1. Parallel Computing Hardware

[1]

[1] M. Ganesh, MATH440/540, SPRING 2018

Parallel Scientific Computing[2]

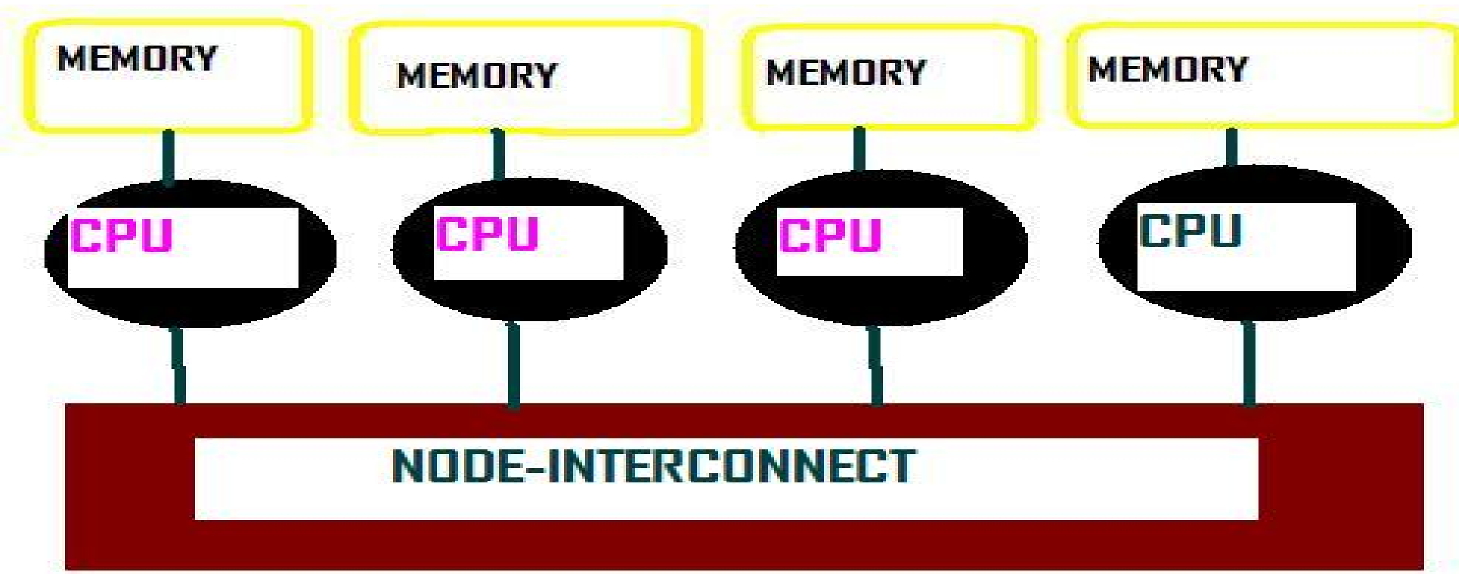
Solve a mathematical system, modeling a scientific problem (requiring a large number of floating point operations - many independent), *faster* by using *parallel programming* in a *parallel computing environment*.

- **Parallel computing environment: A Multiple-core/processor system that supports parallel programming**
- **Parallel programming: Programming in a language that supports concurrency explicitly**
- **Hardware architectures in Parallel computing environment:**
  - ***Multicomputer* Distributed Memory Systems – DMP (distributed memory parallelization)**
  - ***Multiprocessor* Shared Memory Systems – SMP (symmetric multiprocessing)**
  - ***Cluster* Hierarchical Memory Systems – a hybrid of DMP and SMP**

[2] M. Ganesh, MATH440/540, SPRING 2018

Multicomputer Distributed Memory Systems[3]

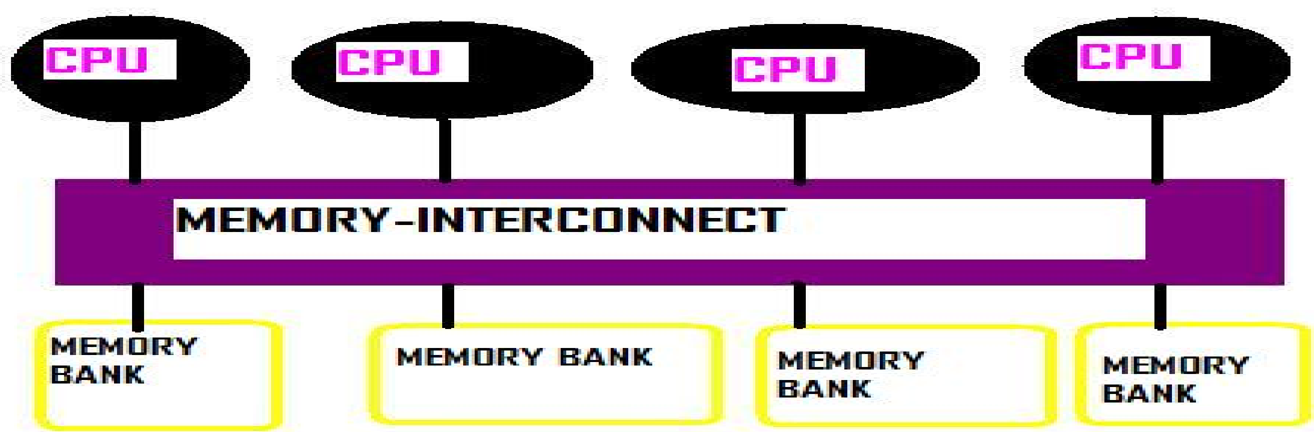
- Desktops/nodes/CPUs are coupled by a node-interconnect
- Each processor mainly has access mainly to its own memory
- Non-uniform memory access (NUMA)
- Node-interconnect includes several Ethernet links/switches with slow speed 10MB/sec connection to faster connection with Gigabit Switches.



[3] M. Ganesh, MATH440/540, SPRING 2018

Multiprocessor Shared Memory Systems[4]

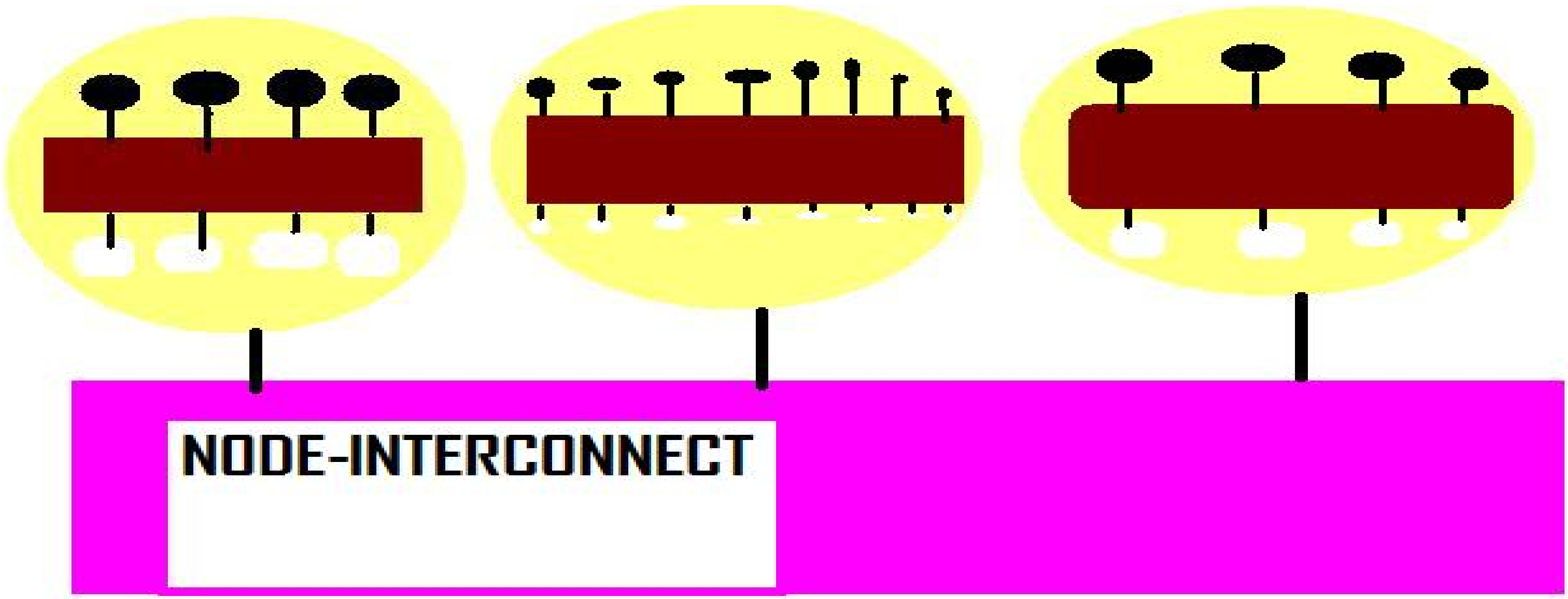
- All CPUs are connected to all memory banks with same memory-interconnect speed
- Uniform Memory Access (UMA)
- Symmetric Multi-Processing
- Memory interconnect:
  - ⋆ Bus: One CPU can block the memory access of the other CPU
  - ⋆ Crossbar: Independent access from each CPU



[4] M. Ganesh, MATH440/540, SPRING 2018

Cluster Hierarchical Memory Systems[5]

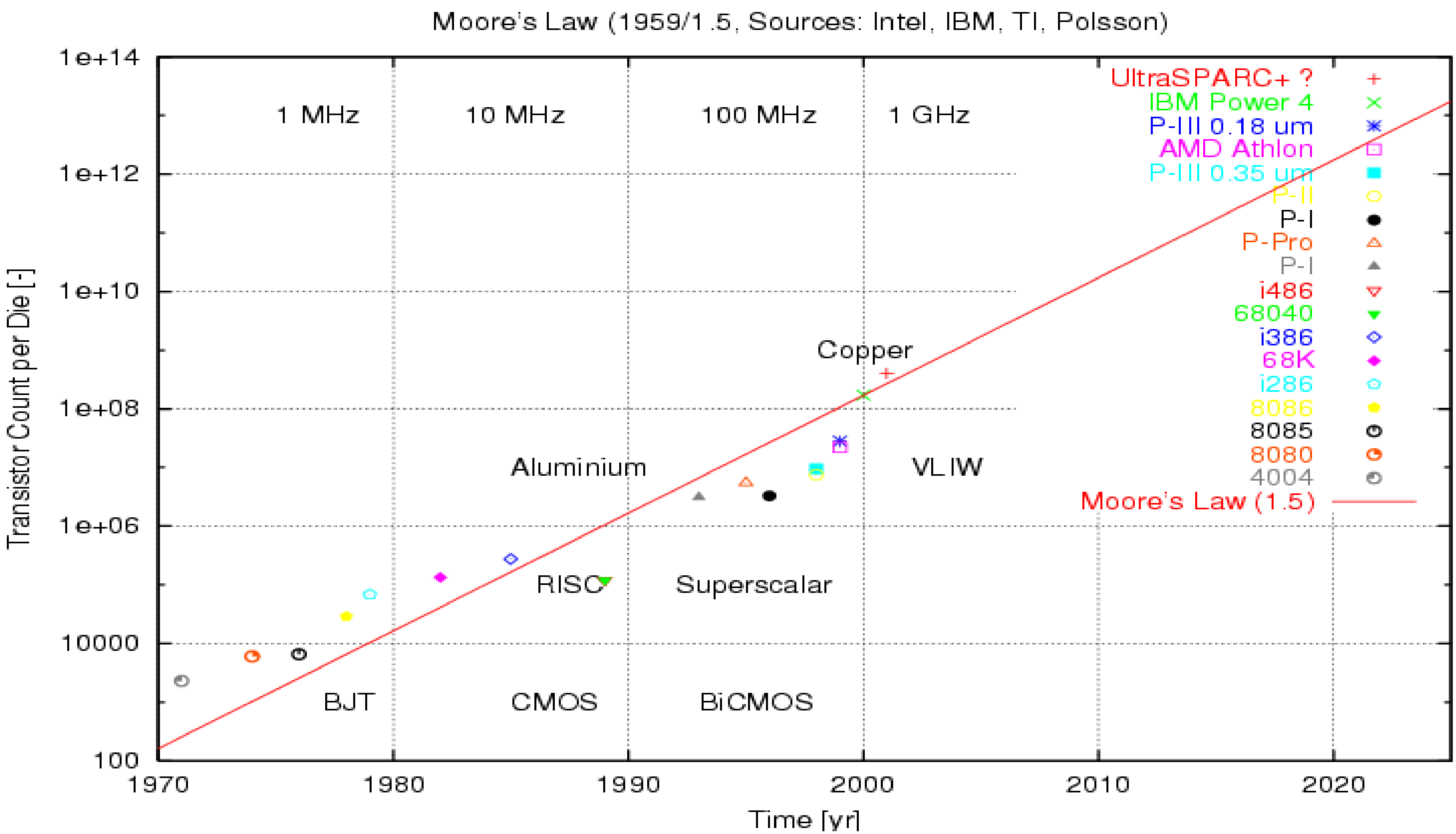
- Almost all current HPC (high-performance computing) systems are clusters of SMP nodes.
- SMP inside each of the node
- DMP on the node interconnect

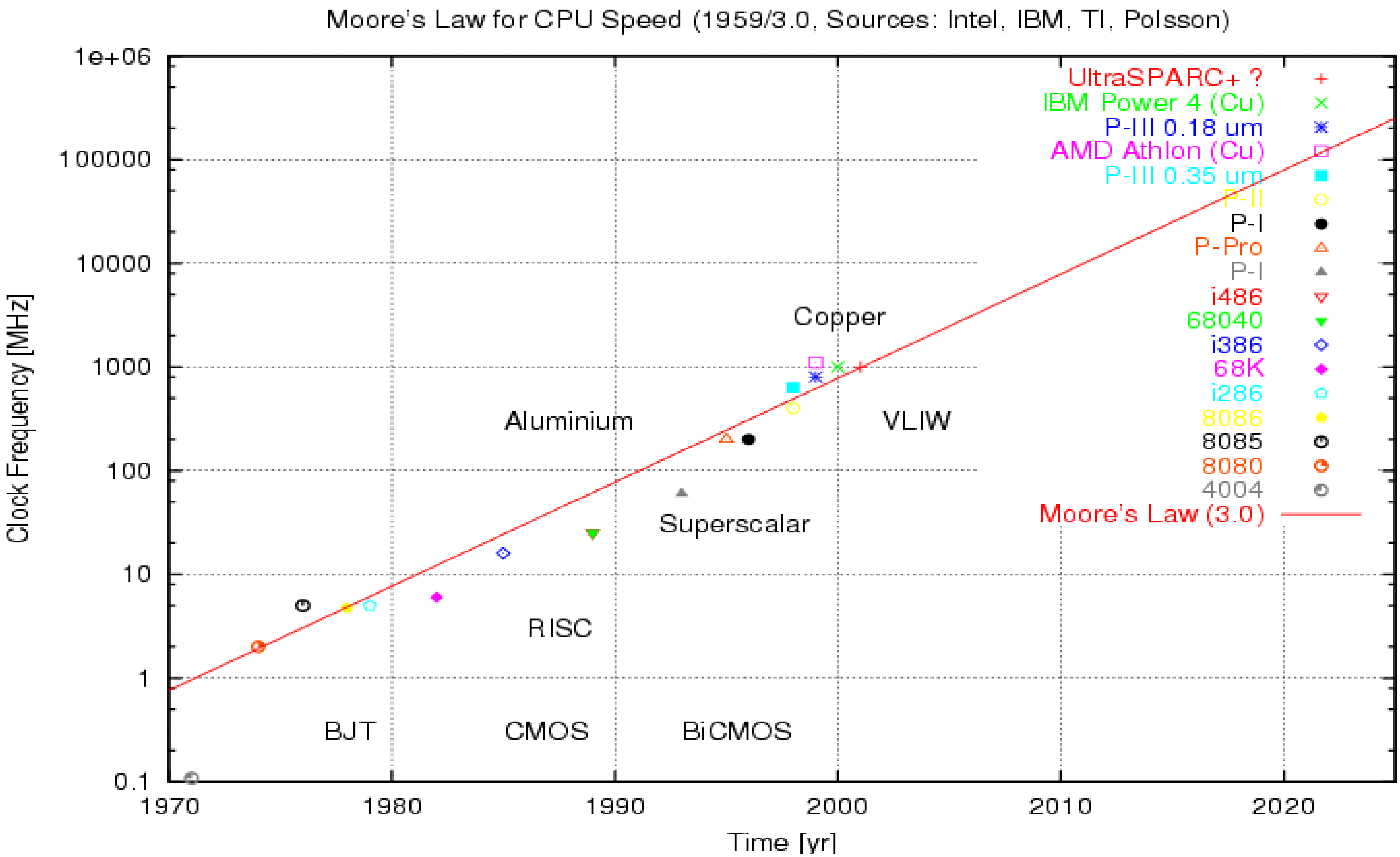


[5] M. Ganesh, MATH440/540, SPRING 2018

Why Parallel Computing Environment?

- Moore's Law: The number of transistors on a chip will double approximately every 18 month




Moore's Law for CPU Speed (1959/3.0, Sources: Intel, IBM, TI, Polsson)
Clock Frequency [MHz]
Time [yr]
1e+06
100000
10000
1000
100
10
1
0.1
1970
1980
1990
2000
2010
2020
UltraSPARC+ ?
IBM Power 4 (Cu)
P-III 0.18 um
AMD Athlon (Cu)
P-III 0.35 um
P-II
P-I
P-Pro
P-I
i486
68040
i386
68K
i286
8086
8085
8080
4004
Moore's Law (3.0)
Copper
Aluminium
VLIW
Superscalar
RISC
BJT
CMOS
BiCMOS

- [6]**Single core processors are about a million (10^6) times faster than that about 50 years ago.**

- **Why not several trillions times faster(already achieved)?:**

  - ⋆ **Given more transistors architects' failed to build faster single core CPUs**

  - ⋆ **Due to power considerations, clock rate can increase only slowly**

  - ⋆ **Architects instead are succeeding by putting several cores on a chip, leading to multi-core processors**

  - ⋆ **Today even a single CPU machine is a highly parallel system (whether we like it or not!)**

  - ⋆ **Burden more on user/programmer level: parallel programming**

[6]M. Ganesh, MATH440/540, SPRING 2018

- **[7]About 50 years ago, a computer system performance was at best about 10^3 (kilo) FLoating-point Operations Per Second (FLOP/S).**

- **Currently the best computer system performance is over 10^{15} (peta) FLOP/S! (Future: exa-, zetta-, yotta-: $10^{18}, 10^{21}, 10^{24}$-FLOP/S.)**

- **That is, some of the current supercomputers are *trillion* (10^{12}) times faster, compared to about 50 years ago, while single processors are only a *million* (10^6) times faster than that 50 years back!!**

- **How is this possible?**

- **Parallel cluster computing environment, with thousands of cores**

- **Task: Solve scientific computing problems on a cluster computer using several thousands of cores (instead of using just one processor)**

- **To this end, parallel scientific computing is essential**

[7] M. Ganesh, MATH440/540, SPRING 2018

Brief 21st century parallel hardware history [8]

- **In 2000, the fastest computer in the world was IBM's ASCI White**
- **Peak performance of ASCI White was 12 TFLOP/S (TeraFLOP/S, that is 12×10^{12} FLOP/S)**
  - ⋆ **ASCI White had 512-nodes**
  - ⋆ **Total of 8192 processors (that is, 16 CPUs per node)**
  - ⋆ **ASCI White had 6.2 Terabytes of memory**
- **In June 2000, Hitachi SR 8000-F1/112 was ranked 5th in performance**
- **The peak performance of Hitachi was about 2 TFLOP/S**
  - ⋆ **The Hitachi machine had 168-nodes**
  - ⋆ **8-CPUs per node (total of of 1344 processors)**
  - ⋆ 1.3 **TB memory**

[8] M. Ganesh, MATH440/540, SPRING 2018

- [9] **Building of a world fastest supercomputer (in 2002), started in 1999**

- **The supercomputer called Earth Simulator was the fastest computer in the world from 2002 to 2004**

- **The peak performance of the Earth Simulator was about 35 TFLOP/S**

- **The NEC built Earth Simulator comprised:**
  - $\star$ 640**-nodes**
  - $\star$ **Total of** 5120 **processors (that is,** 8**-processors per node)**
  - $\star$ 16**GB memory per node (that is, total of about** 10**TB memory)**

- **IBM's first in Blue Gene Series, Blue Gene/L, achieving about** 360 **TFLOP/S in late 2004, replaced the Earth Simulator as the fastest computer.**

- **In May 2008, NEC announced building of a new Earth Simulator**

[9] M. Ganesh, MATH440/540, SPRING 2018

[10]State-of-the-art hardware/performance 2008-2016

- **The buzz HPC phrase in 2008 was: Peta-scale Computing**
- **That is, to build and carry out parallel scientific computing on a parallel computer with 10^{15} FLOP/S performance**
- **The Peta-FLOP/S barrier was overcome in May 2008 using the Los Alamos Lab supercomputer, called the Roadrunner, built by IBM.**
- **The cost of Roadrunner was over \$100 million**

- **Roadrunner was a hybrid cluster computer**
- **Hybrid because it was built using two distinct class of processors: AMD Opteron and CELL (Cell Broadband Engine Architecture).**
- **[11]Roadrunner was the fastest computer in the world (in 2008-2009):**
  - ⋆ **Comprising $6,480$ dual-core AMD Opteron (1.8 GHz) processors and $12,960$ CELL (IBM PowerXcell 8i; 3.2GHz) chips**

[10]M. Ganesh, MATH440/540, SPRING 2018
[11]M. Ganesh, MATH440/540, SPRING 2018

- ⋆ **Each CELL processor had 9 cores and hence Roadrunner had a total of** $129,600$ **cores (**$12,960$ **Opteron cores +** $116,640$ **CELL cores)**

- ⋆ **Comprising of** 296 **racks with** 18 **connected units**[12]

- ⋆ **Each connected unit has** 180 **Triblades (with Infiniband switches)**

- ⋆ **Each Triblade (with Opteron, expansion, CELL blades) had:**
  two Opteron processors ($18 \times 180 \times 2 = 6,480$**) with 16G memory and four CELL chips (**$18 \times 180 \times 4 = 12,960$**) with** 16**GB memory.**
  So in total Roadrunner had $18 \times 180 \times 32$**GB (**$= 103.680$**TB) memory**

[12] or 17 connected units; $17 \times 180 \times 4 = 12,240$ CELL chips and $17 \times 180 \times 2 = 6,120$ and extra 442 to make $6,562$ Opteron processors, reported by IBM.

[13]How fast'efficient was Roadrunner (a Peta-FLOP/S performance)?

- **Within the first decade of this century supercomputer power increased 1000-fold**
- **Three of Roadrunner Triblades were as fast as the 1998 supercomputer**
- **A parallel scientific computing job requiring one week on Roadrunner to complete, would have taken the 1998 machine 20 years to finish (i.e., the job wouldn't even completed now, even if it were submitted on the machine in 1998 and continuously running)**
- **If it were possible to achieve such a great improvement in performance within a decade, for example, in fuel efficient car technology, today it would be possible to drive a car with fuel efficiency several thousands of miles per gallon!**
- **If each and every human on the earth were asked to use one calculator each and do simultaneously the amount of FLOP/S Roadrunner can do in 24 hours, it would take more than four decades for the entire human population.**

[13]M. Ganesh, MATH440/540, SPRING 2018

[14]

- ***Roadrunner*, with achieved performance of 1.04 petaflop/s was ranked at number ten in the June 2011 TOP500 List of World's Supercomputers (`top500.org`) (and is no longer in the list of top 10 computers!)**

- **Achieved performance is measured by running the *Linpack* (a Linear Algebra package, `http://www.netlib.org/linpack/`)**

- **For the first time in history, in 2011, all top ten computers in the TOP500 list achieved at least one petaflop/s performance**

- **In June 2011, a Japanese supercomputer, called the *K Computer*, with peak performance of 8 petaflop/s was declared as the world's fastest supercomputer. (This put Japan back on the top spot since the Earth Simulator was dethroned in Nov. 2004.)**

- **The name *K* for the supercomputer was chosen to reflect the Japanese word *Keo*, for 10^{16} (10 petaflop/s)**

[14] M. Ganesh, MATH440/540, SPRING 2018

- [15]**The *K Computer* consists of a total of $705,024$ SPARK64 CPU processing cores (comprising several eight-core CPUs)**

- **Currently (January 16, 2018) the *K Computer* is the tenth fastest supercomputer.**

- **On June 17, 2013, China's Tianhe-2 (Milky Way-2) was declared as the (TOP500 list) fastest supercomputer in the world.**

- **Tianhe-2 was originally scheduled to complete only in 2015 and hence was not expected to surpass in 2014, the previously top ranked USA Department of Energy (DoE) systems, described below.**

- **Until May 2016, Tianhe-2 was the fastest supercomputer in the world, retaining the top position for several consecutive ranking periods. Currently, it is the second fastest supercomputer in the world.**

- **Currently (2018), China's Sunway TaihuLight is the fastest machine.**

[15]M. Ganesh, MATH440/540, SPRING 2018

⋆ [16]**Titan (a Cray machine) and**

⋆ **Sequoia (an IBM BlueGene/Q machine).**

- **Sequoia (with 1,572,864 PowerPC CPU cores) was the first machine in the world to cross the** 10 **petaflop/s sustained performance and achieved** 17.17 **petaflop/s on the Linpack benchmark**

- **Sequoia, currently ranked (Top500 list) as number six, was ranked on June 14, 2012 as the fastest machine in the world**

- **Titan, (with 299, 008 AMD-Opteror-CPU cores accelerated with an almost equal number of NVIDIA-GPU cores) achieved 17.59 petaflop/s on the Linpack benchmark and is currently Titan is one of the most energy efficient supercomputers in the world**

- **Titan is currently ranked (Top500 list) as the fifth fastest machine in the world**

---

[16]M. Ganesh, MATH440/540, SPRING 2018

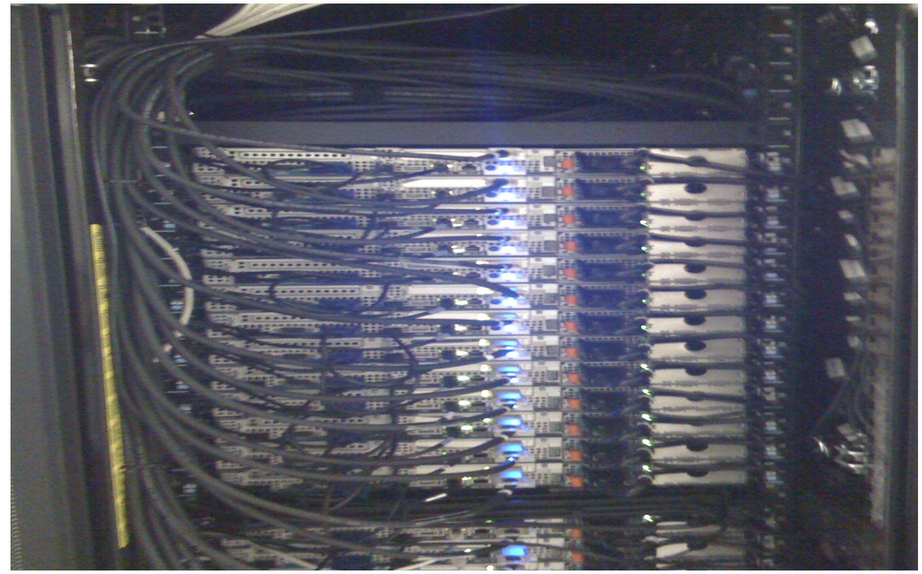
[17]

- **The second fastest machine Tianhe-2 achieved 33.86 petflop/s on the Linpack benchmark**
- **Tianhe-2, a Linux system, has 16,000 compute nodes**
- **Each node has two Intel Xeon (Ivy Bridge) CPUs, 88GB RAM and three 57-core Intel Xeon Phi 8-GB accelerator cards**
- **Thus in total Titan-2 has 1,408,000 RAM (largest ever in a single system) with 3,120,000 cores of which 2,736,000 are Phi-Coprocessor accelerated cores and rest (384,000) CPU cores**

- **The fastest machine Sunway TaihuLight achieved 93.01 petflop/s on the Linpack benchmark**
- **TaihuLight features the custom-designed ShenWei 1.45GHZ SW26010 processors, and has the total of** $10,649,600$ **cores**
- **Each TaihuLight node has 260 cores and 32 GB and the configuration comprises coprocessor, stand-alone processor, stand-alone processor with integrated fabic.**

[17]M. Ganesh, MATH440/540, SPRING 2018

State-of-the-art CSM HPC hardware (2008-2017): RA, MIO, BlueM [18]



- **The CSM supercomputer RA was set up in 2008**
- **RA was part of the TOP500 supercomputer list in Spring 2008**
- **RA had 17 TFLOP/S sustained (and 23 TFLOP/S peak) performance**
- **RA had 268-nodes with each node comprising two quad-core processors and hence had a total of** $268 \times 8 = 2144$ **cores**

[18] M. Ganesh, MATH440/540, SPRING 2018

[19]

- **RA compute nodes were connected by Infiniband switches**

- **All, but 12, RA nodes had INTEL Clovertown 2.67 GHz processors and the 12 nodes were equipped with Xeon 3.4 GHz processors**

- 184 **RA nodes had** 16**GB RAM each**

- **Other** 84 **RA nodes each had** 32**GB RAM**

- **Hence RA had a total of** $184 \times 16 + 84 \times 32 = 5.632$ **TB memory**

- **RA had** 300**TB disk space and same size tape backup**

- **RA nodes and infrastructure were bought in 2008 using funds from the NSF, CSM, and NREL**

- **RA nodes are no longer available for use**

[19] M. Ganesh, MATH440/540, SPRING 2018

[20]

- **Another class of HPC cluster of compute nodes (with environment similar to that of RA) in CSM is called MIO**
- **The concept of MIO was developed in 2010**
- **MIO nodes are owned by individual CSM faculty/groups**
- **In Summer 2011, six CPU MIO nodes and one GPU MIO node were bought by the MCS (Math & CS) group for MCS faculty and CSM students enrolled in MATH and CSCI courses**
- **Each MCS-MIO CPU node consists of** 12 **cores (with a dual Intel Xeon X5670 six-core processor) and** 24**GB RAM (**$6 \times 4GB$**)**
- **In Summer 2015, eight AMS-CPU MIO nodes were were bought and each node comprises** 24 **cores (with a dual Intel Xeon E5-2680 12-core processor) and** 64**GB RAM (**$4 \times 16GB$**)**
- **All MIO nodes are connected by Infiniband switches**

[20] M. Ganesh, MATH440/540, SPRING 2018

[21]

- **Currently, the total number of CPU cores in MIO exceeds 1300 with performance of over 20 Tflop/s**

- **In addition MIO has two GPU nodes, with a total of 2304 cores with peak performance of 7.23Tflop/s The first GPU node was bought in 2010 with 960 GPU cores and in Summer 2011 the MCS group bought a new GPU node filling in remaining of the 2304 GPU cores**

- **In Summer 2013, two new Intel Phi coprocessor enhanced nodes were purchased and these two nodes became operational as part of MIO in August 2013**

- **Each Phi nodes consists of two Intel Xeon 2.3Ghz hexacore CPUs and four Phi 5510P coprocessor cards, each having 60 cores. Thus each of the current CSM Phi nodes have 12 CPU cores and 240 coprocessor Phi-cores**

- **In 2016, two IBM Power 8 GPU enhanced nodes were added to Mio. Each node has 20 Power cores and two Nvidia K80 two-GPU cards**

[21] M. Ganesh, MATH440/540, SPRING 2018

22

- **Students in this course have priority to use the above mentioned nineteen MIO nodes (two Phi + one GPU + two Power + fourteen CPU nodes)**

- **In addition, (for homework, assignments and projects) students in this course are required to use the AMS Sayers Linux Lab computing environment that was set up in August 2015.**

- **The Sayers Lab consists of 24 desktop computers (16 in CH215 and 8 in CH275) (connected by Gigabit switches) with each desktop equipped with an Intel Xeon E3-1271 quad core processor and 16GB memory**

- **BlueM and MIO nodes and lab machines use the Linux operating system**

- **RA was replaced in September 2013 by a CSM supercomputer called BlueM**

[22] M. Ganesh, MATH440/540, SPRING 2018

23

- **BlueM was located at NREL and moved to CSM in Fall 2017**
- **BlueM is an energy efficient five rack machine with IBM dual architecture, split into two units (AuN and MC2)**
- **AuN (Golden) is an iDataPlex machine 144 node machine with each node comprising dual Intel 8-cores processors and 64 GB RAM.**
- **In total, AuN has 2,304 CPU-cores, 9216 GB RAM and achieves 50 TFlops performance**
- **Thus a single AuN node has four times more RAM and three times faster and four time than a standard RA node. AuN environment is similar to that of RA/MIO and hence should be easier to port codes developed on RA/MIO**
- **MC2 (MC^2, Energy) is an IBM Blue Gene Q,** 104 **TFlops,** 512 **node,** $8,192$ **PowerPC (A2 I7) CPU-core,** 8192 **GB RAM machine with each node having** 512 **cores and** 16 **GB RAM**
- MC^2 **of BlueM is designed for large count, highly scalable jobs and has multilevel parallelism.** MC^2 **is not available for use anymore.**

[23]M. Ganesh, MATH440/540, SPRING 2018