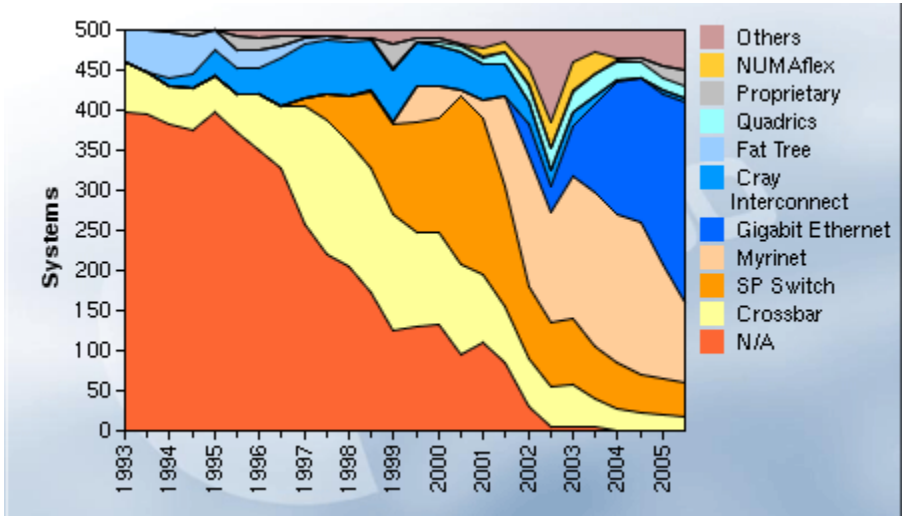
Earth Simulator (above)



Earth Simulator (below)

The Top 500 List

- This is a list of the fastest computers in the world
  - www.top500.org
  - By definition, it's a great place for supercomputer spotting
- Be very careful about the term "fastest", though
  - Speed is measured by one benchmark program, *LINPACK*
  - LINPACK doesn't care how well the system is connected
  - Shared memory isn't useful for good LINPACK results
- With that caveat, it's well worth a look



Japanese Systems: Fujitsu, NEC, Hitachi

- Japanese supercomputers mostly use vector processors with shared or (most recently) distributed memory architecture
- Fujitsu shipped its first vector system (Facom 230) in 1976:
  - This was followed by the VP200 (1982), VP400 (1985), VP2000 (1989), VP2600(1990), and distributed memory VPP500(1992), VPP700(1996), and VPP5000(1999)
  - Recently, Fujitsu has been selling scalar NUMA systems
- NEC got a late start on vectors, but caught up very quickly:
  - SX-1 (1983), SX-2 (1985), SX-3 (1990), SX-4 (1994), SX-5 (2000), and distributed memory SX-6 and SX-7 (2002), Earth Simulator (2002), and SX-8 (2006)
  - Tadashi Watanabe is the chief architect of these systems
- Hitachi has sold few supercomputers outside Japan
  - An exception is the distributed memory SR8000 (2000)

Japanese National Projects

- The National Superspeed Computer Project (1981)
  - About ¥23 billion over 9 years
  - Intended to produce a supercomputer
  - Two projects: architecture and technology
  - Companies: Fujitsu, NEC, Hitachi, Mitsubishi, Oki, Toshiba
- The Fifth Generation Project (1982)
  - About ¥55 billion over 10 years
  - Centered at a new laboratory: ICOT
  - Focused on artificial intelligence: inference, vision, speech
  - MCC in Austin, TX was a US industrial response

DARPA Strategic Computing Program

- This program was intended to accelerate the development and application of parallel computing
- It funded several industrial development projects over the period 1983-1993:
  - Thinking Machines: Connection Machine (1-bit SIMD array)
  - Bolt, Beranek, and Newman: Butterfly (shared memory)
  - Intel: Touchstone series (distributed memory)
  - Tera: MTA (shared memory)
  - Cray: T3D (distributed memory)
- Government agencies were encouraged to try these systems

The DOE ASCI Program

- The Department of Energy was directed to use computer simulation to replace nuclear weapons testing
- It has procured a series of distributed memory systems:

System Name	Laboratory	Manufacturer	Date
Red	Sandia	Intel	1995
Blue Pacific	Livermore	IBM	1996
Blue Mountain	Los Alamos	SGI	1998
White	Livermore	IBM	2000
Q	Los Alamos	Compaq	2002
Red Storm	Sandia	Cray	2004
Purple	Livermore	IBM	2006

- The performance range is about 100-fold

What Next?

- Parallelism is becoming mainstream
  - Intel and AMD are putting multiple processors on a chip
  - It's only a few now, but just wait!
- This will change the supercomputer playing field, again
  - Parallel application software will emerge
  - Parallel languages and compilers will enable it
  - Parallel architecture will evolve to match
- The world's computers become one giant cluster
  - The cloud links them
  - We must make them safe
- What an exciting field it is with such challenges!

Apologies: Companies Not Mentioned

- Alliant
- Convex
- HP
- MasPar
- Matsushita
- Meiko
- Parsys
- Parsytec
- Quadrics
- Siemens-Nixdorf
- Sun