# Introduction to Parallel Architecture

R. Govindarajan
Indian Institute of Science,
Bangalore, INDIA
govind@iisc.ac.in

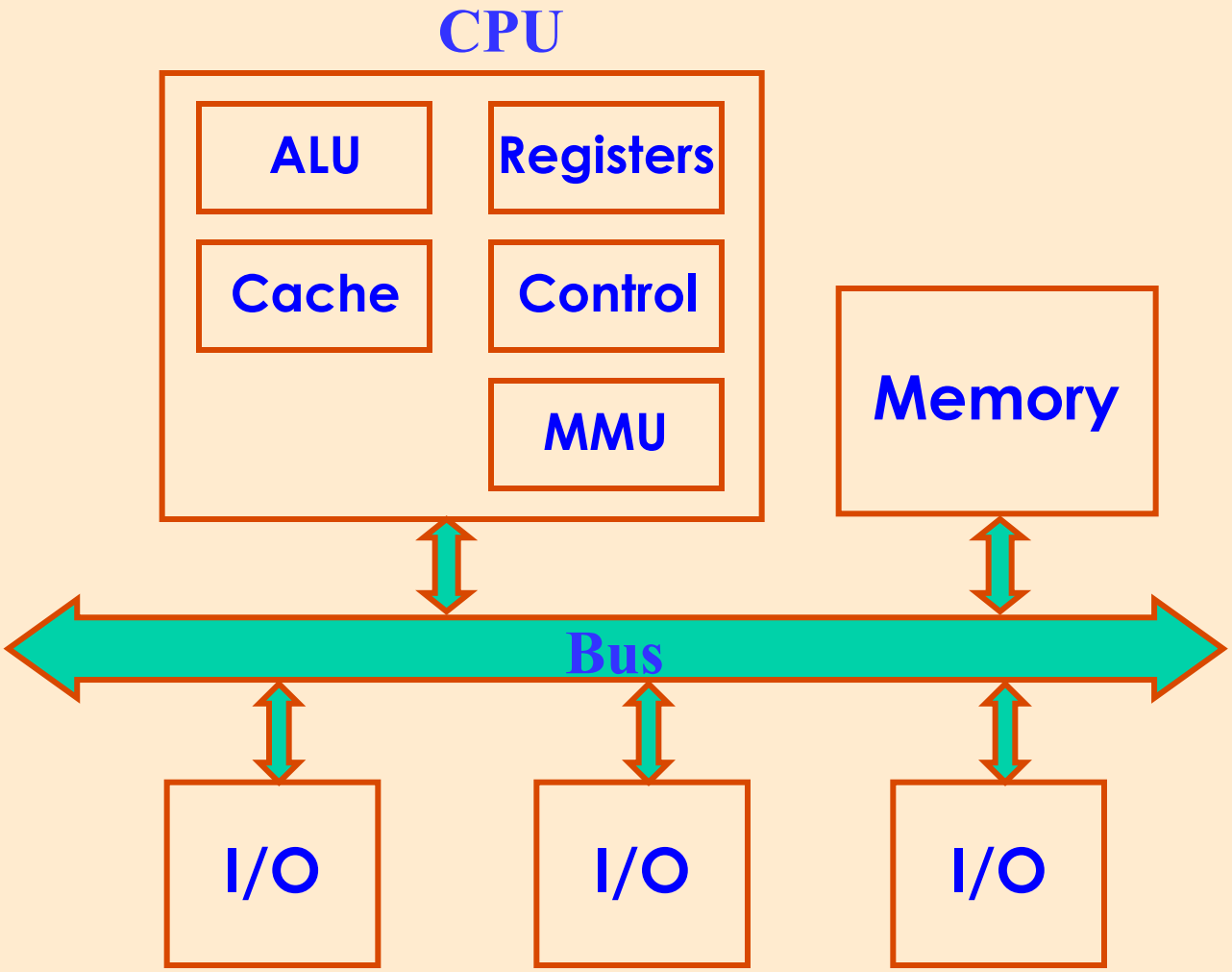
# Overview


- Introduction
- Pipelining, Instruction Level Parallelism
- Multicore Architectures
- Multiprocessor Architecture
  - Shared Address Space
  - Distributed Address Space
- Accelerators – GPUs
- Supercomputer Systems

# Introduction


- Parallelism everywhere
  - Pipelining, Instruction-Level Parallelism
  - Vector Processing
  - Array processors/MPP
  - Multiprocessor Systems
  - Multicomputers/cluster computing
  - Multicores
  - Graphics Processing Units (GPUs) and other Accelerators

# Basic Computer Organization



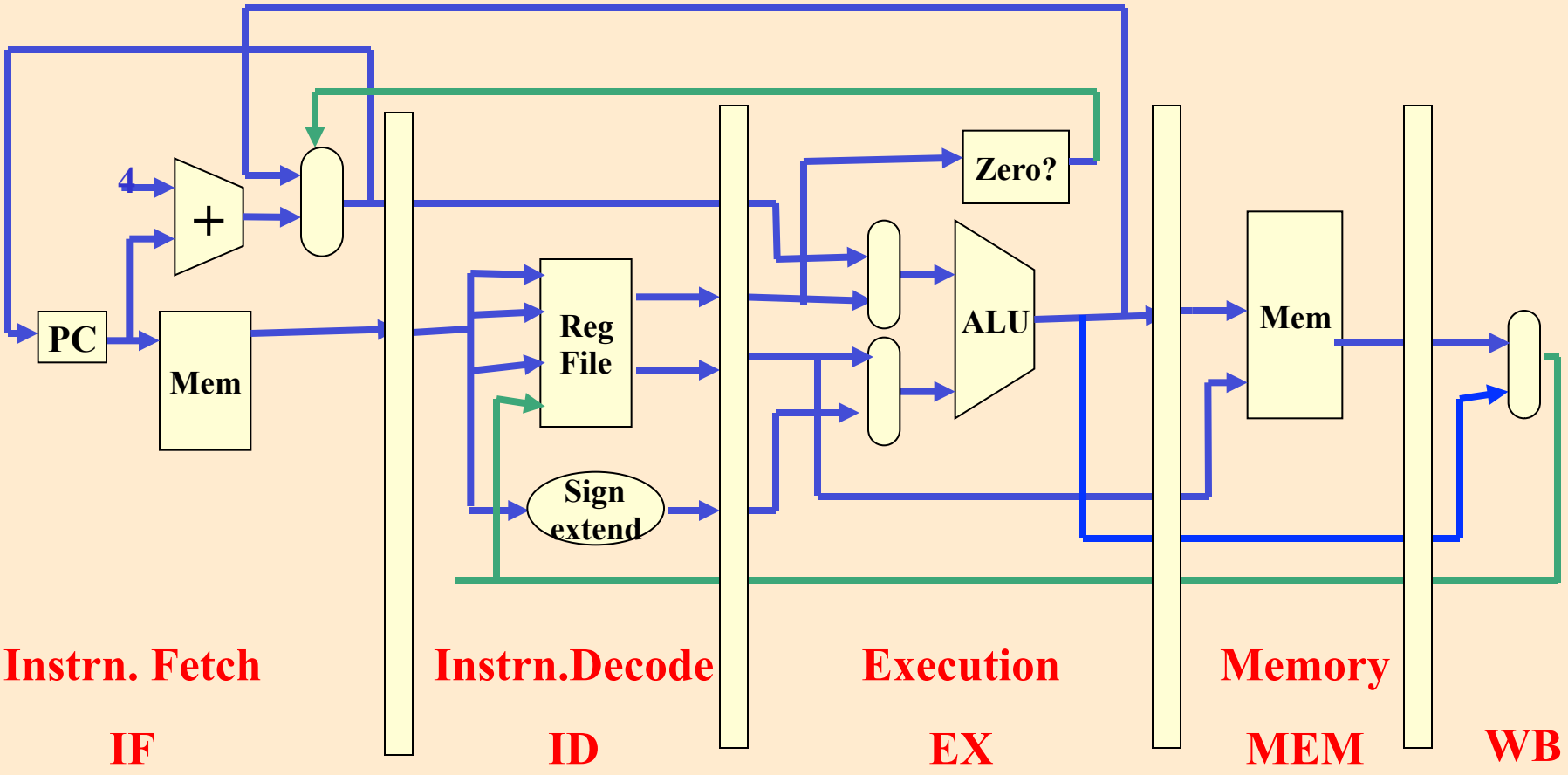
# Overview


- Introduction
- **Pipelining, Instruction Level Parallelism**
- Multicore Architectures
- Multiprocessor Architecture
  - Shared Address Space
  - Distributed Address Space
- Accelerators – GPUs
- Supercomputer Systems

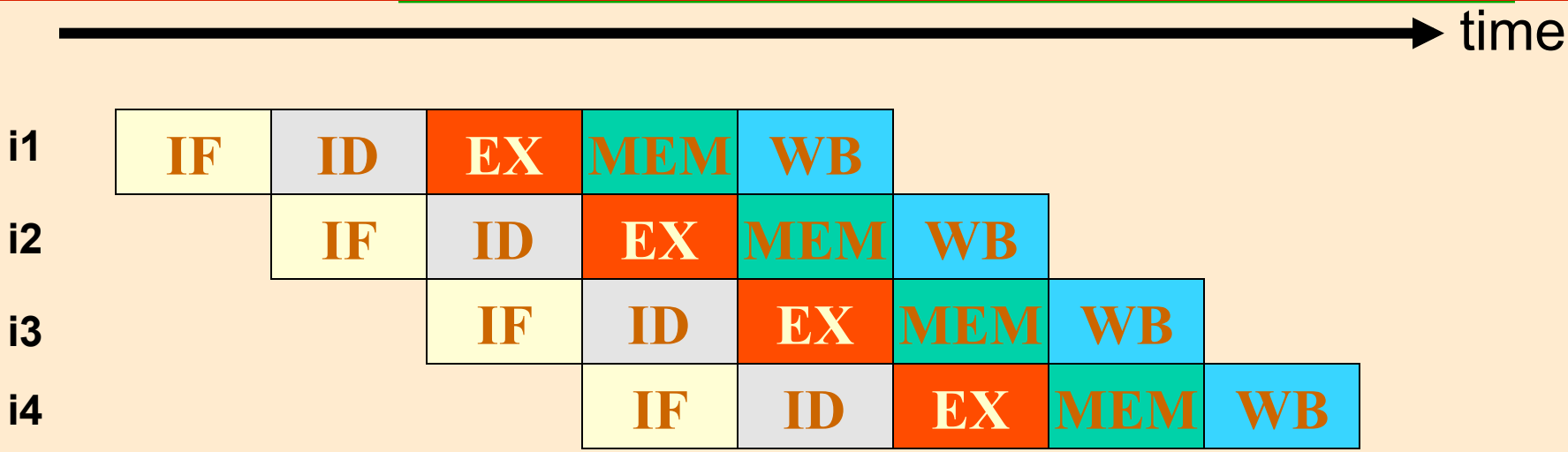
# Pipelined Processor


- Pipelining instruction execution
  - Instrn. Fetch, Decode/Reg.Fetch, Execute, Memory and WriteBack
- Why pipelined exeuction?
  - Improves instruction throughput
  - Ideal : 1 instruction every cycle!

# Processor Datapath



# Pipelined Exeuction

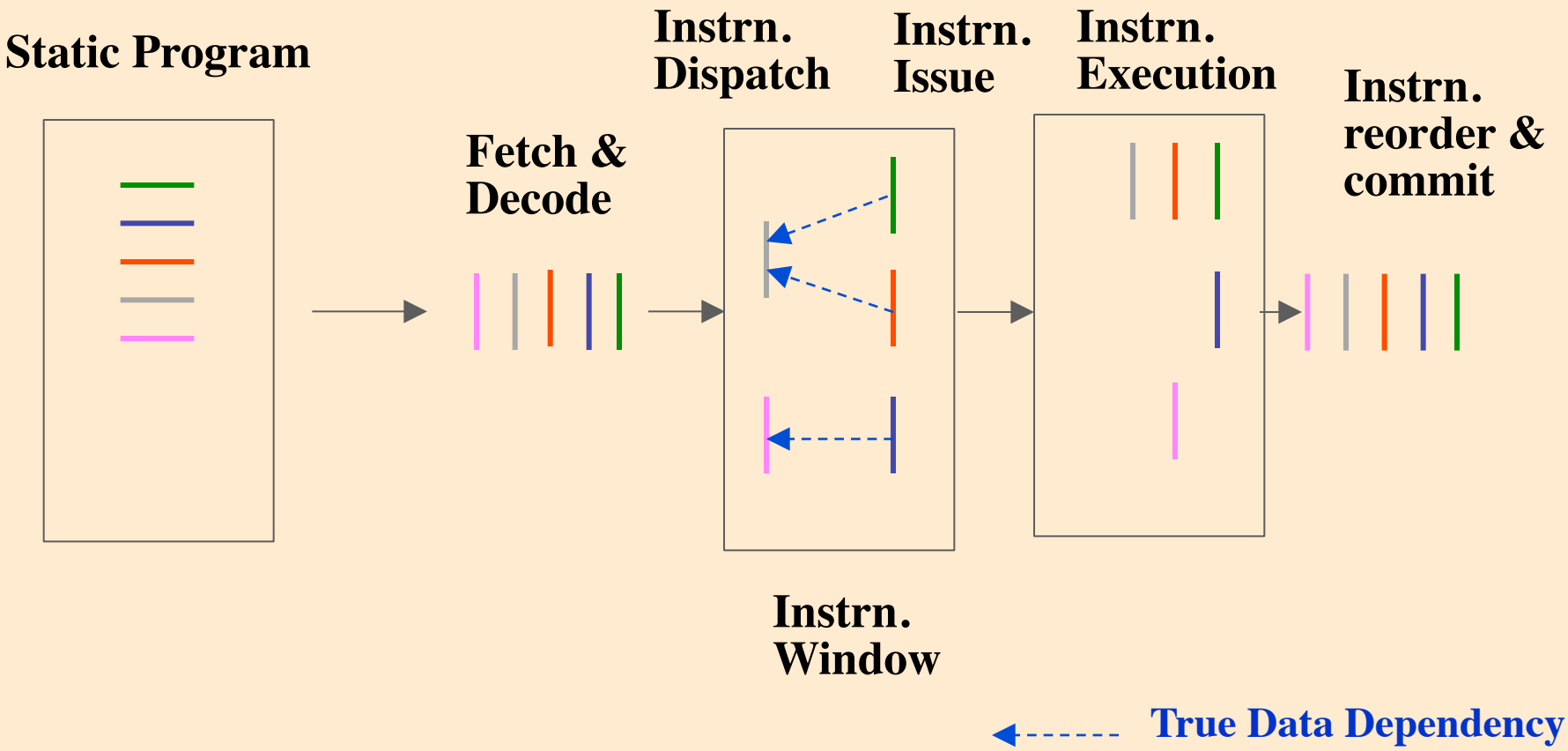



- Execution time of instruction is still 5 cycles, but throughput is now 1 instruction per cycle
- Initial pipeline fill time (4 cycles), after which 1 instruction completes every cycle

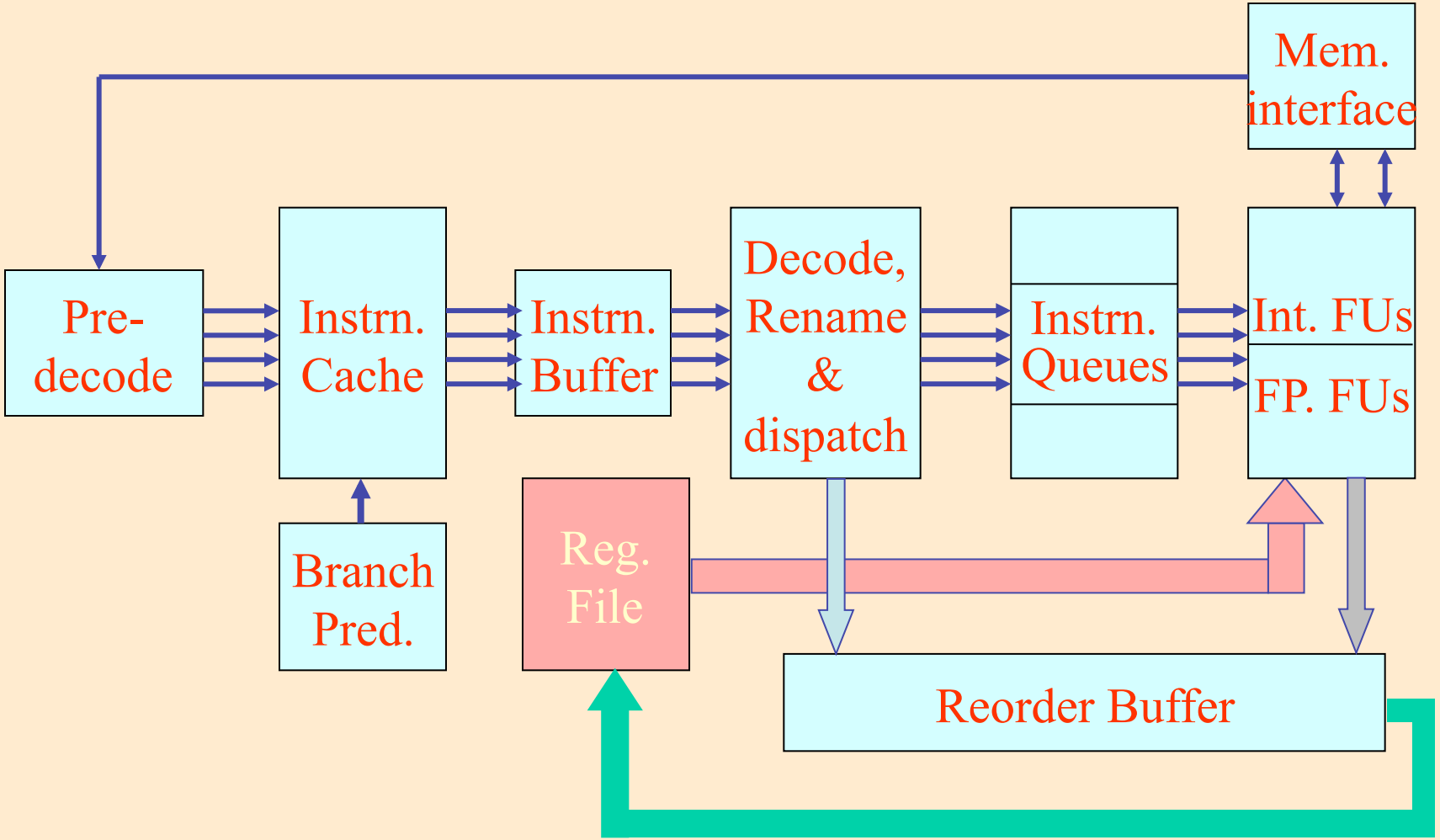
# Instruction Level Parallelism


- Multiple independent instructions issued/ executed together
- Why?
  - Improve throughput (Instrns. Per Cycle or IPC) beyond 1
- How independent instructions are identified?
  - Hardware – Superscalar processor
  - Compiler – VLIW processor

# Superscalar Execution Model



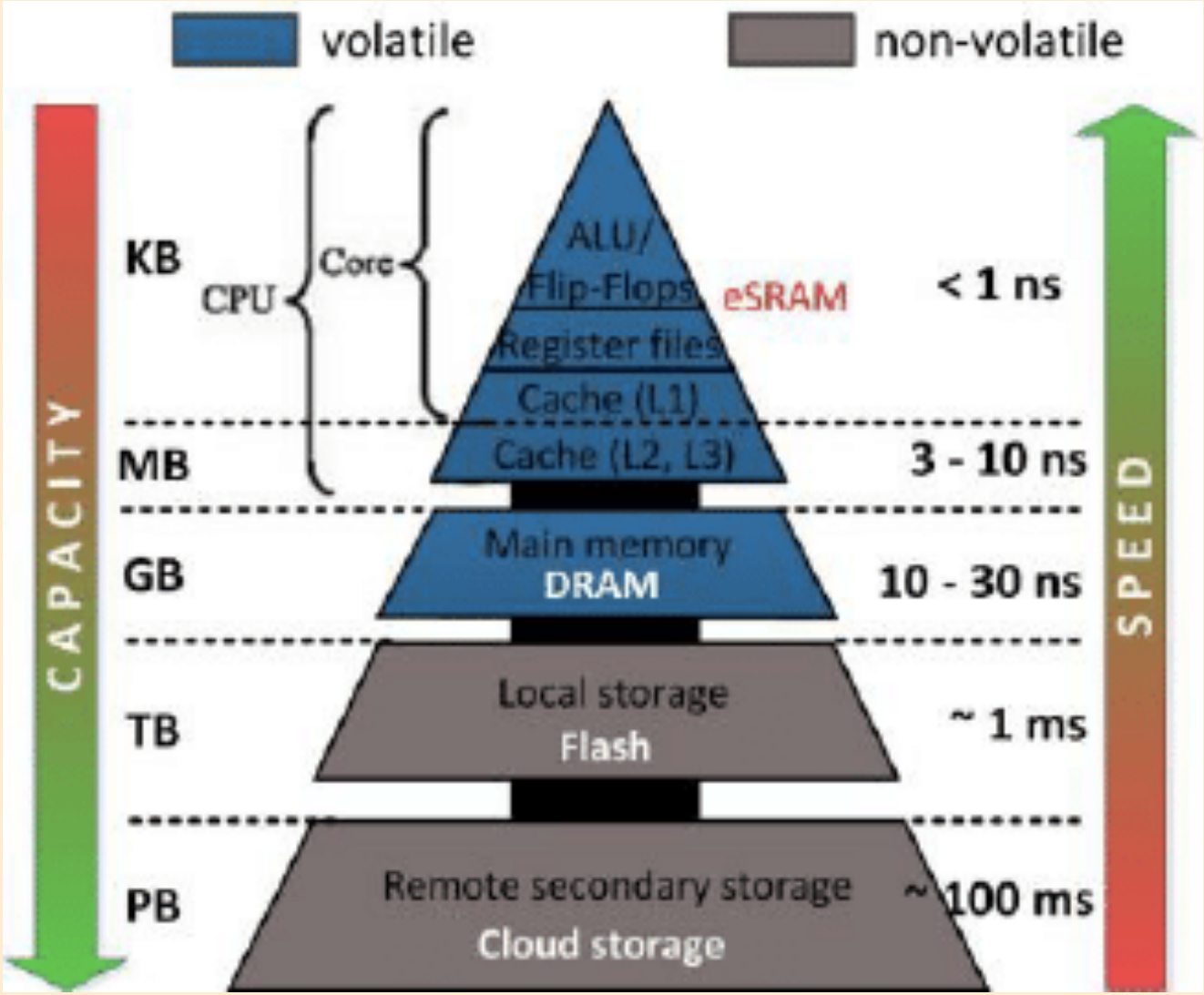
# Superscalar Overview



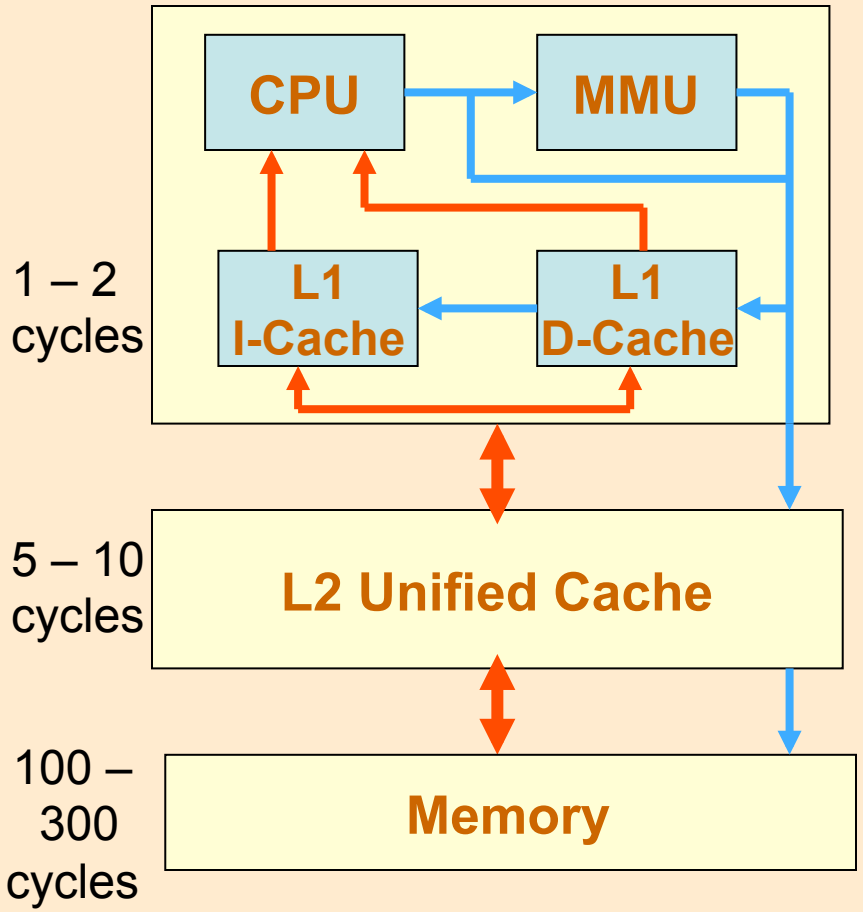
# Memory Hierarchy


- (Pipelined) Instruction execution assumes fetching instruction and data from memory in single cycle.
  - Memory access takes several processor cycles!
- Instruction-level parallelism requires multiple instrn. and data to be fetched in the same cycle.
- Memory hierarchy designed to address this!

# Memory Hierarchy



# Memory Hierarchy : Caches




- Locality of Reference
  - Temporal locality
  - Spatial locality
- Locality in instruction and data
- Extended to L3 and even L4!

- Avg. Memory Access Time (with one level of cache)

  AMAT = hit time of L1 +

  miss-rate * miss-penalty

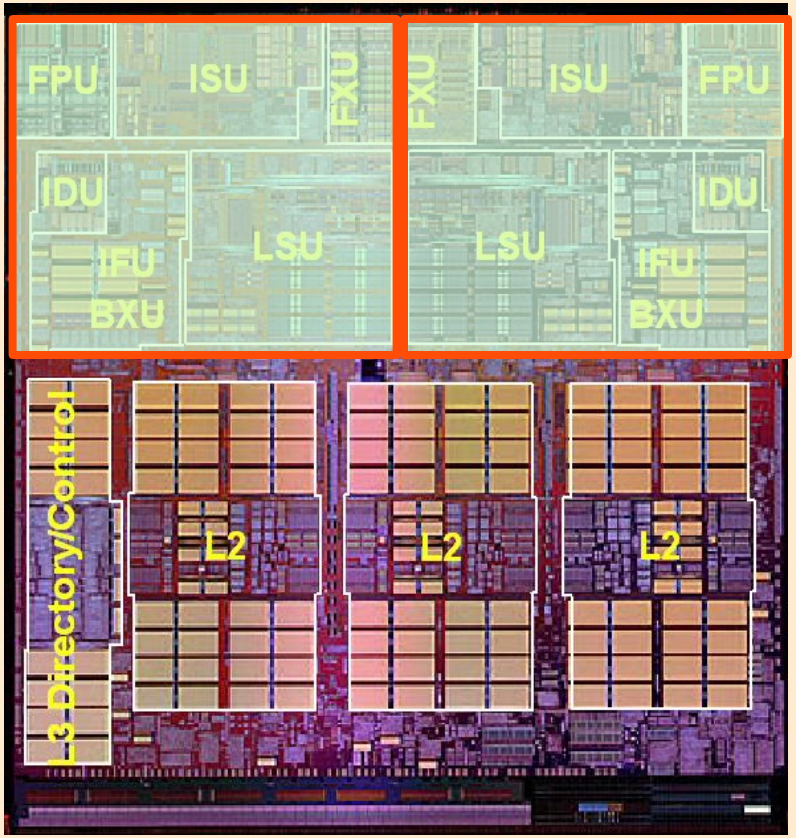
# Overview

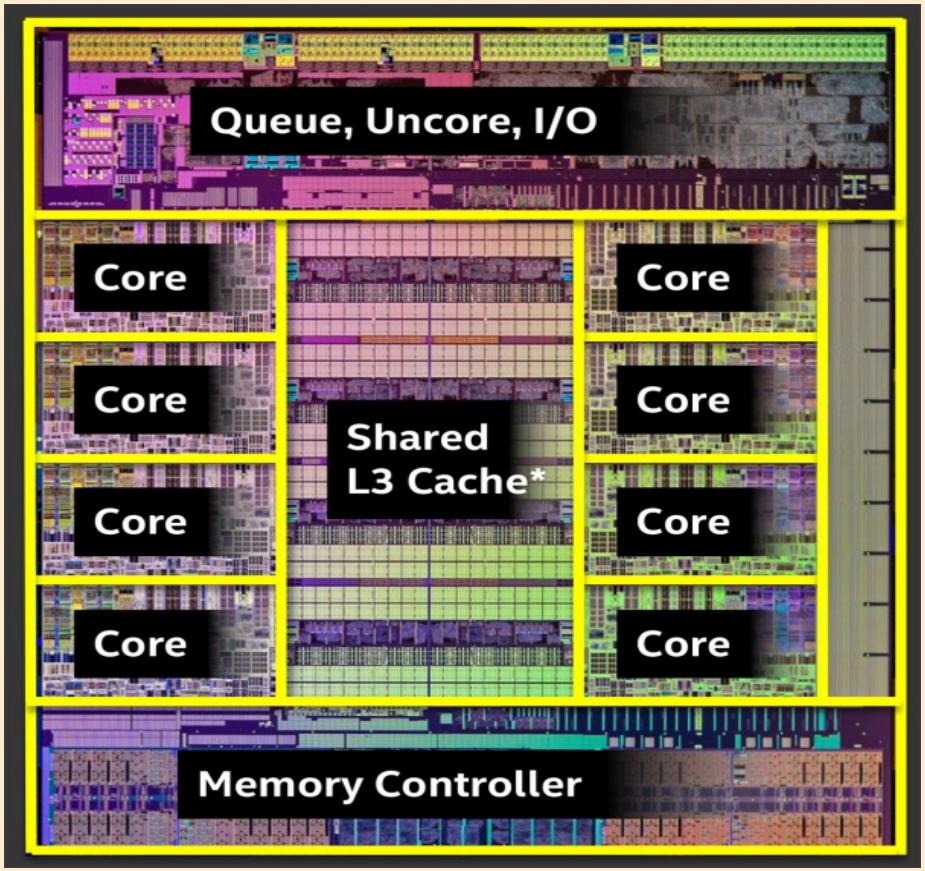

- Introduction
- Pipelining, Instruction Level Parallelism
- **Multicore Architectures**
- Multiprocessor Architecture
  - Shared Address Space
  - Distributed Address Space
- Accelerators – GPUs
- Supercomputer Systems

# Parallelism in Processor

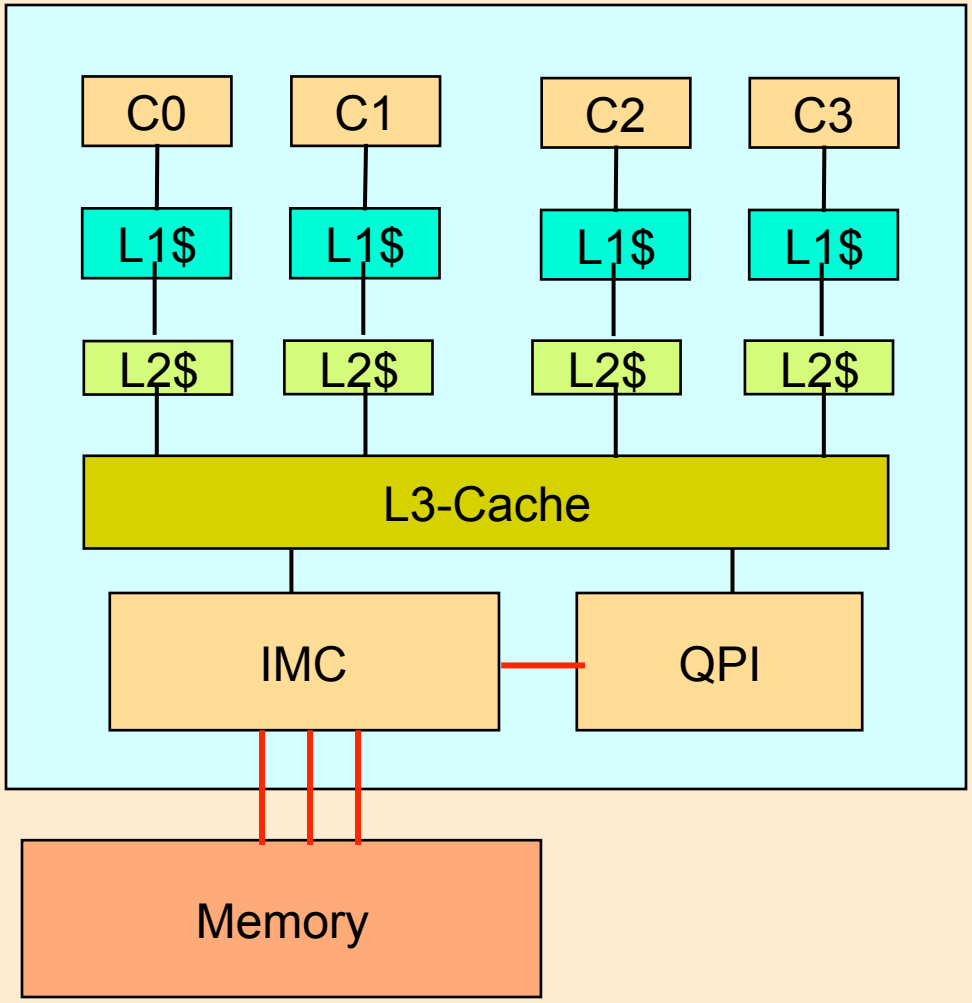

- Pipelined processor
- Instruction-Level Parallelism
- What next? Multicore processors
  - Multiple processors in a single chip
- Why?
  - To improve performance of a single program
  - To execute multiple processes on different cores

# Multicore Processors



# Multicore Processor



# Overview


- Introduction
- Pipelining, Instruction Level Parallelism
- Multicore Architectures
- **Multiprocessor Architecture**
  - **Shared Address Space**
  - **Distributed Address Space**
- Accelerators – GPUs
- Supercomputer Systems

# Classification of Parallel Machines


Flynn's Classification: in terms of number of Instruction streams and Data streams

- *SISD: Single Instruction Single Data*
- *SIMD: Single Instruction Multiple Data*
- *MISD: Multiple Instruction Single Data*
- *MIMD: Multiple Instruction Multiple Data*

# SIMD Machines


- Vector Processors
  - Single instruction on multiple data (elements of a vector – temporal)
- Array Processors
  - Single instruction on multiple data (elements of a vector / array – spatial )
- Modern Processors
  - AVX / MMX instructions
- Graphic Processing Units
  - Multiple SIMD Cores in each Streaming Processors

# MIMD Machines


## Parallel Architecture

- Shared Memory
  - Centralized shared memory (UMA)
  - Distributed Shared Memory (NUMA)
- Distributed Memory
  - A.k.a. Message passing
  - E.g., Clusters

## Programming Models

- What programmer uses in coding applns.
- Specifies synch. And communication.
- Programming Models:
  - Shared address space, e.g., OpenMP
  - Message passing, e.g., MPI

# Shared Memory Architecture

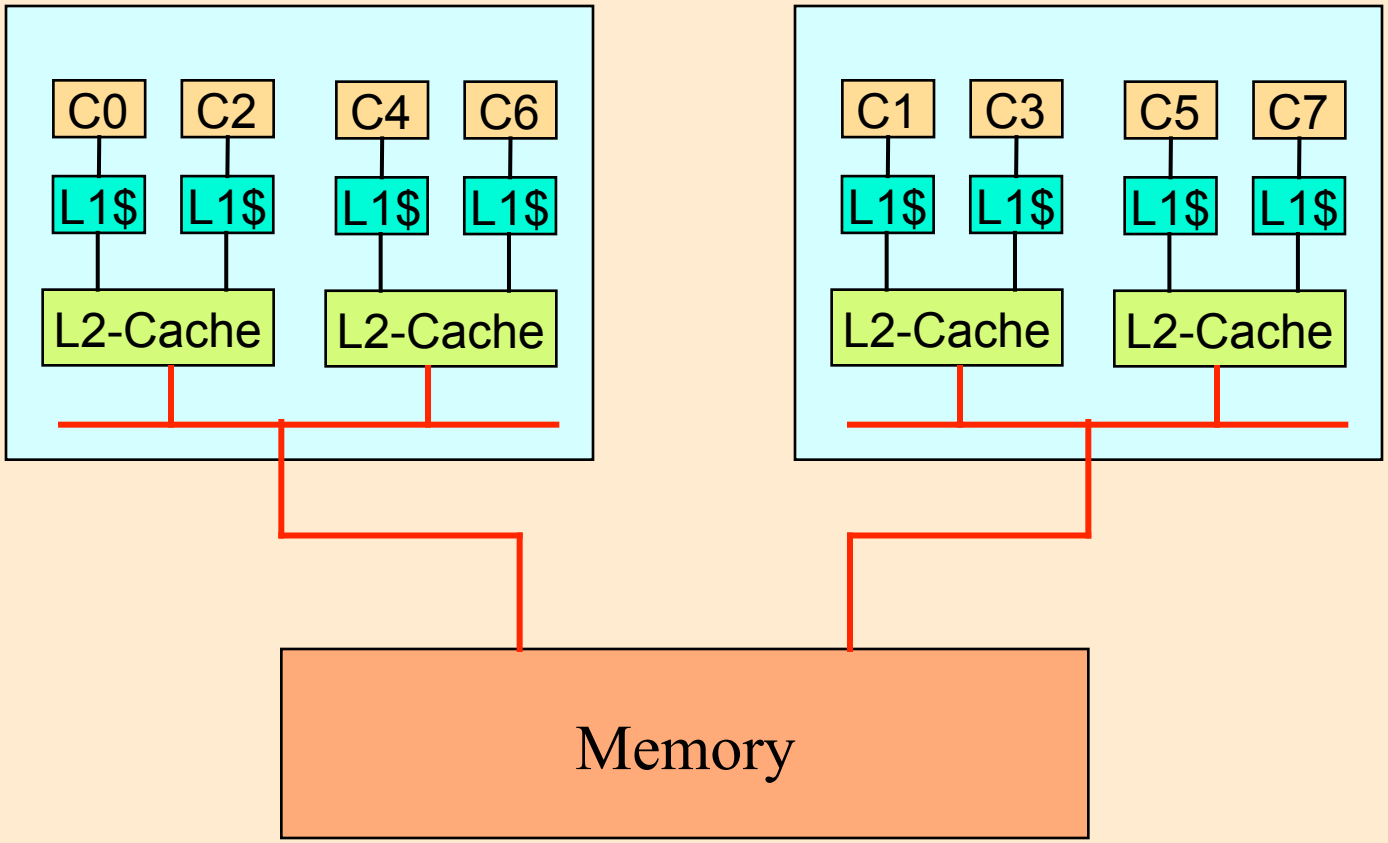

Uniform Memory Access (UMA) Architecture

Non-Uniform Memory Access (NUMA) Architecture

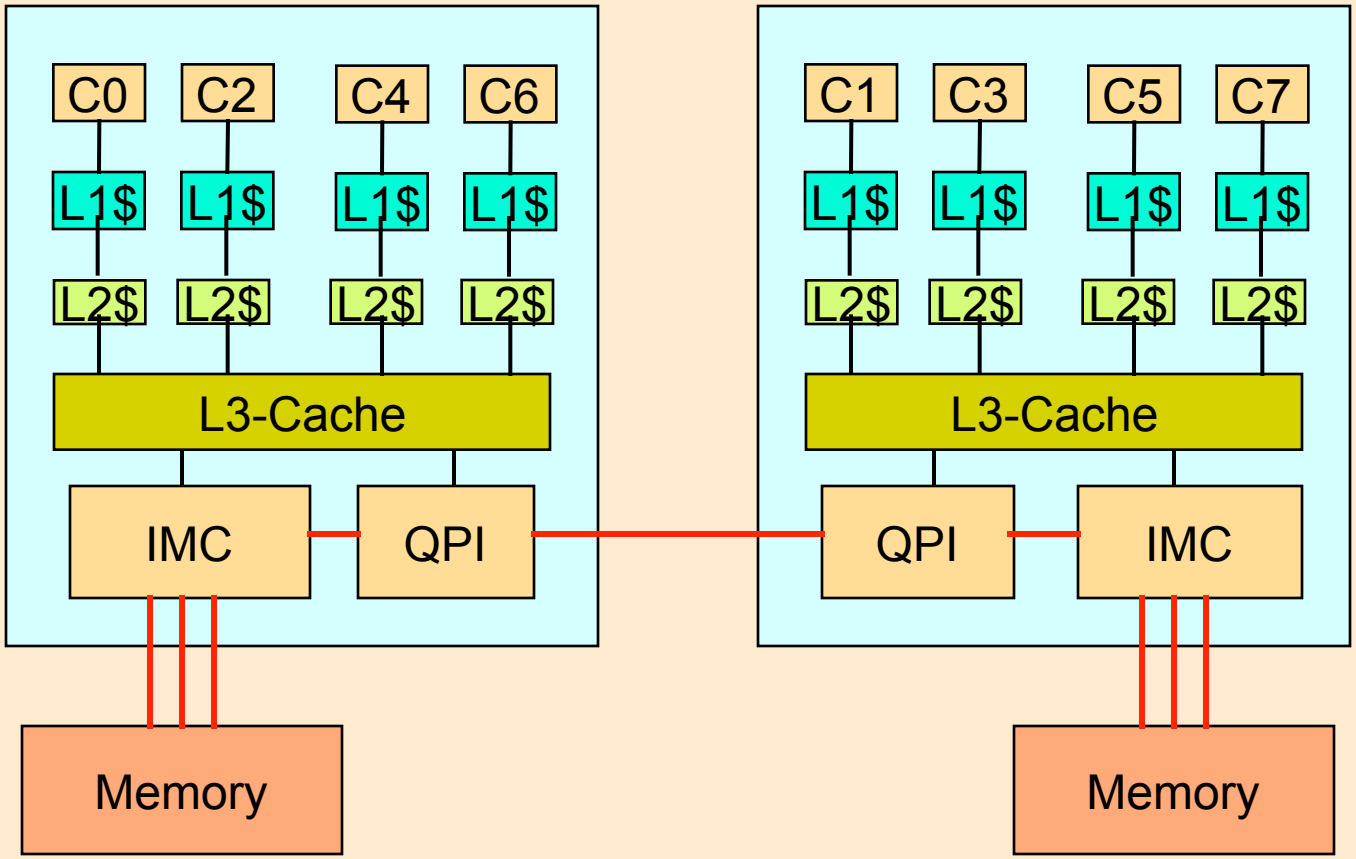
Centralized Shared Memory

Distributed Shared Memory

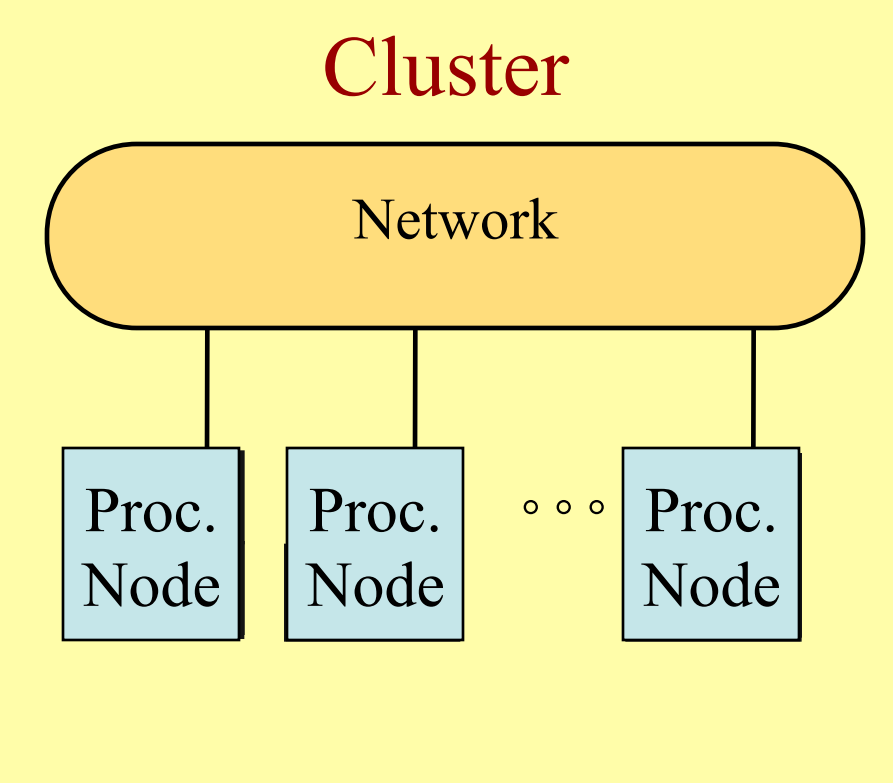
# UMA Architecture



# NUMA Architecture

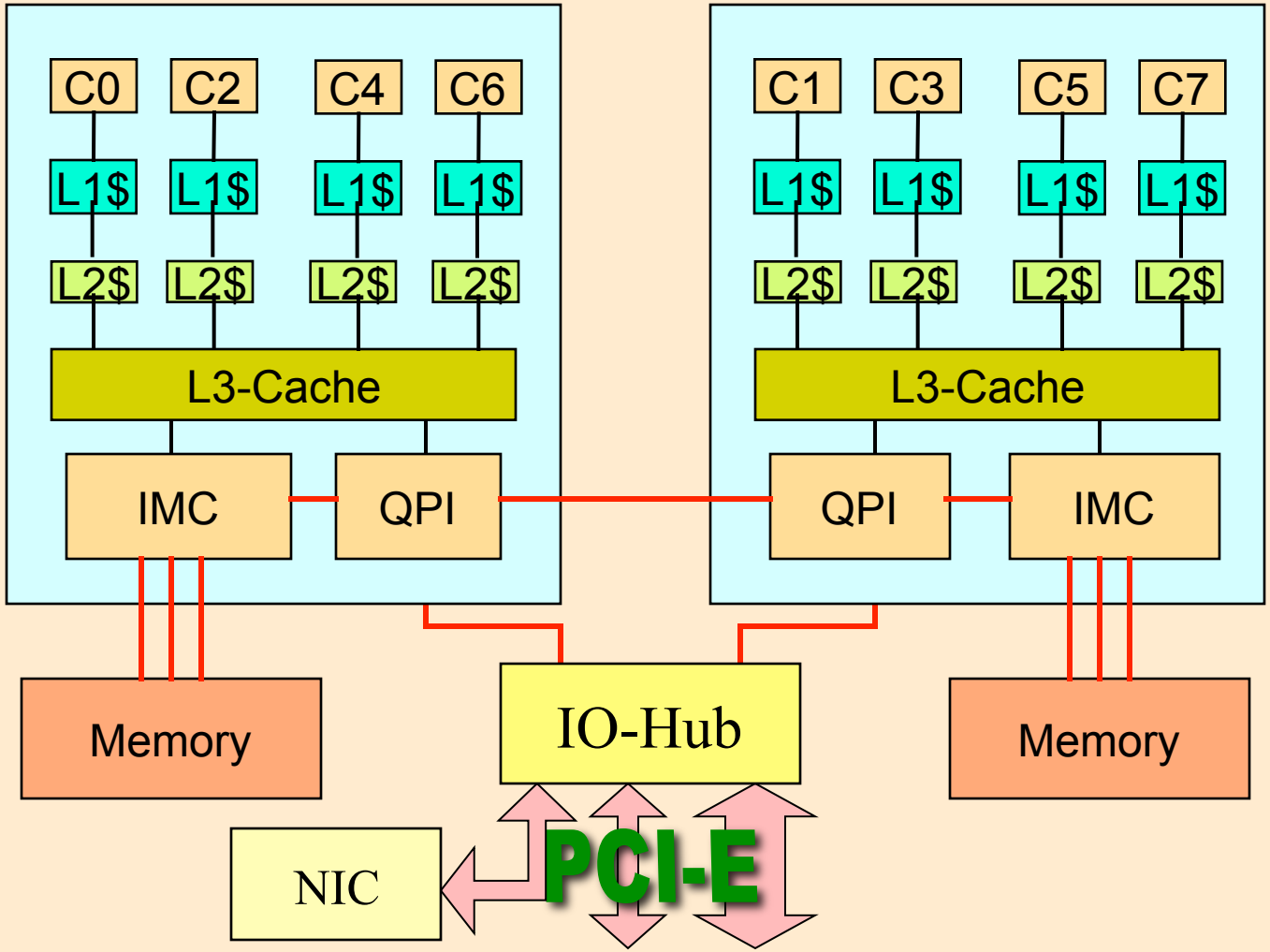


# Distributed Memory Architecture

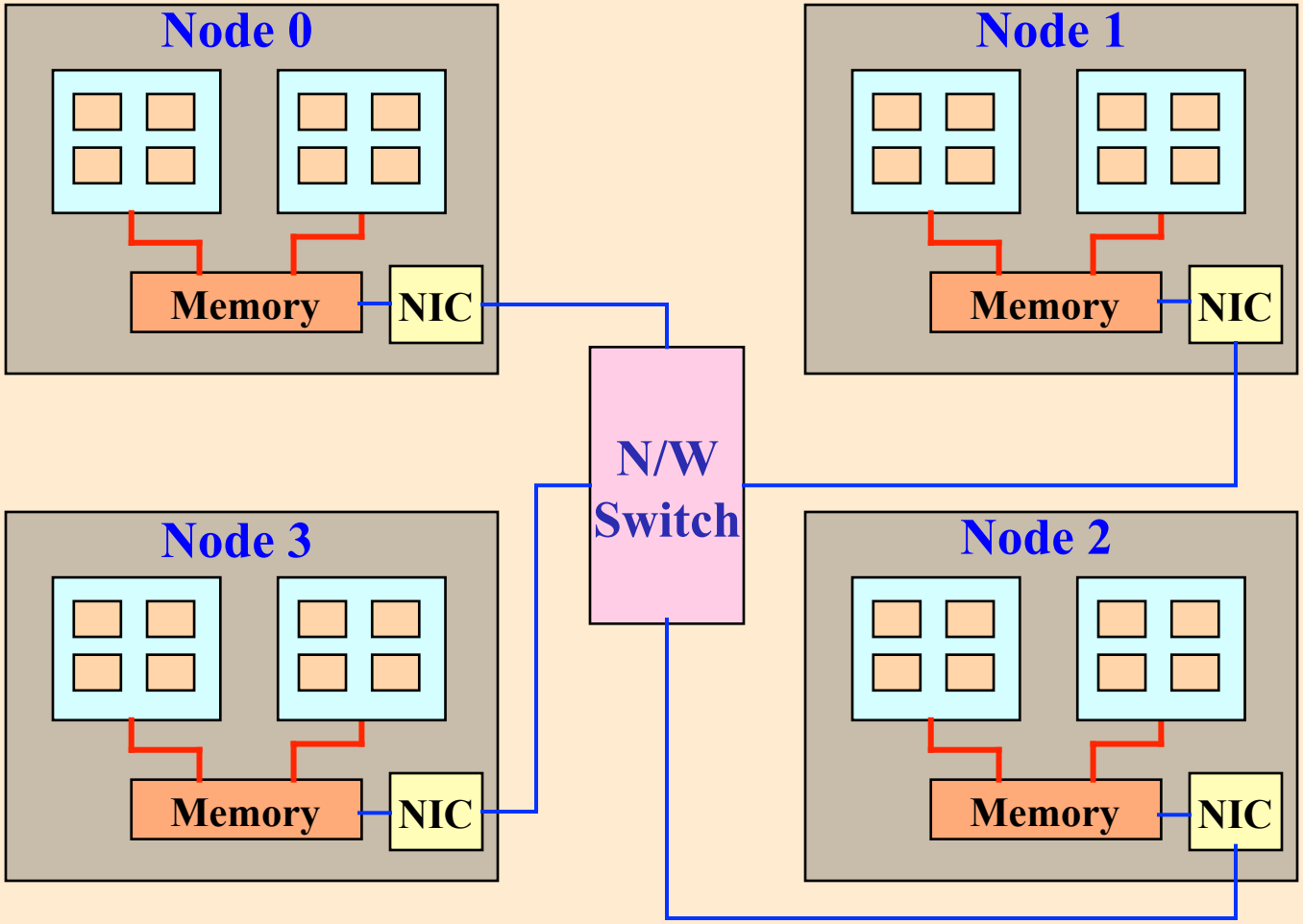



- Message Passing Architecture
  - Memory is private to each node
  - Processes communicate by messages

# NUMA Architecture



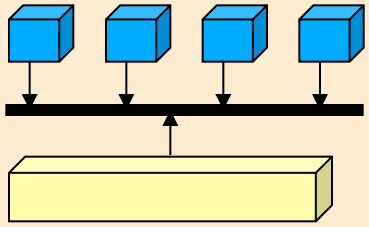
# Distributed Memory Architecture



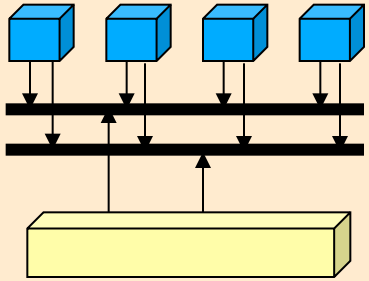
# Interconnection Network


- Processors and Memory modules connected to each other through Interconnect Network
- Indirect interconnects: nodes are connected to interconnection medium, not directly to each other
  - Shared bus, multiple bus, crossbar, MIN
- Direct interconnects: nodes are connected directly to each other
  - Topology: linear, ring, star, mesh, torus, hypercube
  - Routing techniques: how the route taken by the message from source to destination is decided

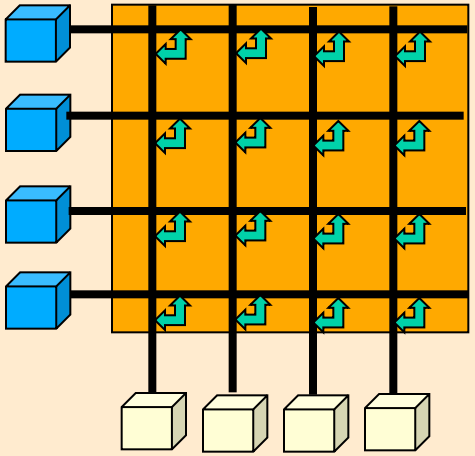
# Indirect Network Topology



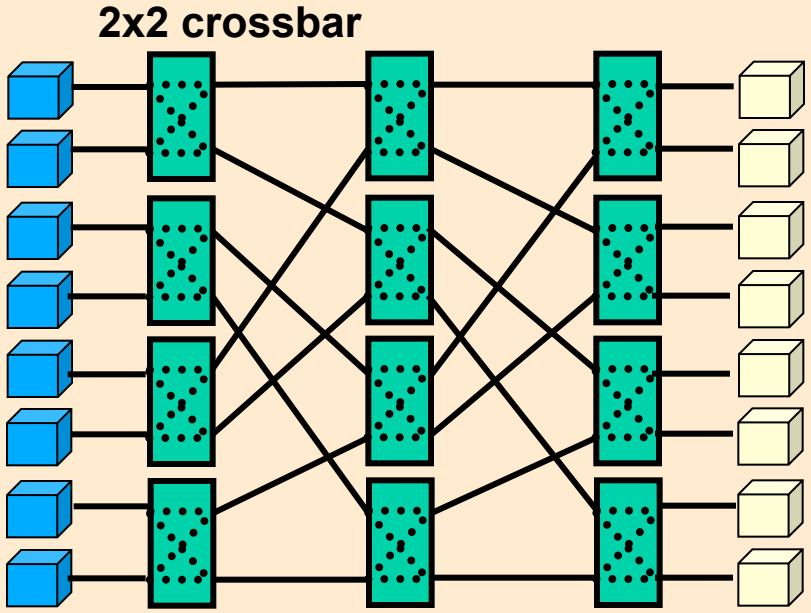
Shared bus



Multiple bus

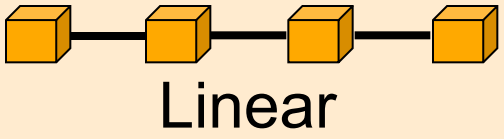


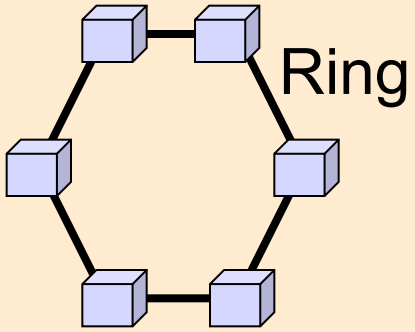
Crossbar switch

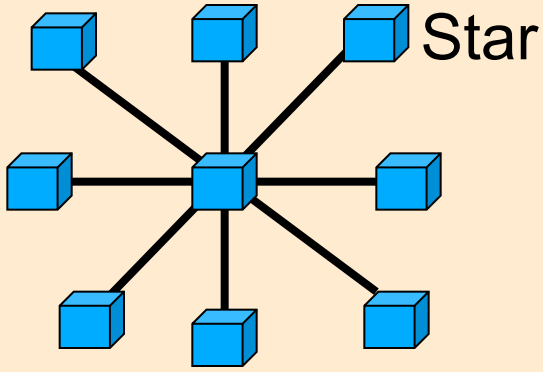


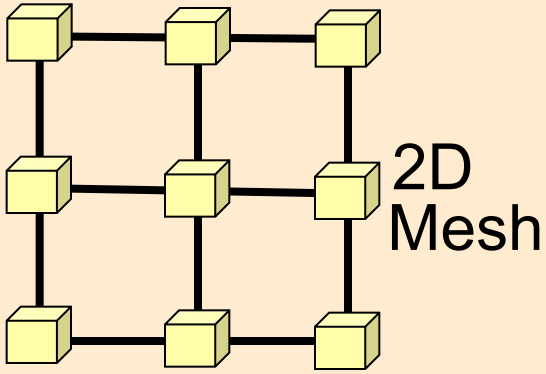

Multistage Interconnection Network

# Direct Interconnect Topology

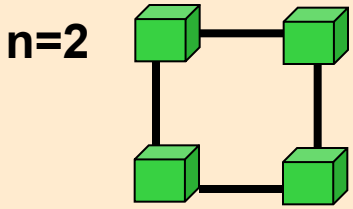


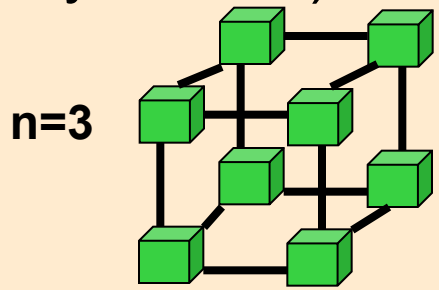



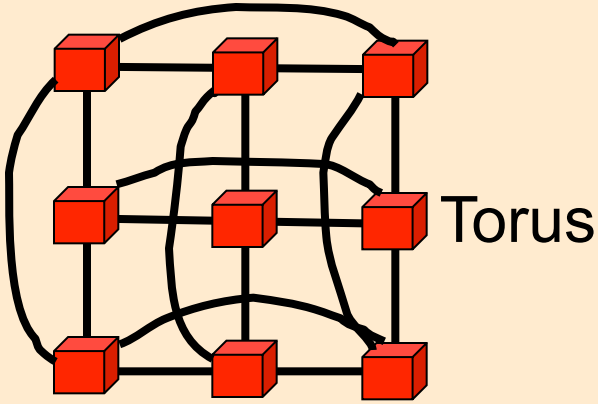




Hypercube (binary n-cube)



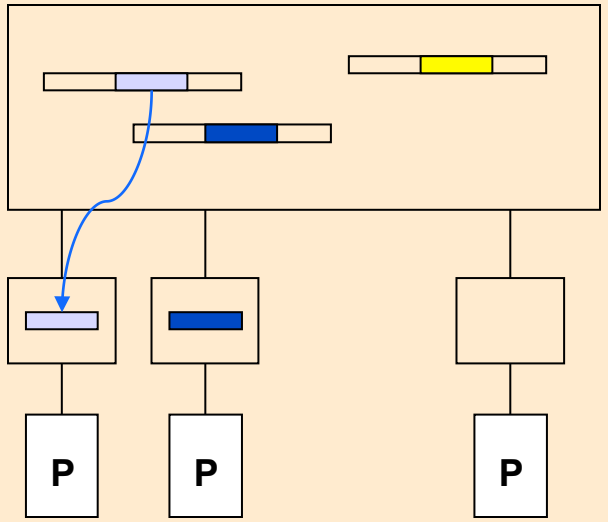




Atria-CC-

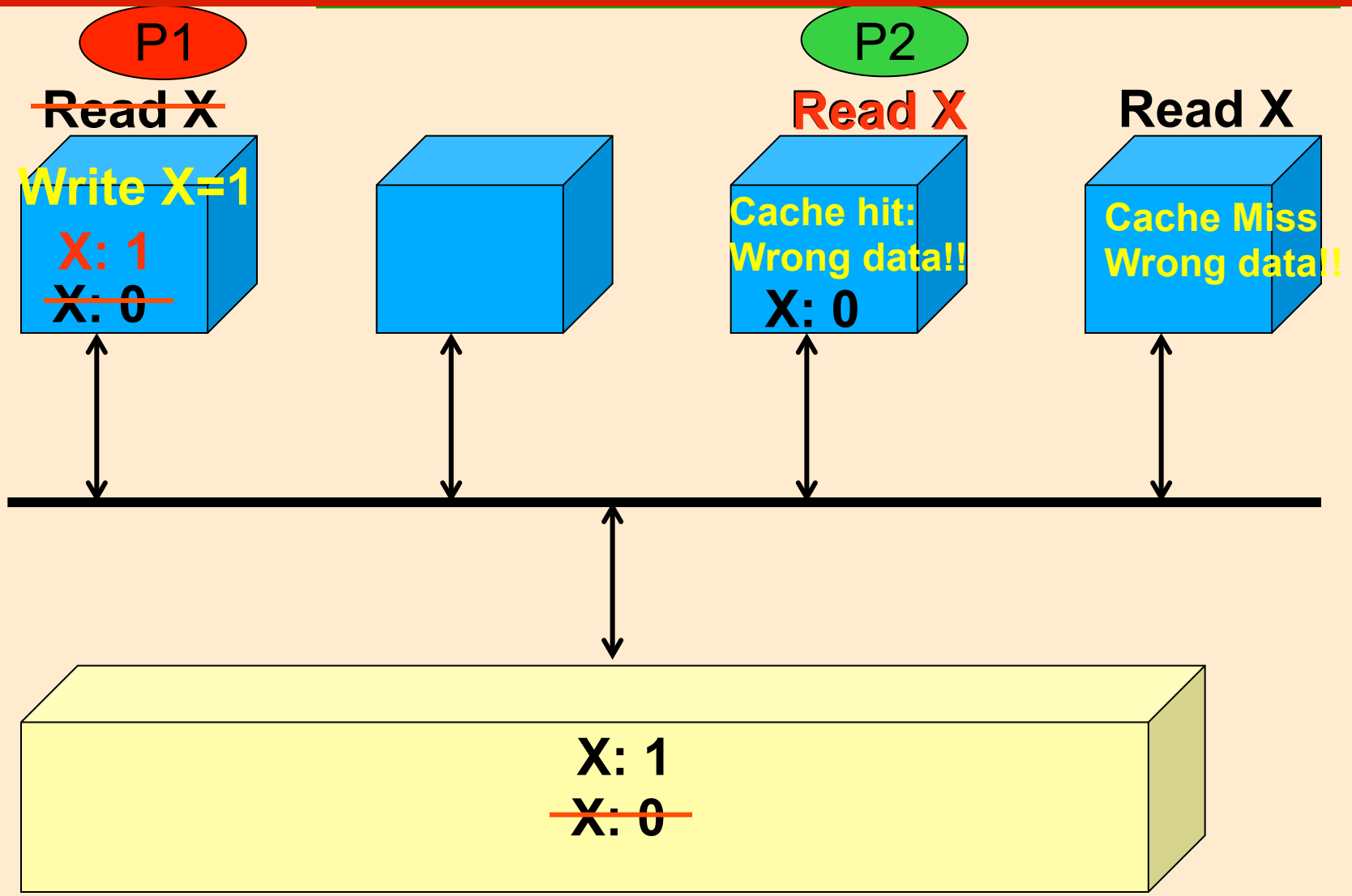
# Caches in Shared Memory

- Reduce average latency
  - automatic replication closer to processor
- What happens when store & load are executed on different processors?

  ⇒ Cache Coherence Problem



# Cache Coherence Problem



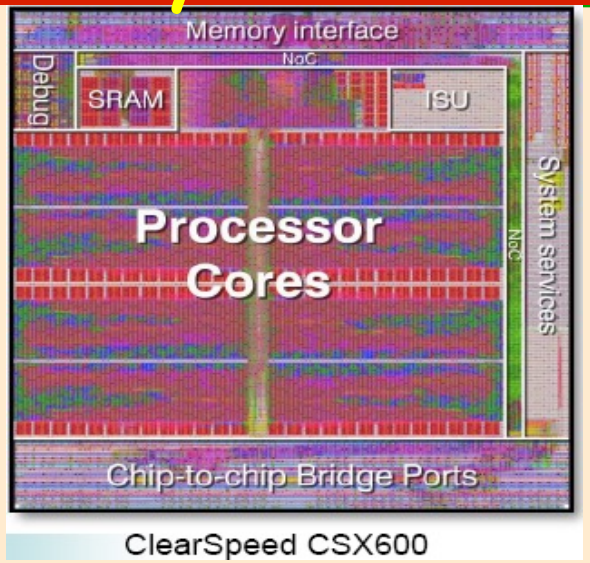
# Cache Coherence Solutions


- Snoopy Protocol: shared bus interconnect where all cache controllers monitor all bus activity
  - Cache controllers to take corrective action based on traffic in the interconnect network
  - Corrective action: update or invalidate a cache block
- Directory Based Protocols: Cache controllers maintain info. of shared copies of cache block
  - Send invalidation/update message to copies

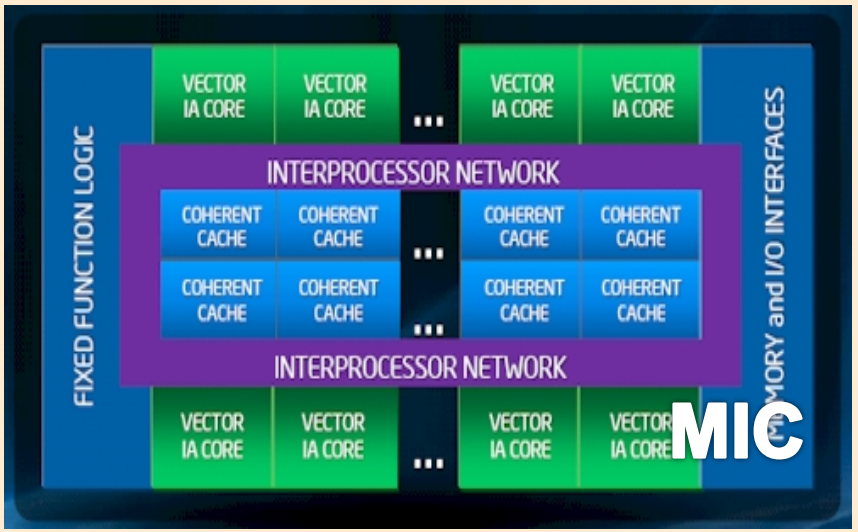
# Overview

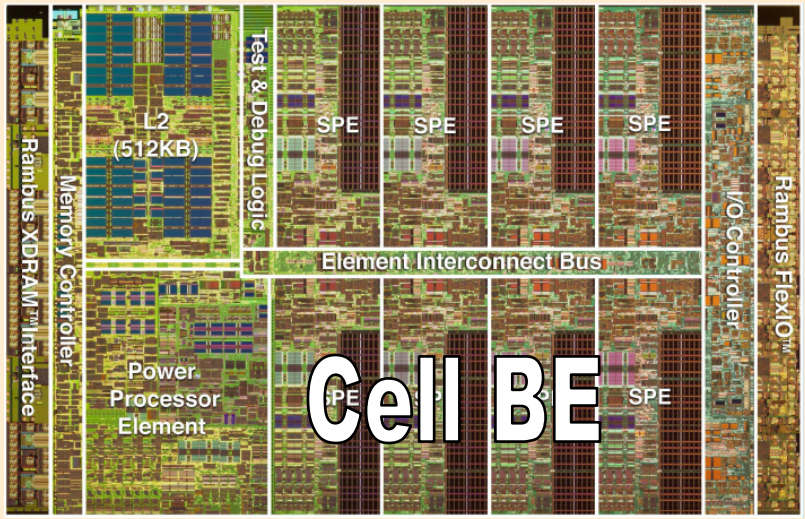

- Introduction
- Pipelining, Instruction Level Parallelism
- Multicore Architectures
- Multiprocessor Architecture
  - Shared Address Space
  - Distributed Address Space
- **Accelerators – GPUs**
- Supercomputer Systems

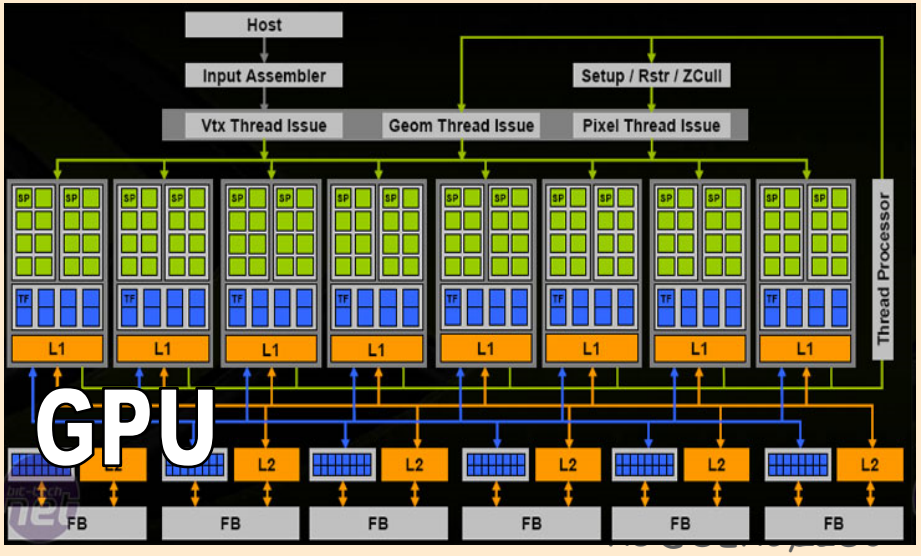
# Accelerators and Manycore Architectures



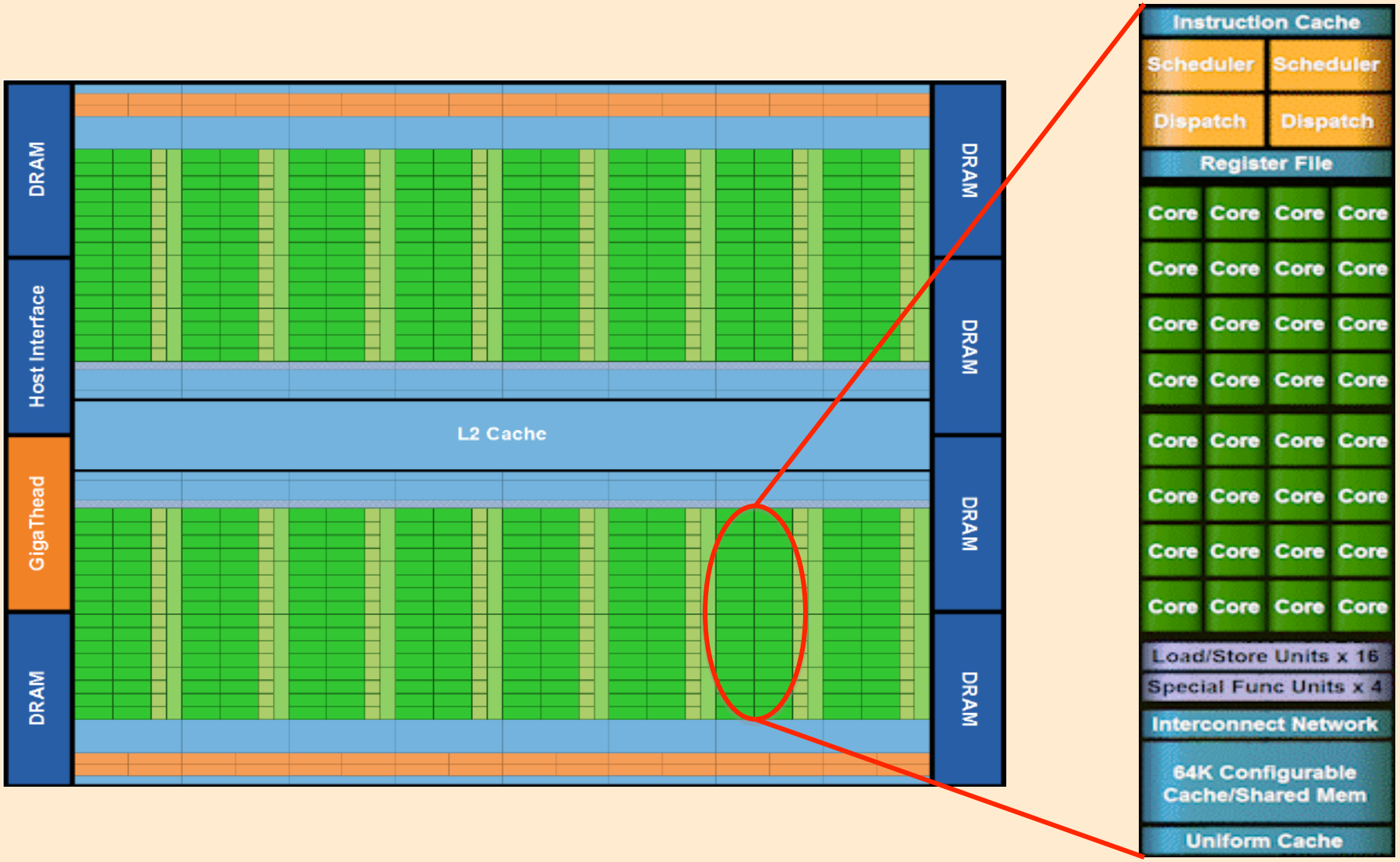
ClearSpeed CSX600



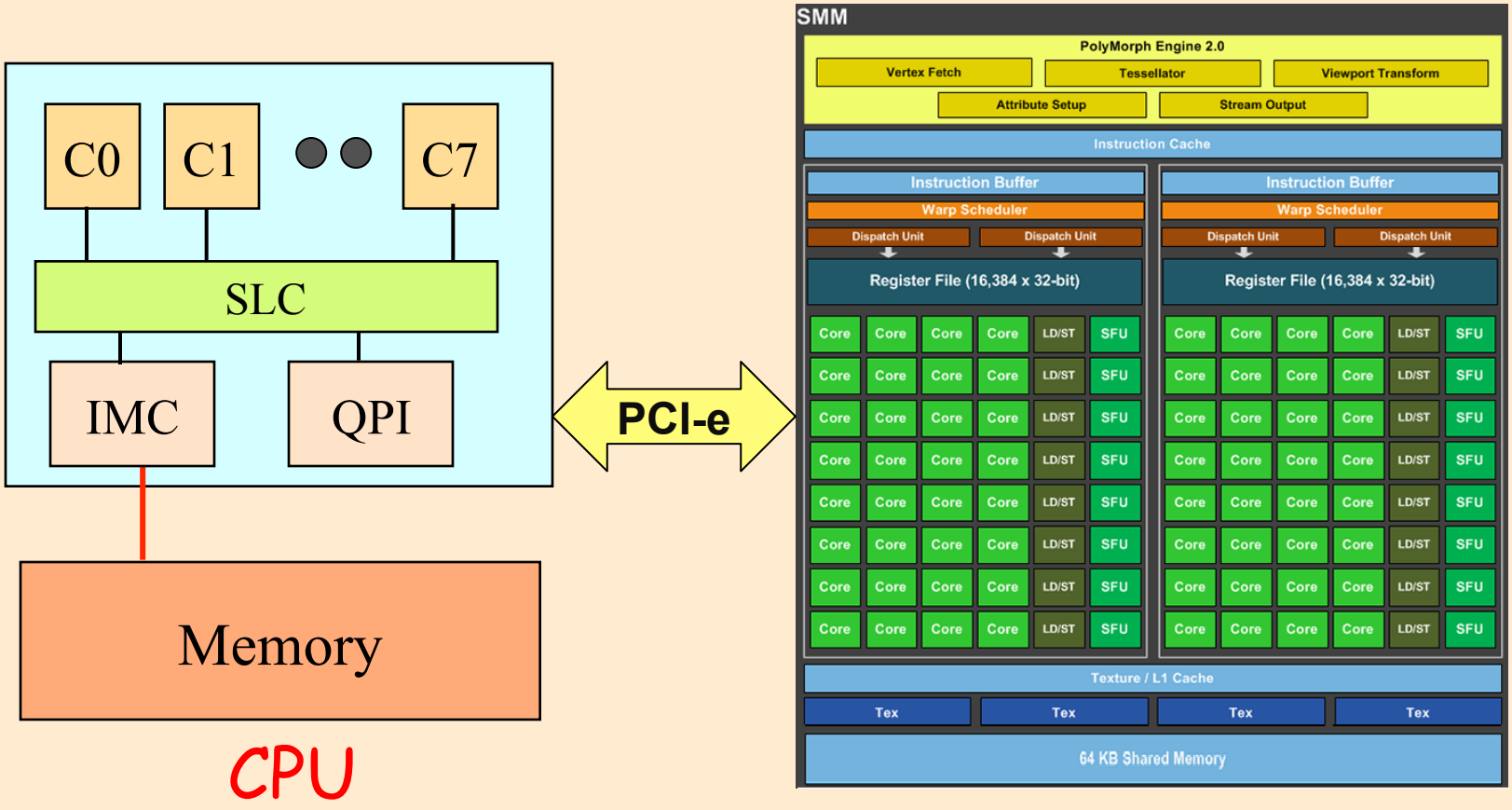
# Accelerator - Fermi S2050



# Combining CPU and GPU Arch.

- 8 CPU cores @ 3 GHz
- 0.38 TFLOPS

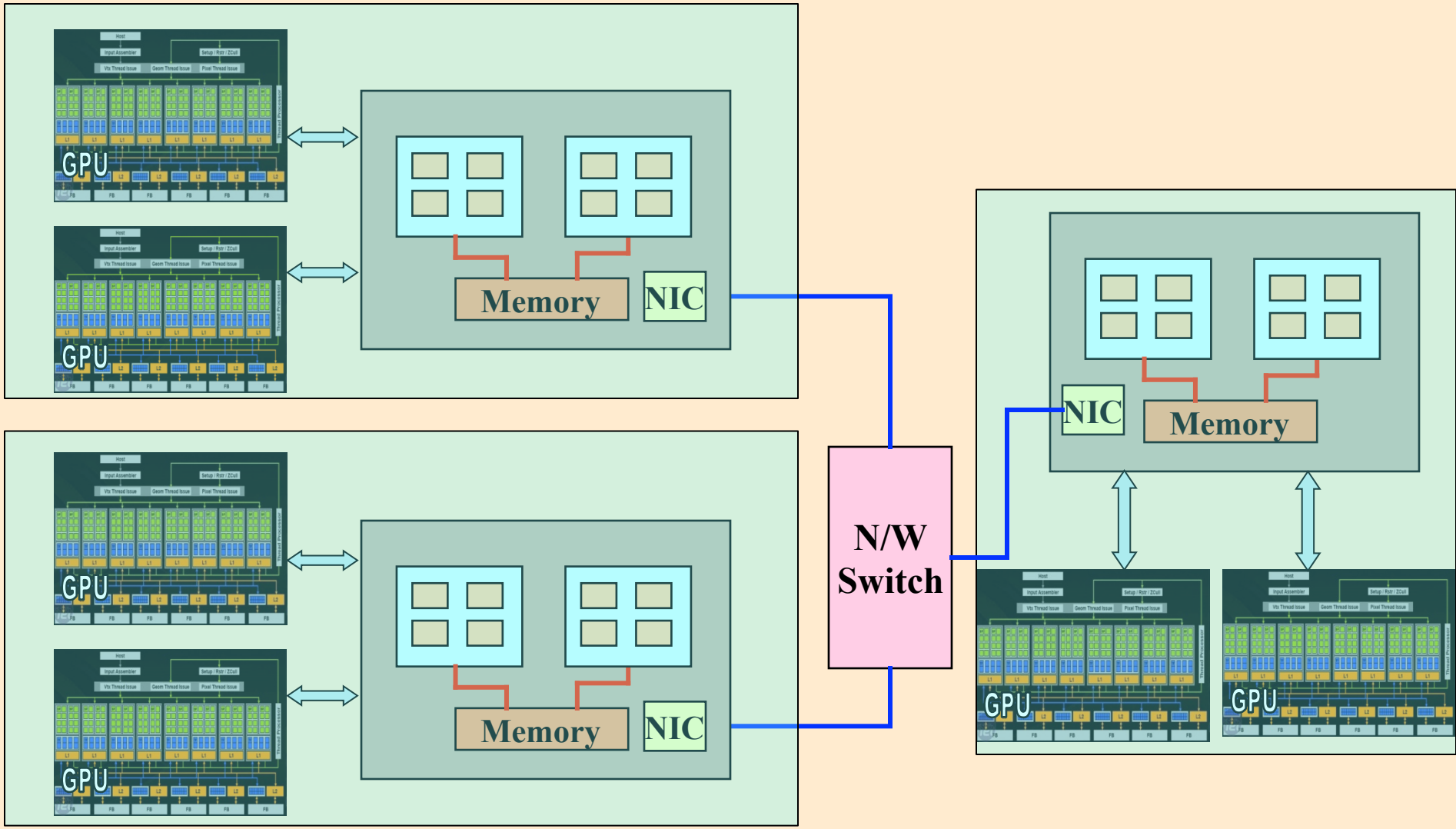
- 2880 CUDA cores @ 1.67 GHz
- 1.5 TFLOPS (DP)




CPU

GPU

# Heterogeneous Clusters with GPUs



# Overview

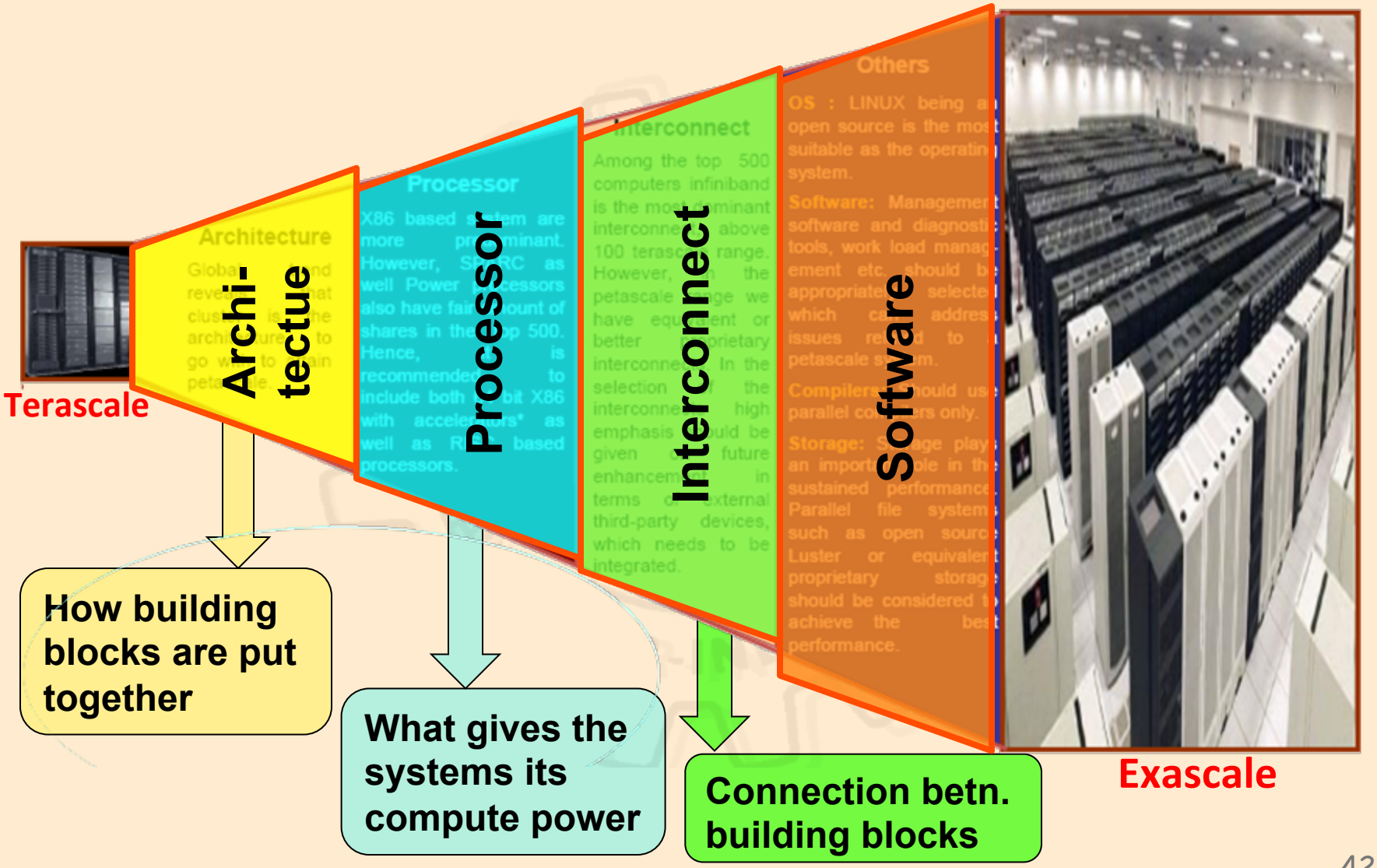

- Introduction
- Pipelining, Instruction Level Parallelism
- Multicore Architectures
- Multiprocessor Architecture
  - Shared Address Space
  - Distributed Address Space
- Accelerators – GPUs
- **Supercomputer Systems**

# What is a Supercomputer?


- A hardware and software system that provides close to the maximum performance than can currently be achieved.
- What was a supercomputer a few (5) years ago, is probably an order of magnitude slower system compared to today's supercomputer system!

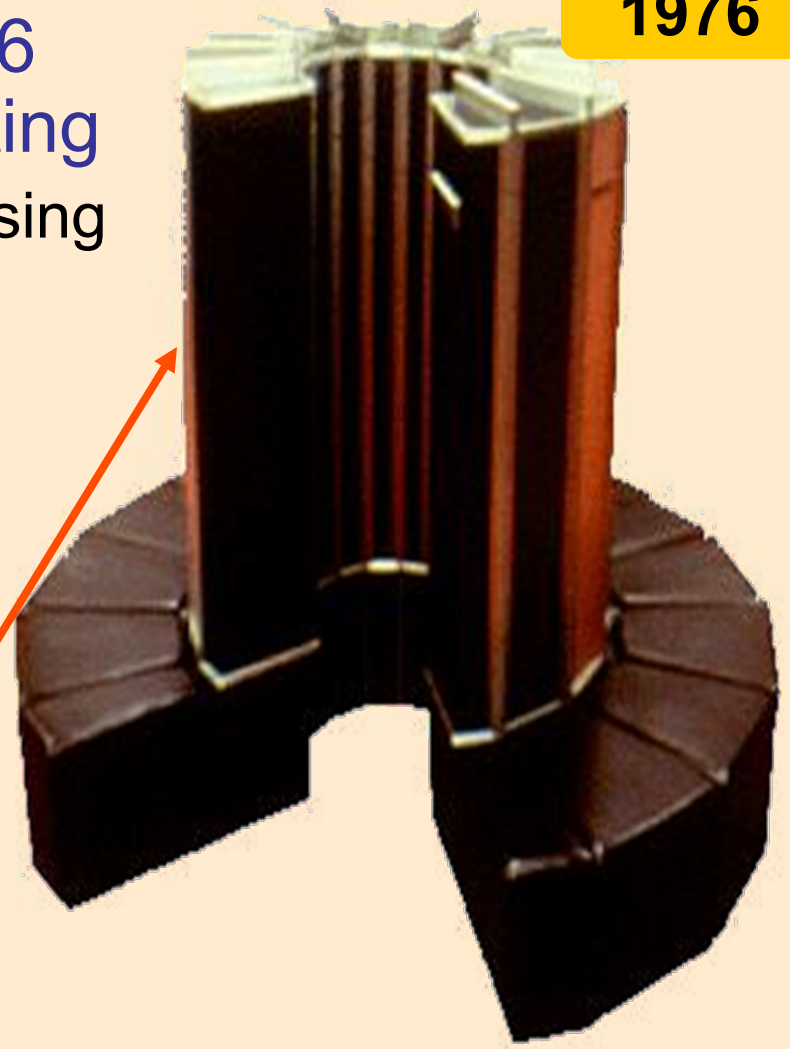
Therefore, we use the term "high performance computing" also to refer to Supercomputing!

# Components of a Supercomputer



# Era of Supercomputing

- Introduction of Cray 1 in 1976 ushered era of Supercomputing
  - –Shared memory, vector processing
  - –Good software environment
  - –A few 100 MFLOPS peak
  - –Cost about $5 million




# Performance of Supercomputer


- What are the top 10 or top 500 computers?
  - www.top500.org
  - Updated every 6 months
  - Measured using Rmax of Linpack (solving Ax = b )
- What is the trend?

| Year | Performance (GFLOPS) | |
|---|---|---|
| | #1 | |
| 1993 | 59.7 | 0.422 |
| 2018 | 143,500,000 | 874,800 |

2,403,685 x Impr.!!

~2,397,824 processors!

# The TOP 500 (Nov. 2018)


| Rank | Site | Manufacturer | Computer | Country | Cores | Rmax [Pflops] | Power [MW] |
|---|---|---|---|---|---|---|---|
| 1 | Oak Ridge National Labs, DOE/SC/ORNL | IBM | Summit: IBMPower9 22c, Nvidia V100,Mellanox EDR | USA | 2,397,824 | 143.50 | 9.78 |
| 2 | DOE/NSA/LLNL | IBM | Sierra: IBMPower9 22c, Nvidia V100,Mellanox EDR | USA | 1,572,480 | 94.64 | 7.43 |
| 3 | National SuperComputer Center in Wux | NRCPC | Sunway TaihuLight Sunway SW26010, 260C, 1.45 GHz | China | 10,649,600 | 93.01 | 15.37 |
| 4 | National SuperComputer Center in Tianjin | NUDT | Tianhe-2, NUDT TH MPP, Xeon E5 2691 and Xeon Phi 31S1 | China | 4,981,760 | 61.44 | 18.48 |
| 5 | Swiss National Supercomputing | Cray | PizDaint Cray XC-50, Xeon E5-2690, 12C (2.6GHz) + NVIDIA Tesla P100 | USA | 387,872 | 21.23 | 2.30 |
| 6 | DOE/NNSA/LANL/SNL | Cray | Cori, Cray XC-40, Intel Xeon E52689, 16c, Aries | USA | 979,072 | 20.16 | 7.58 |
| 7 | AIST, Japan | Fujitsu | Intel Xeon 6148, 20c, Tesla V100 SXM2, Infiniband EDR | Japan | 391,680 | 19.88 | 1.65 |
| 8 | SuperMUC-NG | Lenovo | Xeon 8174, 24c, Omnipath | Germany | 305,856 | 19.47 | |

Top 500 List : www.top500.org

# TOP SUPERCOMPUTERS INDIA

| Rank | Site | System | Cores/Processor Sockets/Nodes | Rmax (TFlops) | Rpeak (TFlops) |
|---|---|---|---|---|---|
| 1 | Indian Institute of Tropical Meteorology(IITM), Pune | Cray XC-40 class system with 3315 CPU-only (Intel Xeon Broadwell E5-2695 v4 CPU ) nodes with Cray Linux environment as OS, and connected by Cray Aries interconnect.<br>OEM: Cray Inc., Bidder: Cray Supercomputers India Pvt. Ltd. | 119232/ /3312 | 3763.9 | 4006.19 |
| 2 | National Centre for Medium Range Weather Forecasting (NCMRWF), Noida | Cray XC-40 class system with 2322 CPU-only (Intel Xeon Broadwell E5-2695 v4 CPU ) nodes with Cray Linux environment as OS, and connected by Cray Aries interconnect<br>OEM: Cray Inc., Bidder: Cray Supercomputers India Pvt. Ltd. | 83592//2322 | 2570.4 | 2808.7 |
| 3 | Supercomputer Education and Research Centre (SERC), Indian Institute of Science (IISc), Bangalore | Cray XC-40 Cluster (1468 Intel Xeon E5-2680 v3 @ 2.5 GHz dual twelve-core processor CPU-only nodes, 48 [Intel Xeon E5-2695v2 @ 2.4 Ghz single twelve-core processor+Intel Xeon Phi 5120D] Xeon-phi nodes, 44 [Intel Xeon E5-2695v2 @ 2.4 Ghz single twelve-core processor+NVIDIA K40 GPUs] GPU nodes) w/ Cray Aries Interconnect. HPL run on only 1296 CPU-only nodes.<br>OEM: Cray Inc., Bidder: Cray Supercomputers India Pvt. Ltd. | 36336C + 2880ICO + 126720G/ 3028C + 48ICO + 44G/ 1560C + 48ICO + 44G | 901.51 (CPU-only) | 1244.00 (CPU-only) |
| 4 | Indian Institute of Tropical Meteorology, Pune | IBM/Lenovo System X iDataPle DX360M4, Xeon E5-2670 8C 2.6 GHz, Infiniband FDR<br>OEM: IBM/Lenovo, Bidder: IBM India Pvt. Ltd. | 38016/ / | 719.2 | 790.7 |
| 5 | Indian Lattice Gauge Theory Initiative, Tata Institute of Fundamental Research (TIFR), Hyderabad | Cray XC-30 cluster (Intel Xeon E5-2680 v2 @ 2.8 GHz ten-core CPU and 2688-core NVIDIA Kepler K20x GPU nodes) w/Aries Interconnect<br>OEM: Cray Inc., Bidder: Cray Supercomputers India Pvt. Ltd. | 4760C + 1279488G/ 476C + 476G/ 476C + 476G | 558.7 | 730.00 |
| 6 | Indian Institute of Technology, Delhi | HP Proliant XL230a Gen9 and XL250a Gen9 based cluster (Intel Xeon E5-2680v3 @ 2.5 GHz dual twelve-core CPU and dual 2880-core NVIDIA Kepler K40 GPU nodes) w/Infiniband<br>OEM: HP, Bidder: HP | 10032C + 927360G/ 836C + 322G/ 418C + 161G | 524.40 | 861.74 |
| 7 | Center for Development of Advanced Computing (C-DAC), Pune | Param Yuva2 System (Intel Xeon E5-2670 (Sandy Bridge) @ 2.6 GHz dual octo-core CPU and Intel Xeon Phi 5110P dual 60-core co-processor nodes) w/Infiniband FDR<br>OEM: Intel, Bidder: Netweb Technologies | 3536C + 26520 ICO/ 442C + 442 ICO/ 221C + 221 ICO | 388.44 | 520.40 |
| 8 | Indian Institute of Technology ( IITB) Bombay | Cray XC-50 class system with 202 CPUs (Intel Xeon Skylake 6148 CPU @ 2.4GHz) regular node sand connected by Cray Aries interconnect.<br>OEM: OEM: Cray Inc., Bidder: Cray Supercomputers India Pvt. Ltd. | 8000/400/200 | 384.83 | 620.544 |
| 9 | CSIR Fourth Paradigm Institute (CSIR-4PI), Bangalore | HP Cluster Platform 3000 BL460c (Dual Intel Xeon 2.6 GHz eight core E5-2670 w/Infiniband FDR)<br>OEM: HP, Bidder: HCL Infosystems Ltd. | 17408/2176/1088 | 334.38 | 362.09 |
| 10 | National Centre For Medium Range Weather Forecasting, Noida | IBM/Lenovo System X iDataPlex DX360M4, Xeon E5-2670 8C 2.6 GHz, Infiniband FDR<br>OEM: IBM/Lenovo, Bidder: IBM India Pvt. Ltd. | 16832/ / | 318.4 | 350.1 |

# Supercomputing Systems & Applications are Challenging!