Chap.7 Multicores,Multiprocessors, and clusters

- 7.1 Introduction
- 7.2 The Difficulty of Creating Parallel Processing Programs
- 7.3 Shared Memory Multiprocessors
- 7.4 Clusters and Other Message-Passing
- 7.5 Multiprocessors
- 7.6 Hardware Multithreading

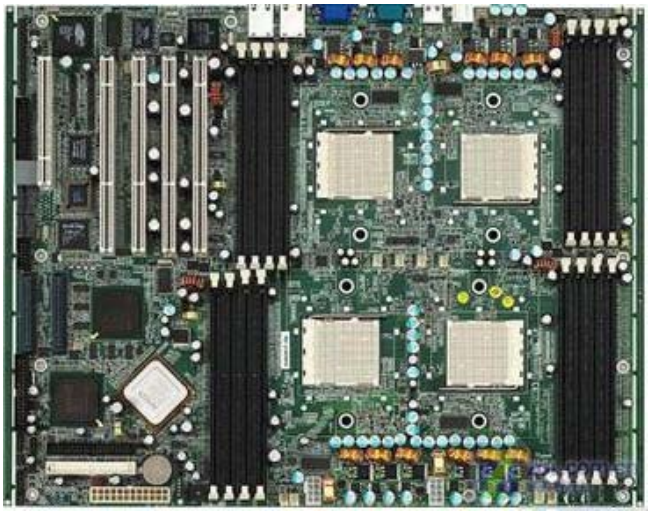
本次课程主要内容

- 认识 Multiprocessor 多处理器
- 认识 Multicore 多核（微处理器）
- 认识 Cluster 集群

认识 Multiprocessor 多处理器

- Multiprocessor 多处理器（计算机）：
  拥有至少2个处理器的计算机系统。

multiprocessor A computer system with at least two processors. This is in contrast to a **uniprocessor**, which has one.



认识 Multiprocessor 多处理器

共享主存的多处理器结构

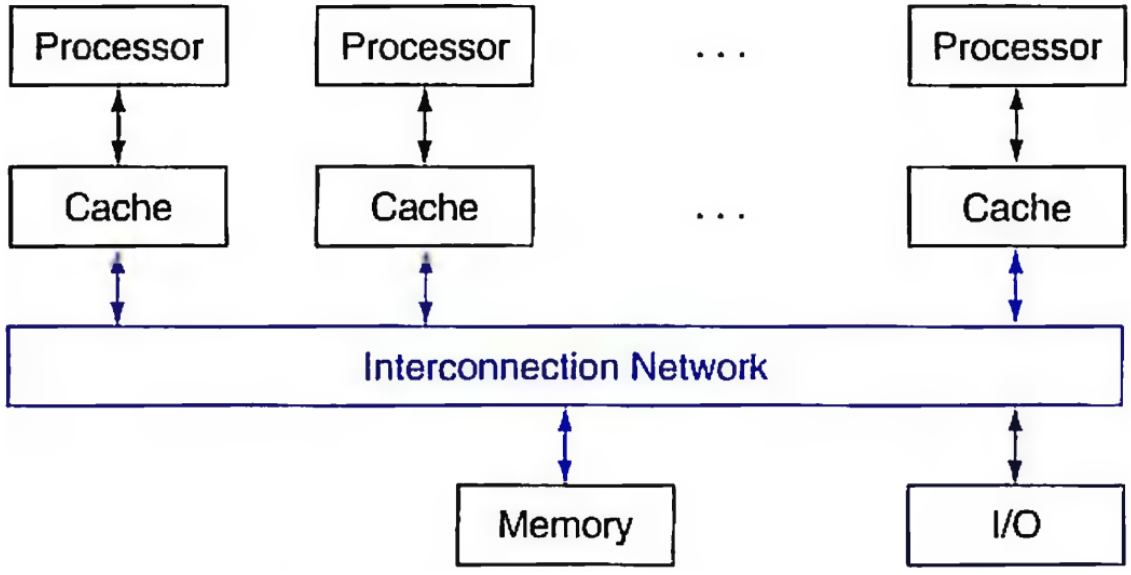


FIGURE 7.2 Classic organization of a shared memory multiprocessor.

认识 Multiprocessor 多处理器

共享主存的多处理器结构：涉及Cache的一致性

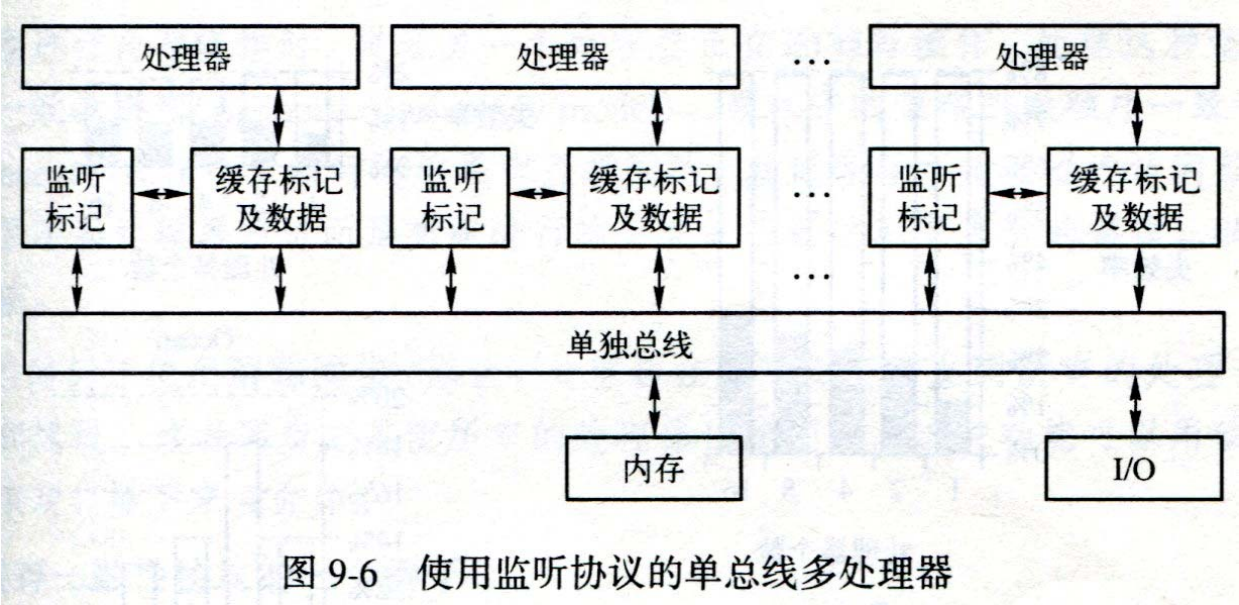


图 9-6　使用监听协议的单总线多处理器

认识 Multiprocessor 多处理器

共享主存的多处理器（SMP）

- Shared Memory Multiprocessor（SMP）
  共享主存多处理器：
  - 给所有处理器提供一个共享的物理地址空间。
  - 每个处理器可以在其自己的虚拟地址空间运行自己独立的程序。
  - 各处理器间可以利用主存中的共享变量进行通讯。

shared memory multiprocessor (SMP) A parallel processor with a single address space, implying implicit communication with loads and stores.

A **shared memory multiprocessor** (SMP) is one that offers the programmer a *single physical address space* across all processors, although a more accurate term would have been shared-*address* multiprocessor. Note that such systems can still run independent jobs in their own virtual address spaces, even if they all share a physical address space. Processors communicate through shared variables in memory, with all processors capable of accessing any memory location via loads and stores. Figure 7.2 shows the classic organization of an SMP.

认识 Multiprocessor 多处理器

通过网络连接的多处理器结构（分布式主存）

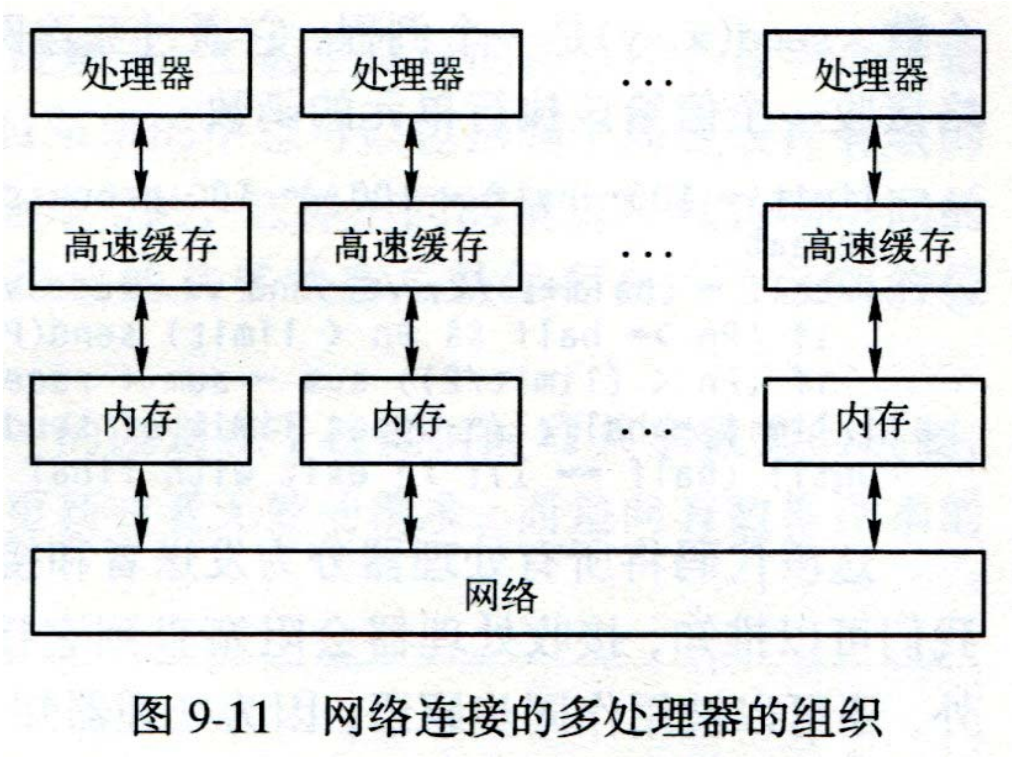


图 9-11　网络连接的多处理器的组织

认识 Multiprocessor 多处理器

各处理器拥有自己的私有地址空间

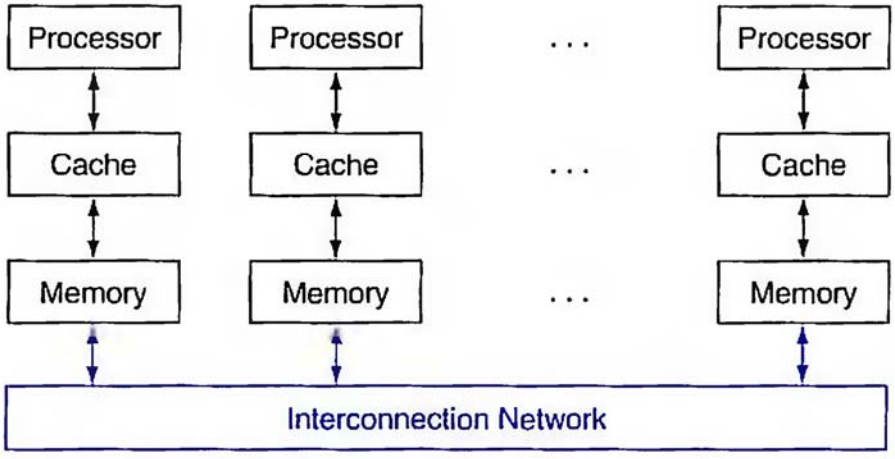


FIGURE 7.4　Classic organization of a multiprocessor with multiple private address spaces, traditionally called a message-passing multiprocessor. Note that unlike the SMP in Figure 7.2, the interconnection network is not between the caches and memory but is instead between processor-memory nodes.

认识 Multiprocessor 多处理器

多处理器可分为三类

- 一致性内存访问（Uniform Memory Access，UMA）：
  不论访问字在哪里，内存访问时间是相同的。
- 非一致性内存访问（NonUniform Memory Access，NUMA）：
  内存访问时间取决于地址。距离较近的访问时间可能会比较短（与连接网络有关）。
- Cache 一致的非一致内存访问（CC-NUMA）：
  NUMA加上Cache一致性确认机制。

uniform memory access (UMA) A multiprocessor in which accesses to main memory take about the same amount of time no matter which processor requests the access and no matter which word is asked.

nonuniform memory access (NUMA) A type of single address space multiprocessor in which some memory accesses are much faster than others depending on which processor asks for which word.

认识 Multiprocessor 多处理器

多处理器间的同步（Synchronization）

- 在没有全局统一地址空间的各处理器间，利用显式消息通讯机制进行处理器间的执行同步。
- 在共享单一地址空间主存方式下，可通过共享变量的锁（Lock）机制进行处理器间的同步。

synchronization The process of coordinating the behavior of two or more processes, which may be running on different processors.

lock A synchronization device that allows access to data to only one processor at a time.

认识 Multiprocessor 多处理器

硬件设计对同步变量“锁 / Lock”机制的支持

■ **atomic swap**（原子交换）操作：

处理器可以在一个总线操作周期内（或指令周期），读一个地址并写入，该动作一气呵成，不可被打断。可防止其它处理器中途修改该变量。

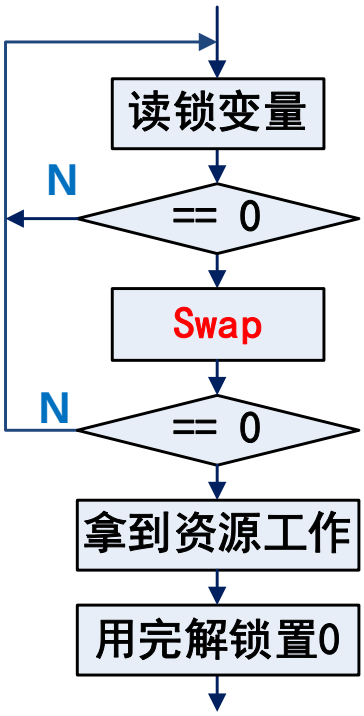
□ 该操作可用于对共享变量（资源）进行加锁。

（**lock variable**, **semaphore**）

硬件设计对同步变量“锁 / Lock”机制的支持

加锁（申请资源）过程：

① 读入锁变量；
② 若该变量值不是0，返回①；
③ 执行swap操作：先读该变量，也即返回的交换值，并给该变量置一非0的“编号”，例如使用者编码；
④ 若交换值（读出值）不等于0，即未抢到资源，返回①；
⑤ 说明已拿到资源，那就工作吧……；
……
⑥ 解锁，即将该变量置0。其它处理器（进程）就可以抢了。



硬件设计对同步变量“锁 / Lock”机制的支持

硬件指令对“锁 / Lock”的支持

- ARM指令：

  swap指令： Atomic swap register and memory

  指令格式：

  swp R1,[R2,#20]

  功能含义：

  R1 = Memory[R2+20]，Memory[R2+20]=R1。

- MIPS指令：由一对2条指令（LL 指令 + SC 指令）完成。

LL	Load Linked Word	Rt = Mem[Rs+offset] LL = 1 LLAdr = Rs + offset
SC	Store Conditional Word	if LL =1 mem[Rxoffs] = Rt Rt = LL

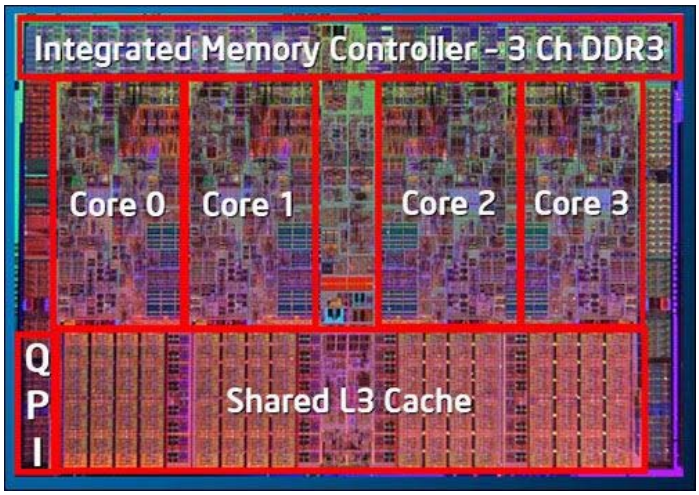
认识 Multicore 多核微处理器

- Multicore 多核（微处理器）：

  在单片集成微处理器上，包含有多个处理器（“核”）。

multicore microprocessor
A microprocessor containing multiple processors (“cores”) in a single integrated circuit.

Intel Core i7



例：Intel Core i7的高速缓存层次结构

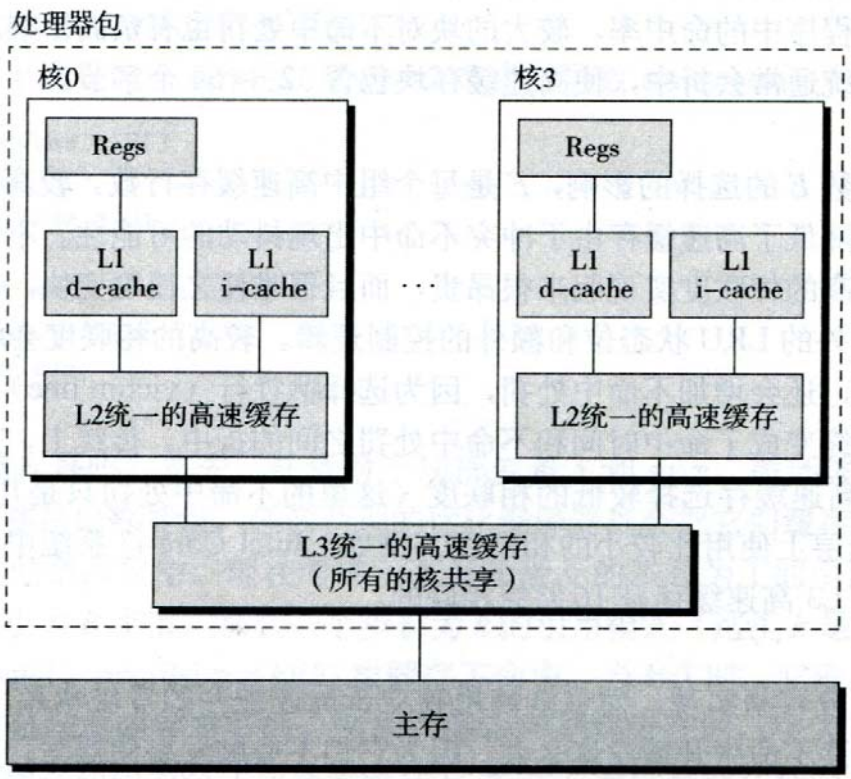
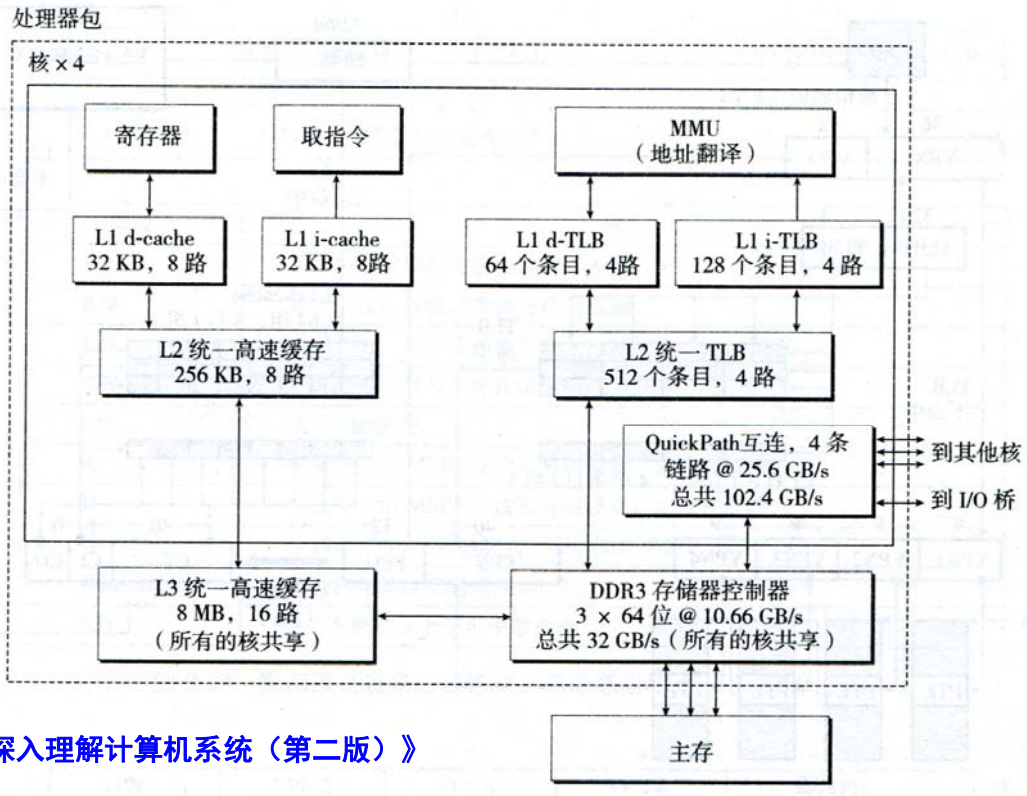


图 6-40　Intel Core i7 的高速缓存层次结构　来自《深入理解计算机系统》

例：Intel Core i7的存储器系统



来自《深入理解计算机系统（第二版）》

图 9-21　Core i7 存储器系统

认识 Cluster 集群

■ Cluster 集群（计算机）：
利用局域网互联一组计算机构成的一个大型多处理器系统。

cluster A set of computers connected over a local area network (LAN) that functions as a single large multiprocessor.

认识 Cluster 集群

计算机集群

■计算机集群简称集群是一种计算机系统， 它通过一组松散集成的计算机软件和/或硬件连接起来高度紧密地协作完成计算工作。

■在某种意义上，它们可以被看作是一台计算机。

■集群系统中的单个计算机通常称为节点，通常通过局域网连接，但也有其它的可能连接方式。

■集群计算机通常用来改进单个计算机的计算速度和/或可靠性。

■一般情况下集群计算机比单个计算机，比如工作站或超级计算机的性能价格比要高得多。

■ 参考第三版教材9.8节Google PC集群系统实例。

认识 Hardware Multithreading 硬件多线程技术

片内硬件多线程技术

- **Hardware Multithreading allows multiple threads to share the functional units of a single processor in an overlapping fashion.**
- **允许多个线程以交替执行的方式共享处理器内的功能部件。**

hardware multithreading
Increasing utilization of a processor by switching to another thread when one thread is stalled.

认识 Hardware Multithreading 硬件多线程技术

Hardware Multithreading 的3种方式

- **fine-grained multithreading，细粒度多线程**
  在每条指令后切换线程，多个线程教程处理。
- **coarse-grained multithreading，粗粒度多线程**
  在遇到进程阻塞（等待开销较大）时，切换线程。
- **Simultaneous MultiThreading，SMT，同时多线程**
  利用具有多发射、动态调度能力的处理器实现线程级并行，同时实现指令级并行。

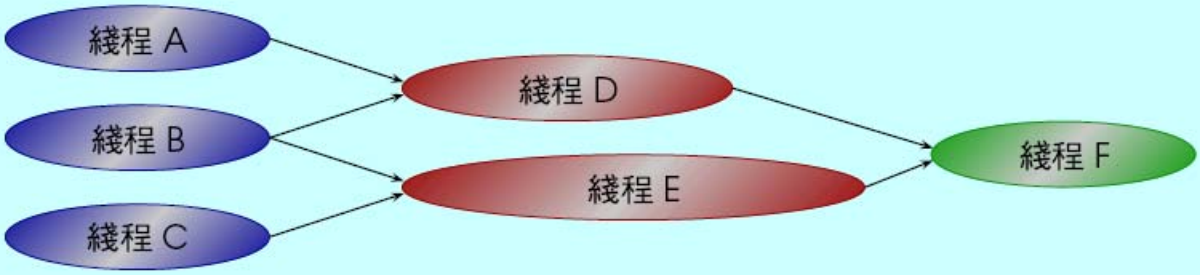
fine-grained multithreading
A version of hardware multithreading that suggests switching between threads after every instruction.

coarse-grained multithreading
A version of hardware multithreading that suggests switching between threads only after significant events, such as a cache miss.

simultaneous multithreading (SMT)
A version of multithreading that lowers the cost of multithreading by utilizing the resources needed for multiple issue, dynamically schedule microarchitecture.

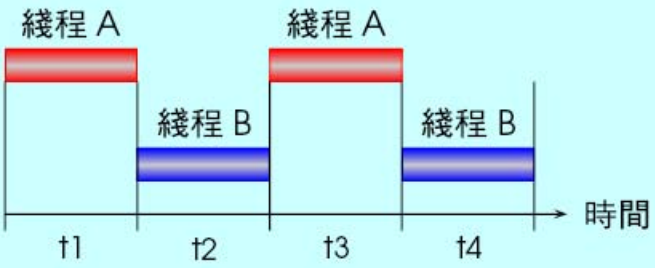
线程

- 一個綫程是指一段程序在執行過程中不依賴其它綫程產生的數據

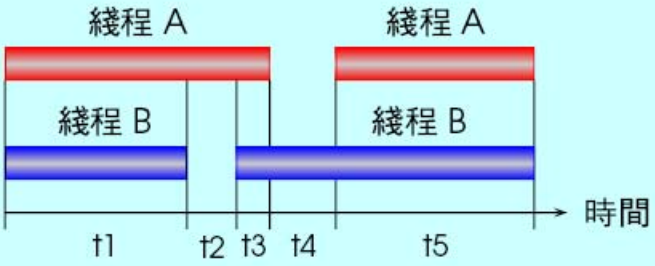


- 綫程 D 和綫程 E 從理論上講可以并行執行
- 綫程 F 祇有等綫程 D 和綫程 E 均執行完才能開始執行

多线程CPU



(a) 單綫程 CPU 順序執行多綫程

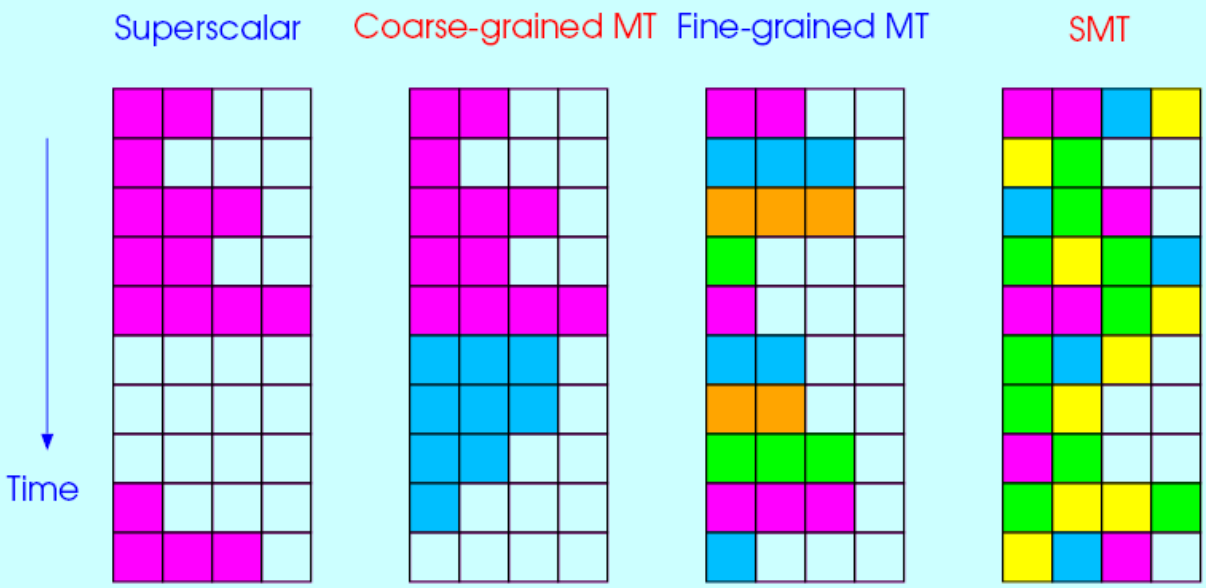


(b) 多綫程 CPU 并行執行多綫程

Multithreading

- Multiple threads share the functional units of a single processor in an overlapping fashion
  1. Fine-grained multithreading
     - Switches between threads on each instruction, causing the execution of multiple threads to be interleaved
  2. Coarse-grained multithreading
     - Switches threads only on costly stalls, such as cache misses
  3. Simultaneous multithreading (SMT)
     - Uses the resources of a multiple-issue, dynamically scheduled processor to exploit TLP (thread level parallelism), at the same time it exploits ILP (instruction level parallelism)

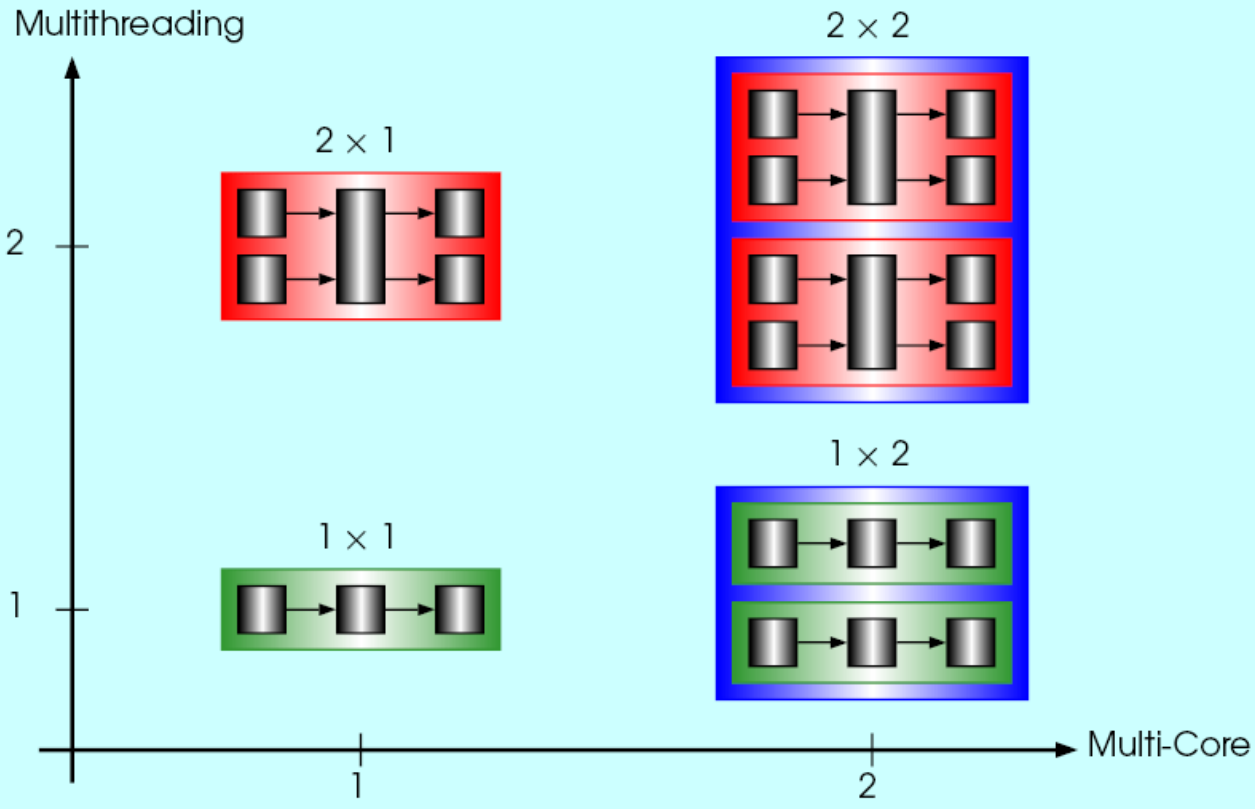
Superscalar and Multithreading



Multithreading CPU and Multi-Core CPU

- Multithreading CPU
  - Multiple virtual processors
  - Functional units are shared
- Multi-core CPU
  - Multiple cores on a single chip
    - = Chip Multiprocessors (CMP)
  - Functional units are not shared
  - Multiple cores may share L2 Caches
  - A core may be a multithreading CPU
    - ⇒ Multithreaded multicore CPU
  - Dual-core: 2 cores on a single chip

Multithreaded Multi-Core CPUs



Hardware Multithreading 的3种方式

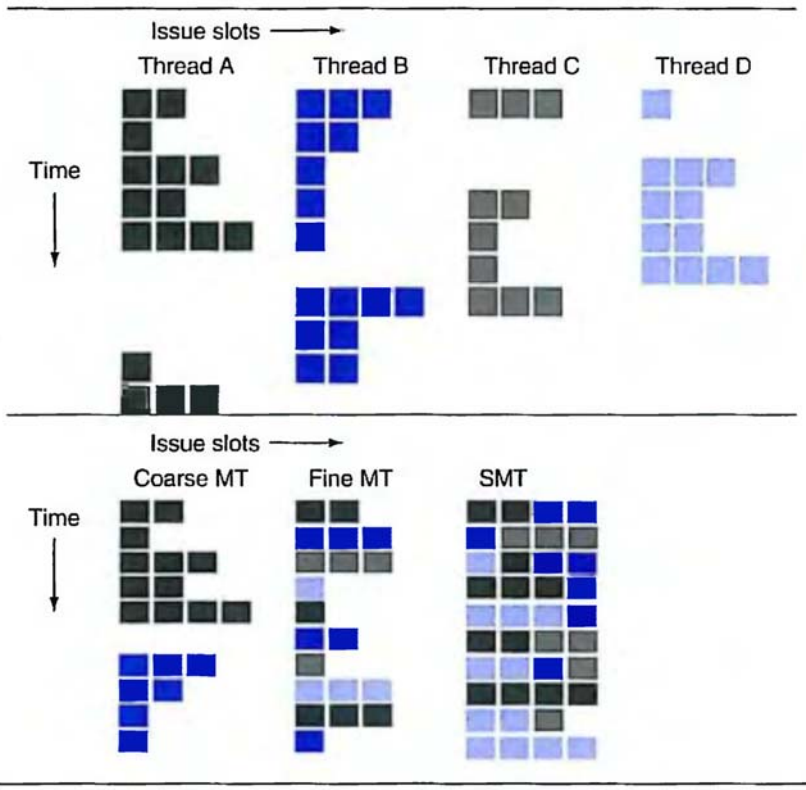
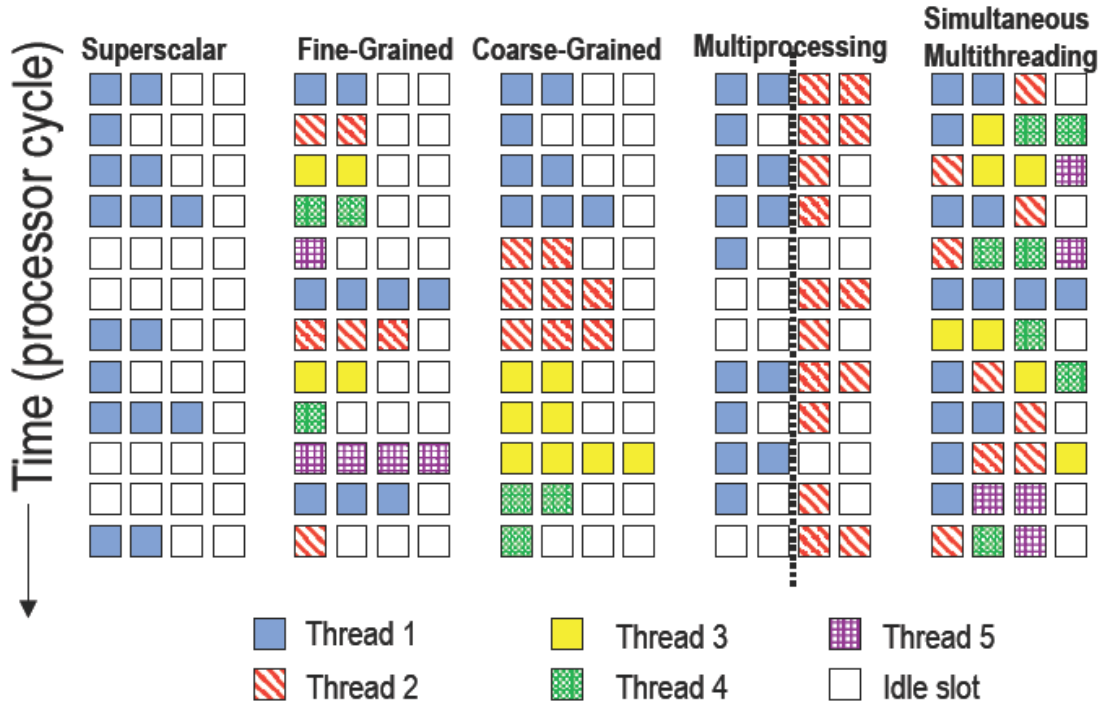


FIGURE 7.5 How four threads use the issue slots of a superscalar processor in different approaches. The four threads at the top show how each would execute running alone on a standard

Hardware Multithreading 方式比较

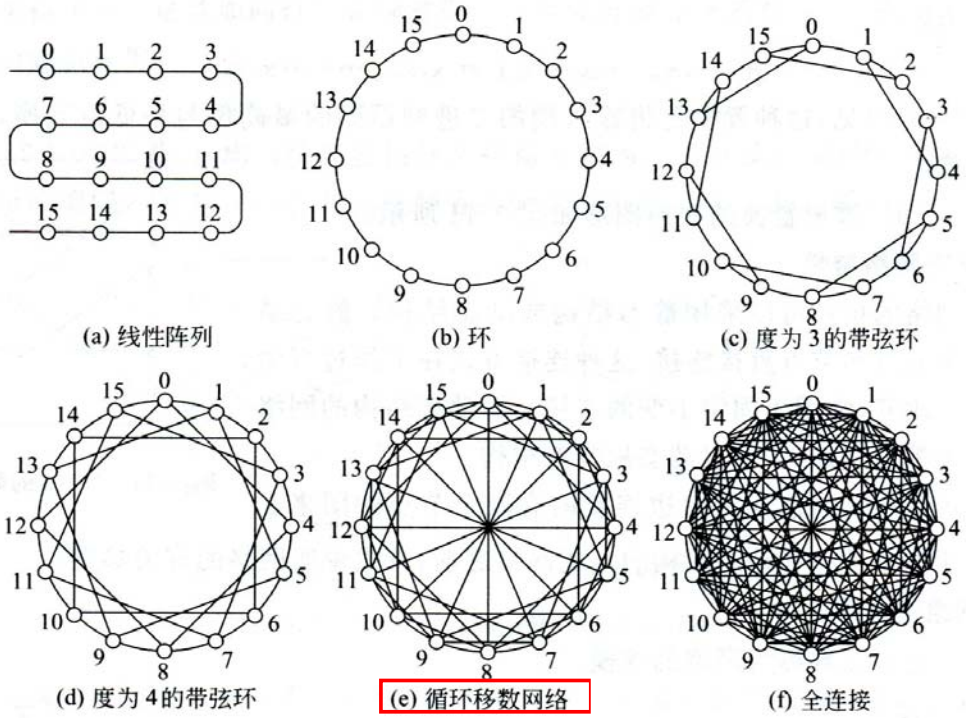
Summary: Multithreaded Categories



April 6, 2010 CS152, Spring 2010 来源：UC-Berkeley

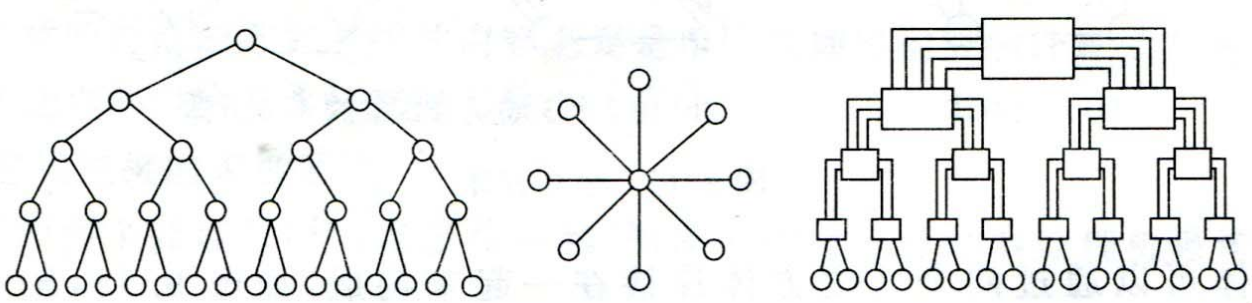
Introduction to Multiprocessor Network Topologies

多处理器互联网络拓扑结构 – 几种线性网络



Introduction to Multiprocessor Network Topologies

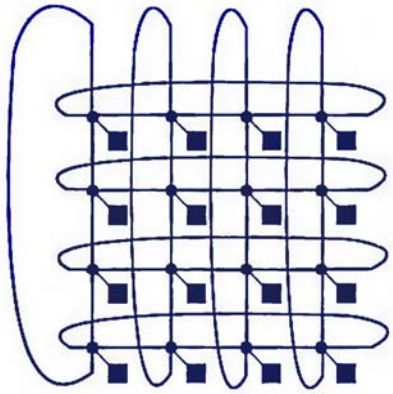
多处理器互联网络拓扑结构



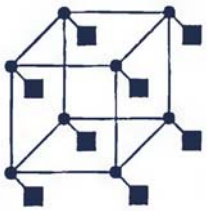
This topology is called a *ring*

Introduction to Multiprocessor Network Topologies

多处理器互联网络拓扑结构 – 2维环网、3-立方体



a. 2-D grid or mesh of 16 nodes

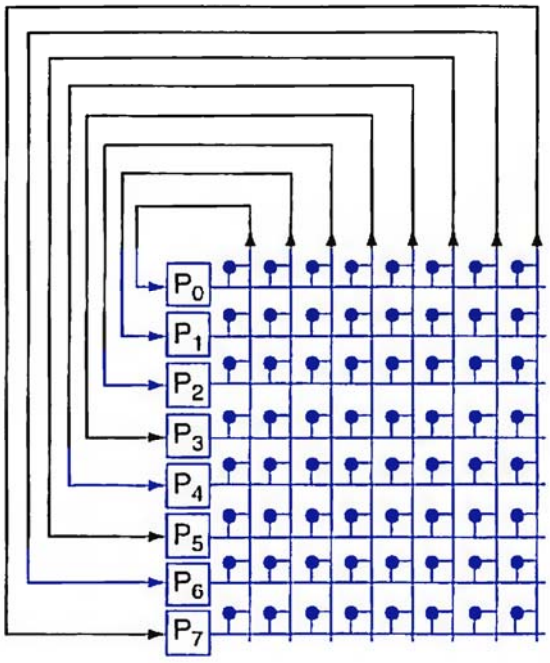


b. *n*-cube tree of 8 nodes ($8 = 2^3$ so $n = 3$)

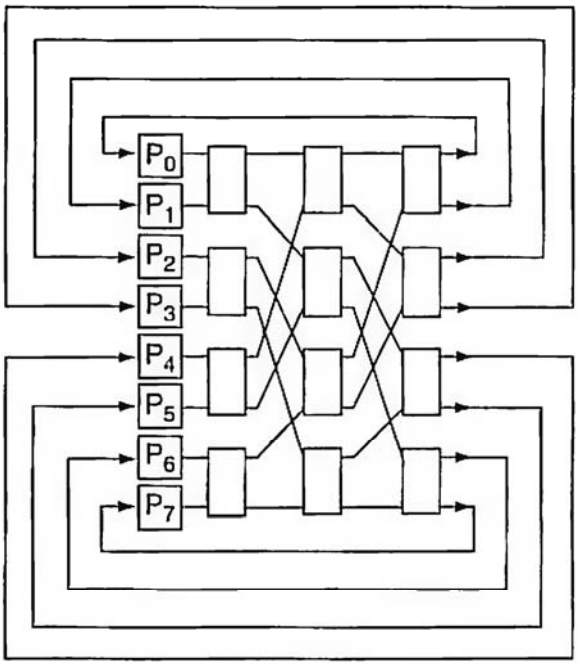
FIGURE 7.9 Network topologies that have appeared in commercial parallel processors. The colored circles represent switches and the black squares represent processor-memory nodes. Even though a switch has many links, generally only one goes to the processor. The Boolean *n*-cube topology is an *n*-dimensional interconnect with 2^n nodes, requiring *n* links per switch (plus one for the processor) and thus *n* nearest-neighbor nodes. Frequently, these basic topologies have been supplemented with extra arcs to improve performance and reliability.

Introduction to Multiprocessor Network Topologies

多处理器互联网络拓扑结构



a. Crossbar

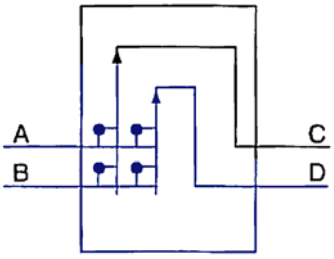


b. Omega network

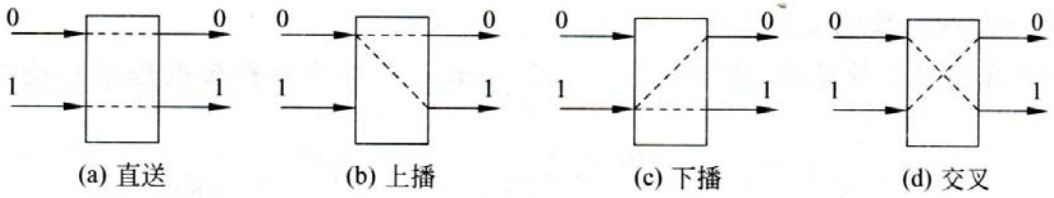
FIGURE 7.10 Popular multistage network topologies for eight nodes.

Introduction to Multiprocessor Network Topologies

多处理器互联网络拓扑结构



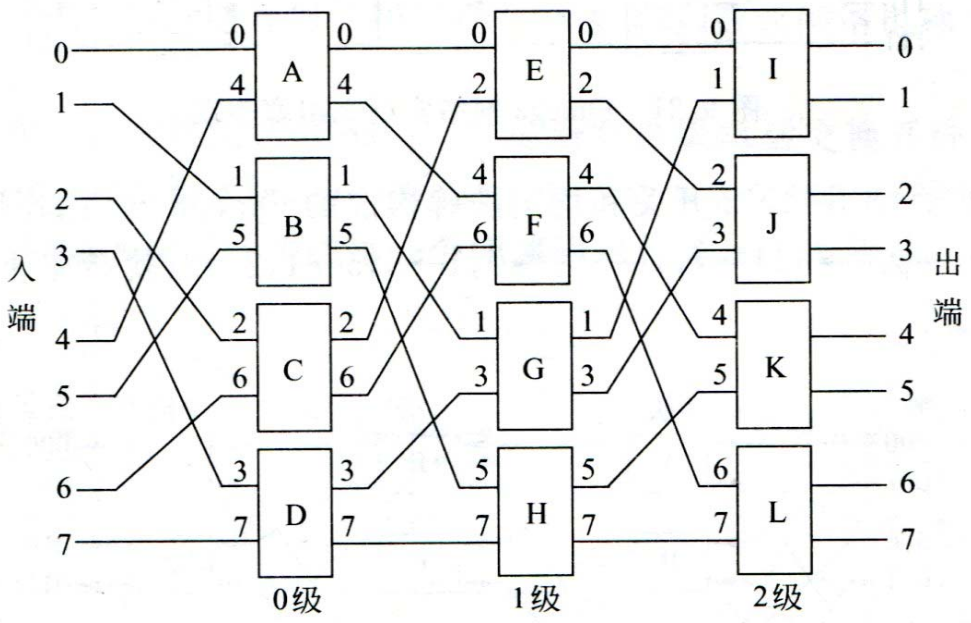
c. Omega network switch box



2 x 2开关模块的4种控制状态

Omega网络/多级均匀洗牌网络

多级均匀洗牌网络亦称Omega网络

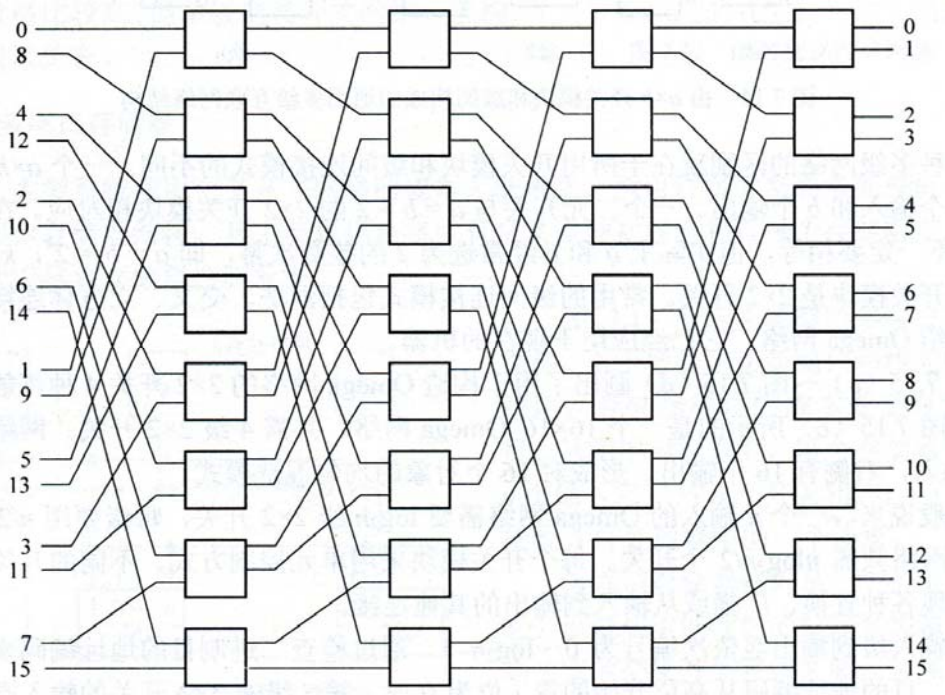


一个N=8用2x2开关和均匀洗牌构成的Omega网络

路径控制：目的地址编码从高位开始，依次控制沿途各级2x2开关。若该位编码为0，该级的输入端与开关上输出端连接，为1与下输出端连接。（同学自行测试）

Omega网络/多级均匀洗牌网络

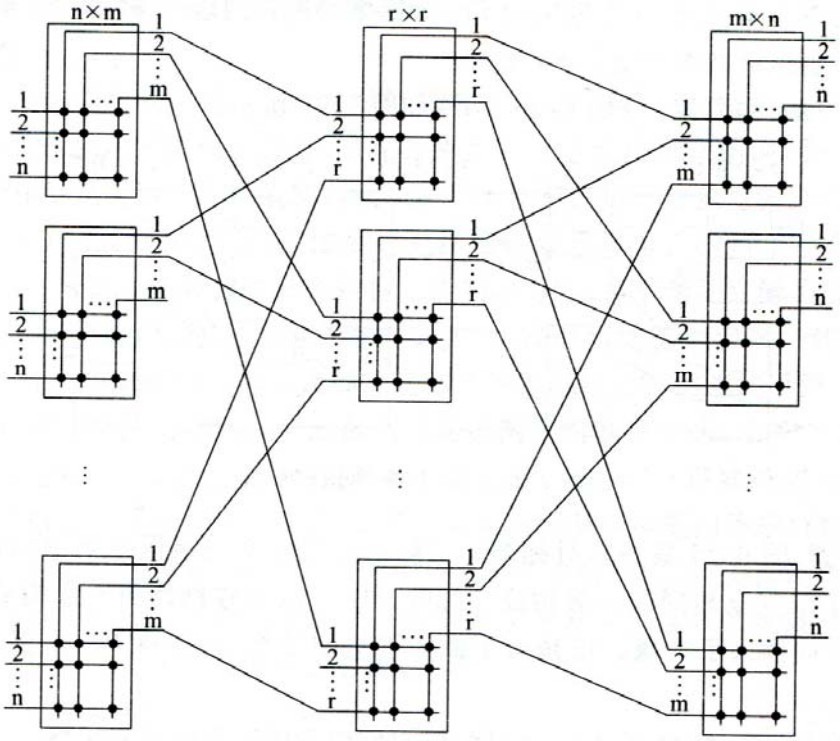
用2x2开关和均匀混洗构成的16x16 Omega 网络



路径控制：目的地址编码从高位开始，依次控制沿途各级2x2开关。若该位编码为0，该级的输入端与开关上输出端连接，为1与下输出端连接。（同学自行测试）

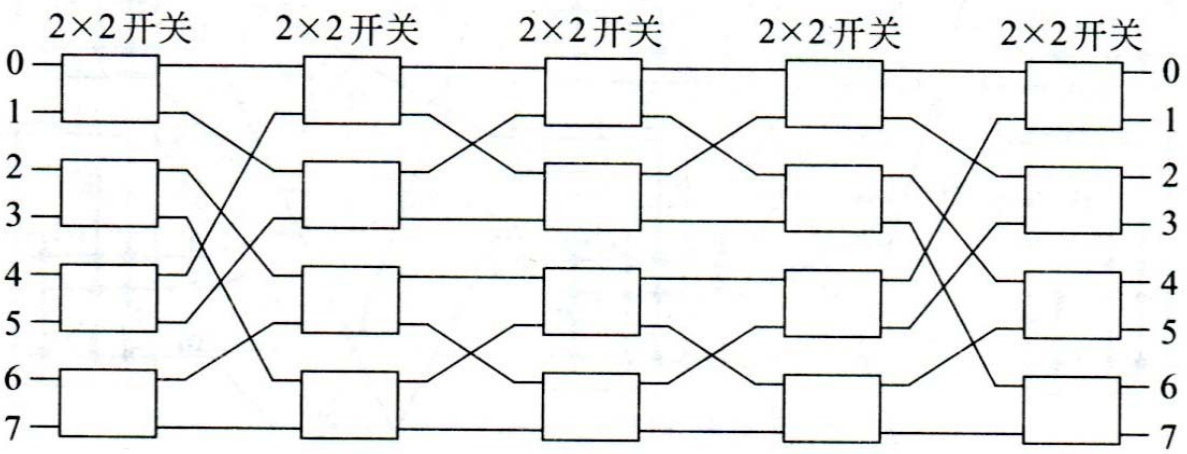
多处理器互联网络拓扑结构

Clos 网络结构



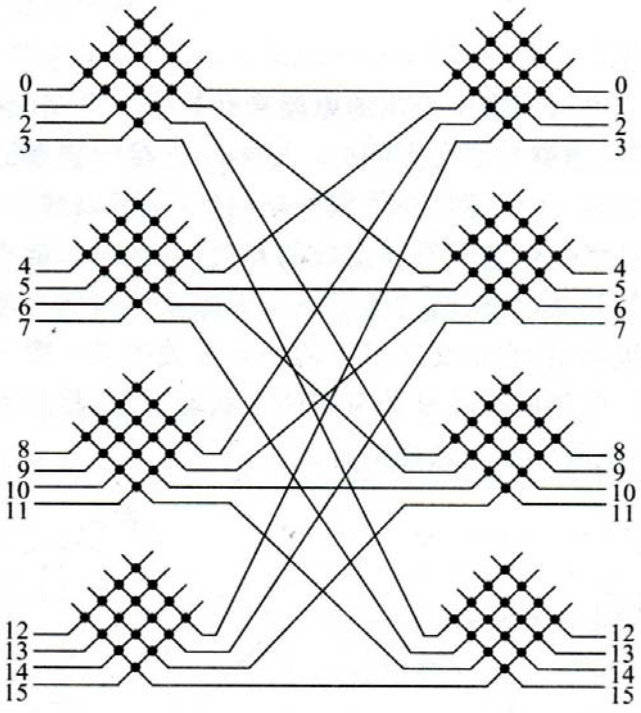
多处理器互联网络拓扑结构

N=8 的 Benes 网络（无阻塞）



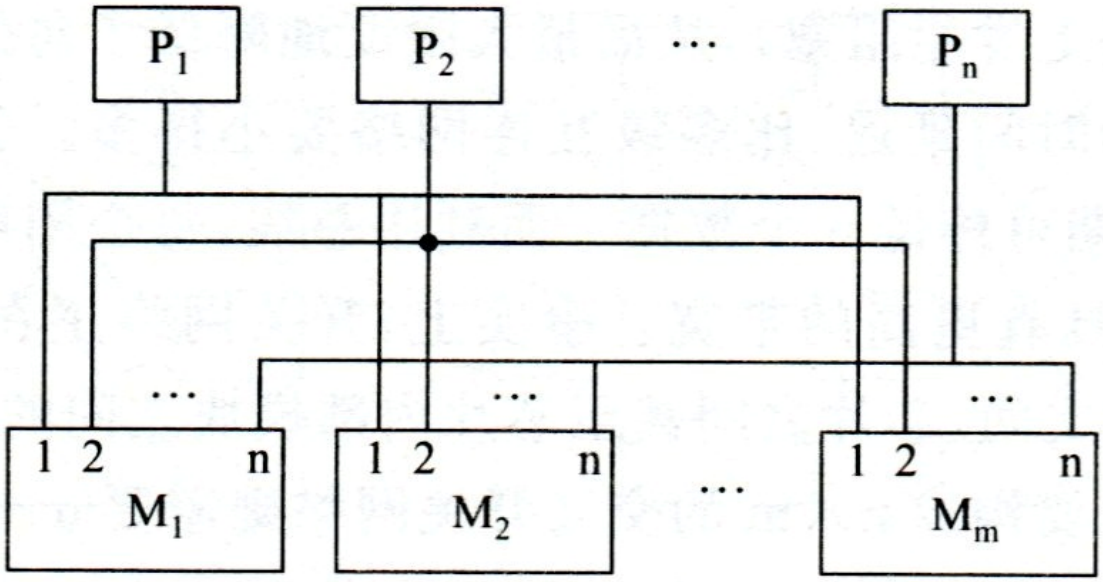
多处理器互联网络拓扑结构

用4x4交叉开关模块构成16x16两级交叉开关网络



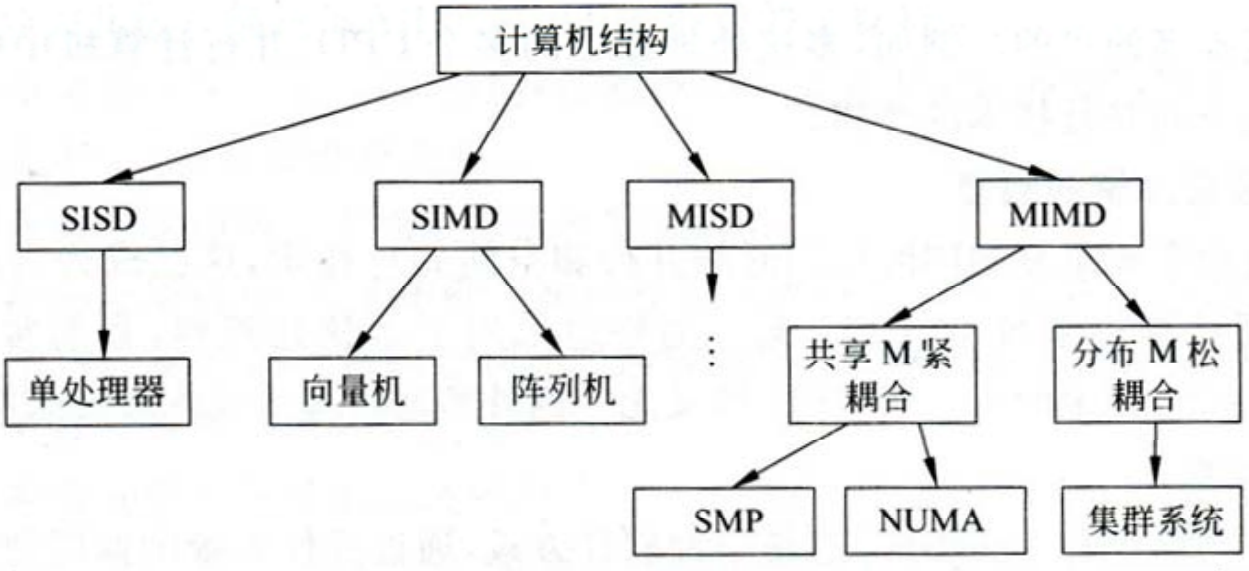
多处理器互联网络拓扑结构

利用多端口存储器构成多机系统



计算机系统结构

并行结构的分类（Flynn分类）



其它知识点

- 总线
- DMA
- 中断
- VLIW、超标量处理器

  （静态多发射、动态多发射（超标量）的概念及其实现原理（考核），教材4.10.2-3节）

- 网格计算
- 云计算

Acknowledgements

- These slides contain material developed and copyright by:
  - —Arvind (MIT)
  - —Krste Asanovic (MIT/UCB)
  - —John Kubiatowicz (UCB)
  - —David Patterson (UCB)
  - —Geng Wang (SJTU)
  - —Yanmin Zhu (SJTU)
  - —Li Yamin(Hosei Univ.)