Cray Supercomputers
Past, Present, and Future

Hewdy Pena Mercedes, Ryan Toukatly
Advanced Comp. Arch. 0306-722
November 2011
Cray Companies


- Cray Research, Inc. bought by Silicon Graphics, Inc (SGI) in 1996.

- Cray Inc. Formed when Tera Computer Company (pioneer in multi-threading technology) bought Cray Research, Inc. in 2000 from SGI.
Seymour Cray

- Joined Engineering Research Associates (ERA) in 1950 and helped create the ERA 1103 (1953), also known as UNIVAC 1103.

- Joined the Control Data Corporation (CDC) in 1960 and collaborated in the design of the CDC 6600 and 7600.

- Formed Cray Research Inc. in 1972 when CDC ran into financial difficulties.
  - First product was the Cray-1 supercomputer
  - Faster than all other computers at the time.
  - The first system was sold within a month for US$8.8 million.
  - Not the first system to use a vector processor but was the first to operate on data on a register instead of memory.
Vector Processor

- CPU that implements an instruction set that operates on one-dimensional arrays of data called *vectors*.

- Appeared in the 1970s, formed the basis of most supercomputers through the 80s and 90s.

- In the 60s the Solomon project of Westinghouse wanted to increase math performance by using a large number of simple math co-processors under the control of a single master CPU.

- The University of Illinois used the principle on the ILLIAC IV. The original design wanted a 1 GFLOP machine with 256 ALUs, but it only had 64 ALUs and could reach only 100 to 150 MFLOPS (Not bad for 1972).
Vector Processor

- CDC had the first practical implementation of a vector processor with the Star-100 system
  - Used memory-to-memory architecture
  - Uses vectors of any length
  - Pipeline had to be very long to allow it to have enough instructions to make up for the slow memory
  - High cost when switching from vectors to individually-located operands
  - Poor scalar performance
Cray 1 (1975)

- Vector operations over 64bit registers instead of directly in memory
- Pipeline parallelism for different instructions
- Vector Instructions could be pipelined (vector chaining)
- 80 Mflops, 80MHZ, 1 MW on 16 interleaved memory banks
- Versions: 1A, 1S, 1M
- One port for read and write
COMPUTATION SECTION
- 64-bit word
- 12.5 nanosecond clock period
- 2's complement arithmetic
- Scalar and vector processing modes
- Twelve fully segmented functional units
- Eight 24-bit address (A) registers
- Sixty-four 24-bit intermediate address (B) registers
- Eight 64-bit scalar (S) registers
- Sixty-four 64-bit intermediate scalar (T) registers
- Eight 64-element vector (V) registers, 64-bits per element
- Four instruction buffers of 64 16-bit parcels each
- Integer and floating point arithmetic
- 128 Instruction codes

MEMORY SECTION
- Up to 1,048,576 words of bi-polar memory
  (64 data bits and eight error correction bits)
- Eight or sixteen banks of 65,536 words each
- Four-clock-period bank cycle time
- One word per clock period transfer rate to B, T, and V registers
- One word per two clock periods transfer rate to A and S registers
- Four words per clock period transfer rate to instruction buffers
- Single error correction - double error detection (SEC-DED)

INPUT/OUTPUT SECTION
- Twelve input channels and twelve output channels
- Channel groups contain either six input or six output channels
- Channel groups served equally by memory (scanned every four clock periods)
- Channel priority resolved within channel groups
- Sixteen data bits, three control bits per channel, and 4 parity bits
- Lost data detection
Cray 1

COMPUTATION SECTION
- Registers
- Functional units
- Instruction buffers

MEMORY SECTION
- 0.25 M or 0.5 M or 1 M
- 64-bit bi-polar words

I/O SECTION
- 12 input channels
- 12 output channels

MCU
MASS STORAGE SUBSYSTEM
FRONT-END COMPUTERS, I/O STATIONS AND PERIPHERAL EQUIPMENT
Cray X-MP (1982)

- Shared memory parallel vector processor (first)
- Better memory bandwidth, two read ports to memory instead of one
- Improved chaining support
- 32 memory banks
- Two read ports, one write port, dedicated port to I/O
- 400MFlops, 105MHz
- Further versions included 1, 2 or 4 processors. Up to 800MFlops
Cray X-MP
## System configuration options

<table>
<thead>
<tr>
<th></th>
<th>X-MP/1</th>
<th>X-MP/2</th>
<th>X-MP/4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mainframe</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPUs</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Bipolar memory (64-bit words)</td>
<td>N/A</td>
<td>N/A</td>
<td>8 or 16M</td>
</tr>
<tr>
<td>MOS memory (64-bit words)</td>
<td>1, 2, 4 or 8M</td>
<td>4, 8 or 16M</td>
<td>N/A</td>
</tr>
<tr>
<td>6-Byte channels</td>
<td>2 or 4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>100-Byte channels</td>
<td>1 or 2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1000-Byte channels</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>I/O Subsystem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/O processors</td>
<td>2, 3 or 4</td>
<td>2, 3 or 4</td>
<td>4</td>
</tr>
<tr>
<td>Disk storage units</td>
<td>2-32</td>
<td>2-32</td>
<td>2-32</td>
</tr>
<tr>
<td>Magnetic tape channels</td>
<td>1-8</td>
<td>1-8</td>
<td>1-8</td>
</tr>
<tr>
<td>Front-end interfaces</td>
<td>1-7</td>
<td>1-7</td>
<td>1-7</td>
</tr>
<tr>
<td>Buffer memory (Mbytes)</td>
<td>8, 32 or 64</td>
<td>8, 32 or 64</td>
<td>64</td>
</tr>
<tr>
<td><strong>Solid-state Storage Device</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory size (Mbytes)</td>
<td>256, 512 or 1024</td>
<td>256, 512 or 1024</td>
<td>256, 512 or 1024</td>
</tr>
</tbody>
</table>
Cray X-MP
Cray-2 (1985)

- 4 processor vector supercomputer
- Memory banks were arranged in quadrants to be accessed at the same time
- B & T registers were replaced with 16 KW block of fast memory called local memory allowing data scattering to increase parallelism
- Foreground processor fed local memory and ran the computer
- Now referred as load/store unit
- First delivered had more memory than all previous machines: 256MW
- 1.9GFlops, 64MW, 4GB main memory
- Used 3-D stack of boards because of high device density
- Liquid cooled
Cray 2
Cray Y-MP (1988)

- Extended address register from 24 to 32
- 2, 4 or 8 vector processors
- Each processor had two FU's with 167MHz speed
- 128 to 512 MB SRAM memory
- Configurable with up to 2 D IOS's and optional SSD up to 4GB capacity
- 333MFlops peak performance per processor
- Y-MP M90 could use up to 32GB of DRAM memory
Cray First Massively Parallel Super Computer Architectures

- **Cray C90 1991**
  - Dual vector pipeline
  - 244 MHz speed
  - Could support up to 16 processors
  - Up to 8GB SRAM main memory

- **Cray T90 1995**
  - Last vector processor manufacture by CRI
  - 450MHz
  - Two wide vector pipeline
  - 1.8 Gflop per processor
  - As J90 had a scalar cache per CPU
  - Up to 32 processors
  - Shared MM up to 8GB SRAM
  - Clock signal distributed by fiber optic
  - 64bit word with 32 CPU could give a stream BW of 360 GB per second
CMOS Flavors

- **Cray XMS 1990**
  - CMOS version of Cray X-MP
  - VMEbus-based I/O Subsystem
  - 18.2 MHz
  - First Cray to support removable disk drives

- **Cray Y-MP EL 1992**
  - CMOS version of Y-MP
  - Up to 4 processors with peak performance of 133 Mflops
  - Up to 1GB of DRAM

- **Cray EL90 1993**
  - Versions with up to 8 processors with up to 256 Mword DRAM

- **Cray J90 1994**
  - 32 Processors with 100MHz
  - 4GB MM
  - 48GB memory performance
  - One chip for scalar, one for vector operations
  - Scalar processor includes a 128W data cache

- **Cray SV1 1998**
  - Included vector cache
  - Included multi streaming (4 processors function as a virtual unit)
  - 300MHz
  - Up to 32 processors with up to 512MB shared memory buses
  - Up to 32 SV1 could be clustered together
Cray T3D (Torus 3-Dimensions) 1993

- **Cray T3D 1993**
  - First massively parallel supercomputer for Cray
  - First use of another company's processor
  - Between 32 and 2048 PE grouped in pairs or nodes of 6 processors
  - Each PE had 150 DEC Alpha 21064 microprocessor with up to 64 MB of DRAM
  - BW of 300 MB/s each direction
  - Designed to be hosted by a Y-MP cabinet model E and rely on its UNICOS OS

- **Cray T3E 1995**
  - Had FROM 8 to 2,176 PE
  - Each PE had up to 2GB of DRAM and a 6 way interconnect router with payload BW of 480 MB/s
  - Selfhosted
Cray Failed Projects

- **Cray 3 1993**
  - First to use gallium arsenide
  - Had a foreground processing system dedicated to I/O with 32 bit processor and 4 synchronous data channels
  - Up to 16 background processors
  - Up to 16GB common memory
  - Background processor had a computation section, control section and local memory
  - 4 to 16 processors at 474MHz

- **Cray 4 1994**
  - 4 to 64 processors at 1GHz
  - 8GB of memory
  - 32 Gflops
  - Went back to B & T registers, due to local memory failure
Cray Research Superservers

- 1991 – A spinoff effort to bring minisupercomputers to the file server and networking markets
- Acquired and modified SPARC-based systems designed by *Floating Point Systems*
- Only produced a few machines – never became successful in the server market
- Eventually renamed *Business Systems Division* before being sold to Sun Microsystems in 1996
Cray Research Superservers

- **APP (1992)**
  - up to 84 processors (Intel i860) arranged in nodes of 12, peak performance of 6.7 GFLOPS
  - acted as a co-processor for SPARC systems

- **S-MP (1992)**
  - eight SPARC processors, 66 MHz
  - supported the APP co-processor

- **CS6400 (1993)**
  - up to 64 SPARC processors, 60 MHz
  - 16 GB RAM, JTAG bus control
  - most successful of the servers, ended up sold to Sun in the acquisition, led to the Sun Enterprise 10000
Cray, Inc.

- *Tera Computer Company* renamed after purchasing Cray Research from SGI, still active today
- Combined Cray Research architectures with systems designed by Tera, NEC Corporation, etc.
- Received funding from the NSA in the mid-2000s, including some classified work on supercomputers
Cray, Inc.

- SX-6 (2001)
  - Adapted from NEC’s own 8-processor SX-6 design
  - 64-bit scalar unit, 72 vector registers of 256-word length
  - Multi-node versions provided up to 8 TFLOPS
  - Ran NEC’s SUPER-UX, a UNIX-based supercomputer OS

  - A more manufacture-friendly upgrade of Tera’s MTA system
  - Improved multi-threading and thread synchronization
Cray, Inc.

- **Red Storm (2004)**
  - A simulation machine designed for the US Dept. of Energy
  - Over 10,000 single-core AMD Opterons at 2.0 GHz (512 devoted to user interface and running a Linux variant)
  - In 2006, ranked the 2nd fastest in the world (101.4 TFLOPS)

- **XT3 (2004)**
  - Commercial adaptation of the Red Storm architecture

- **XT4 (2006)**
  - Updated interconnects, Opterons, DDR2 memory
  - Added FPGA co-processor support
Cray, Inc.

- **X1 / X1E (2003 / 2005)***
  - Combined CMOS processing structure of the SV1, memory structure of the T3E, and liquid cooling of the T90
  - Partially funded by NSA, which supported future systems as well
  - Up to 4096 processors could provide ~150 TFLOPS, although 512 processors is the maximum known (non-classified) implementation

- **XT4 / XT5 (2006 / 2007)***
  - Upgrades to the XT series, using a 3D torus network of Opteron processors
  - The *National Center for Computation Sciences* has two, including “Jaguar” which now runs 224,256 processors at up to 1.75 PFLOPS
  - “Jaguar” was ranked as the world’s fastest in 2009 (after an upgrade)
Cray, Inc.

- Jaguar Specs:

<table>
<thead>
<tr>
<th>Jaguar Specifications</th>
<th>XT5</th>
<th>XT4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Teraflops</td>
<td>2,332</td>
<td>263</td>
</tr>
<tr>
<td>Six-Core AMD Opterons</td>
<td>37,376</td>
<td></td>
</tr>
<tr>
<td>Quad-Core AMD Opterons</td>
<td>7,832</td>
<td></td>
</tr>
<tr>
<td>AMD Opteron Cores</td>
<td>224,256</td>
<td>31,328</td>
</tr>
<tr>
<td>Compute Nodes</td>
<td>18,688</td>
<td>7,832</td>
</tr>
<tr>
<td>Memory (TB)</td>
<td>299</td>
<td>62</td>
</tr>
<tr>
<td>Memory Bandwidth (GB/s)</td>
<td>478</td>
<td>100</td>
</tr>
<tr>
<td>Disk Space (TB)</td>
<td>10,000</td>
<td>750</td>
</tr>
<tr>
<td>Interconnect Bandwidth</td>
<td>374</td>
<td>157</td>
</tr>
<tr>
<td>Floor Space (ft²)</td>
<td>4,352</td>
<td>1,344</td>
</tr>
<tr>
<td>Cooling Technology</td>
<td>Liquid</td>
<td>Air</td>
</tr>
</tbody>
</table>
Cray, Inc.

- **XD1 (2004)**
  - Cray Inc.’s main Entry-Level machine
  - Up to 144 Opterons
  - Replaced PCI with Hypertransport, added Xilinx Virtex-II Pro FPGA accelerators

- **XMT (2006)**
  - Latest version of Tera’s MTA family
  - Supports 8192 processors at 500 MHz, 128 TB RAM

- **CX1 (2008)**
  - Workstation featuring up to 92 Intel cores, Nvidia GPU, RAID storage

- **CX1000 (2011)**
  - Cray machine between Entry-Level and High-End levels, Intel Xeon based
  - Multiprocessor focused (SM) and cluster computer focused (SC) variants
  - Aimed at mid-range engineers, scientists, researchers
Cray, Inc.

- Cray’s most current high-end family:
  - XT6 (2009)
    - Up to 13,000 eight and twelve-core AMD Opteron 6100s
    - Customized *Cray Linux Environment* operating system
    - Pairs of two processors are matched with 64 GB RAM, called one node
  - XE6 (2010)
    - Added more scalable interconnects at the nodes, new version of the OS
  - XK6 (2011)
    - Upgraded to 16-core Opteron 6200, added an Nvidia Tesla GPGPU to every node
    - The first ordered machine was to the Swiss National Supercomputing Centre

- Two XE6 installations are currently ranked as the #6 and #8 fastest supercomputers in the world (via TOP500, summer 2011)
Future Cray Computers

- Cray will be updating the “Jaguar” (XT5) to the new “Titan” (XK6), estimating 10 to 20 PFLOPS performance ($97M contract)

- Ongoing work with NASA, Berkeley Labs, Oak Ridge Labs will produce improved systems and groundbreaking new applications

- New machines in the XK6 and CX1000 (high-end and mid-range) families are expected in the following years (beyond 20 PFLOPS?)