计算机组成与体系结构

2023

7.1 多处理机概念

7.2 多处理机结构

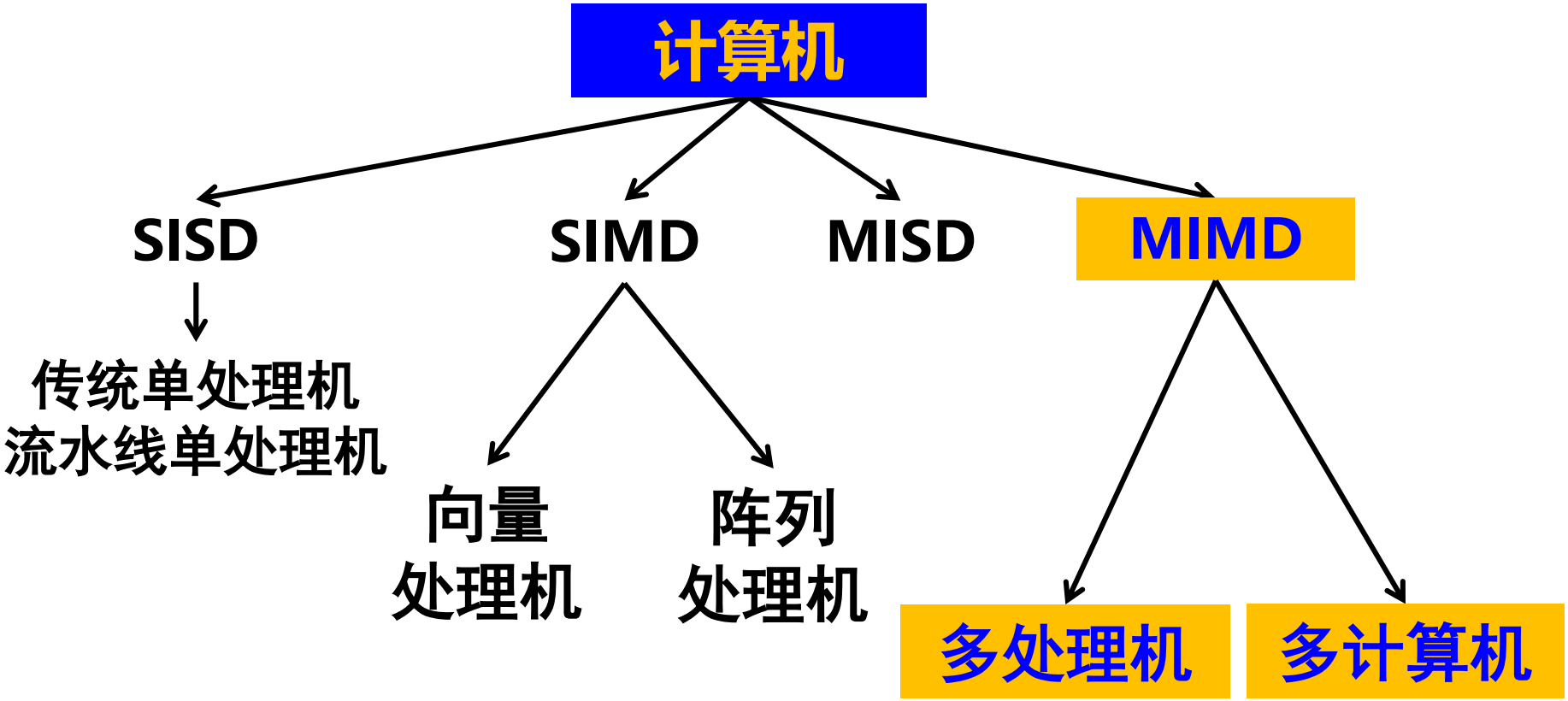
7.3 多核处理器

7.4 多处理机的多Cache一致性

回顾：弗林分类法

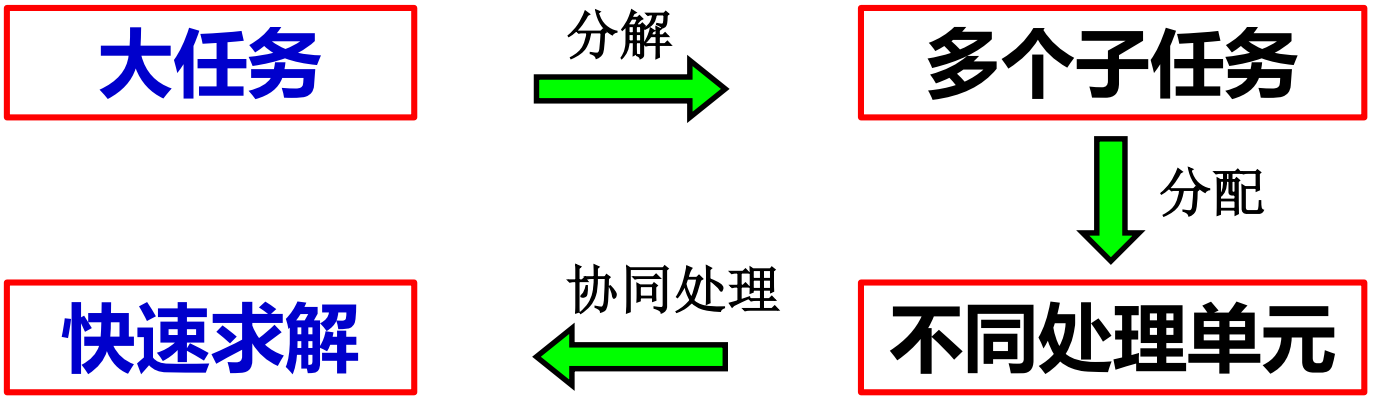
北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 弗林分类法：并行处理能力



7.1.1 多处理机定义

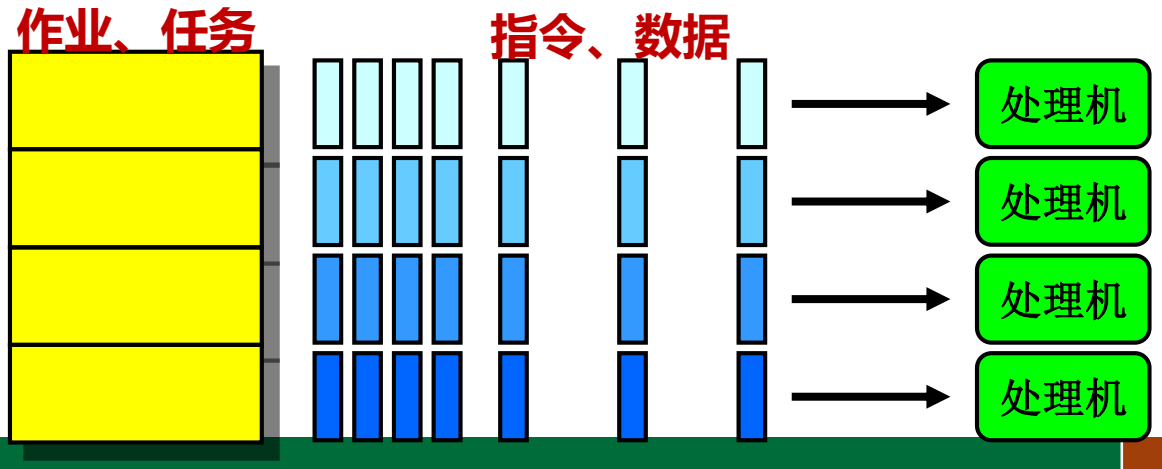
- **多处理：**
  - 在多个处理机上运行同一道程序或作业的不同任务。可以串行，也可以并行。
- **并行处理：并行计算，高性能计算**
  - 同时执行两个或更多个处理的一种计算方法。



7.1.1 多处理机定义

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- **多处理机：**
  - 由两台及以上处理机组成的计算机系统；
  - 各处理机有自己的控制部件、局部存储器，能执行各自的程序，可以共享公共主存和所有外设；
  - 各处理机通过某种形式互连，相互通信；
  - 实现作业、任务、指令、数据等各个级别的并行；
  - 属于MIMD。



7.1.2 多处理机分类

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 根据实现并行技术途径，分为3类:
- 同构型：基于资源重复
  - ➢处理机类型/功能相同
  - ➢并行处理或执行任务
- 异构型：基于时间重叠
  - ➢处理机类型/功能不同
  - ➢重叠处理各任务
- 分布式：`
  - ➢处理机类型/功能相同或不同
  - ➢并行/协同完成任务处理
  - ➢由统一操作系统统一管理资源

7.1.2 多处理机分类

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 根据耦合度分成两类：
- 紧耦合多处理机：直接耦合系统
  - 通过公共硬件（如存储器，I/O系统等）通信
  - 典型：多处理器系统，多核处理器等
  - 多处理器系统：各计算节点共享统一编址的存储器，可通过LOAD和STORE指令访问系统中的存储器，也被称为共享存储多处理器（Shared memory multiProcessor, SMP）系统
  - 一般将紧耦合多处理机称为多处理机

- **松耦合多处理机：间接耦合系统**
  - ➢通过通道、通信线路或网络、消息传递系统通信
  - ➢典型：机群，计算机网络等
  - ➢一般将松耦合多处理机称为多计算机

7.1.3 多处理机特点和主要技术问题

- 多处理机主要技术问题：
  - **结构灵活性与通用性：**以适应、满足多种并行计算算法不同的需求
  - **进程通信方法：**共享内存、互连网络、消息传递机制等
  - **运行模型：**数据并行 + 处理并行
    - 基于共享内存的运行模型
    - 基于消息传递机制的运行模型
  - **并行性表达：**
    - 编译程序自动发现：如 并行C或Fortran
    - 并行程序设计语言：如 HPF(High Performance Fortran)
    - 运行时库：如 LAPACK
  - **并行算法**

7.1 多处理机概念

7.2 多处理机结构

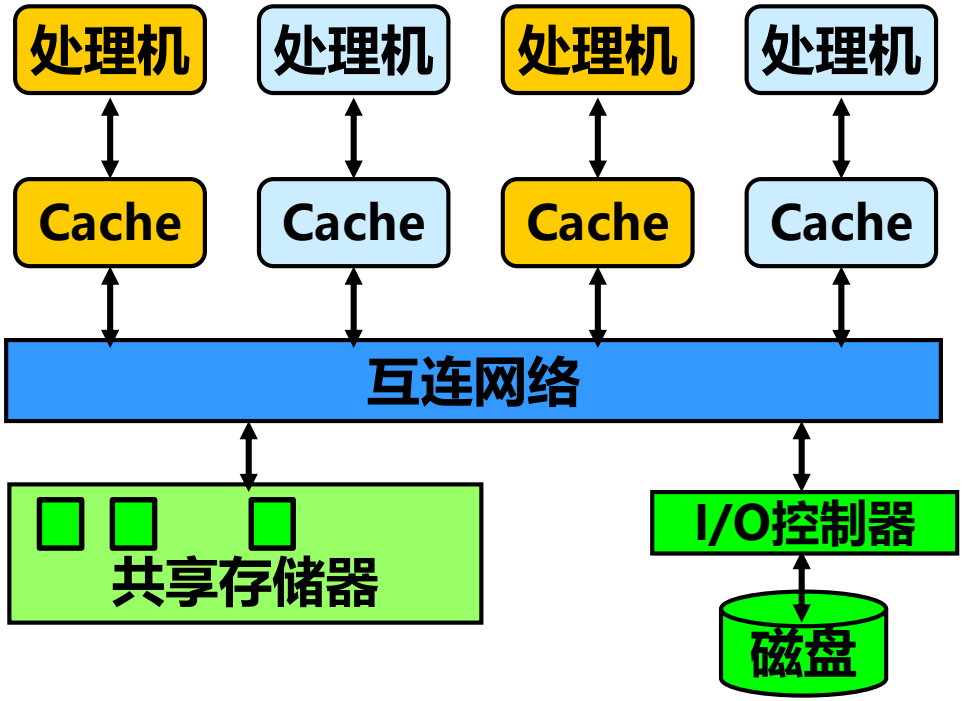
7.3 多核处理器

7.4 多处理机的多Cache一致性

- 从存储器的分布和使用上看，多处理机系统分为两种结构：
  - **共享存储器结构**
  - 均衡存储器访问UMA结构
  - 非均衡存储器访问NUMA结构
    - ccNUMA
    - NCC-NUMA
    - COMA结构
  - **分布式存储器结构**

7.2.1 共享存储器结构

- 各处理机通过互连网络**共享存储器和I/O设备**，并通过共享存储器相互**联系**。



共享存储器结构的多处理机系统

7.2.1 共享存储器结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

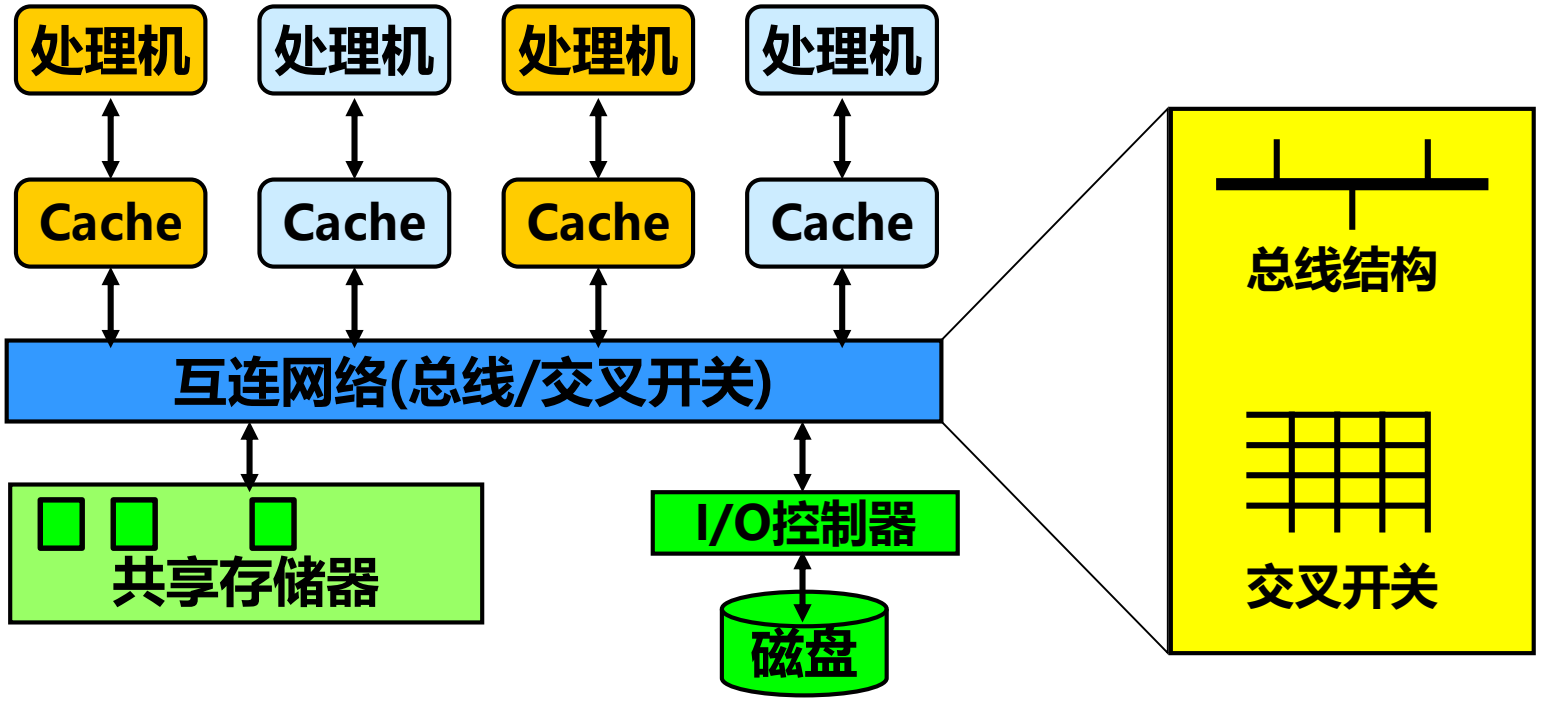
- 特点：
  - 各处理机共享存储器，并通过对共享存储器读/写实现相互通信；
  - 对存储单元的任何修改对其他处理机都是可见的；
  - 延迟低，但扩展性差；

共享数据进入Cache产生了一个新的问题：Cache的一致性问题

1. UMA结构

- UMA = Uniform Memory Access
- 均衡存储器访问结构
- 或 集中式共享存储器 (Centralized Shared-Memory) 结构
- 特点:
  - 各处理机对存储器的访问时间、访问功能相同
- 这种结构的处理机称为对称多处理机 (SMP = Symmetric Multiprocessors)

1. UMA结构



UMA结构的多处理机系统

- 互连网络可以是总线、交叉开关或多级交换网络。
- 大多数的对称多处理机采用总线连接。

1. UMA结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- **优点：**
  - （1）性能提高
  - （2）高可用性
  - （3）增量式增长
  - （4）可扩展性好
  - （5）透明
- **缺点：**
  - 所有处理机对共享存储器的访问都要经过互连网络，当规模较大时，访问延迟较大。因此通常增设大容量本地Cache
- 适合小规模时采用

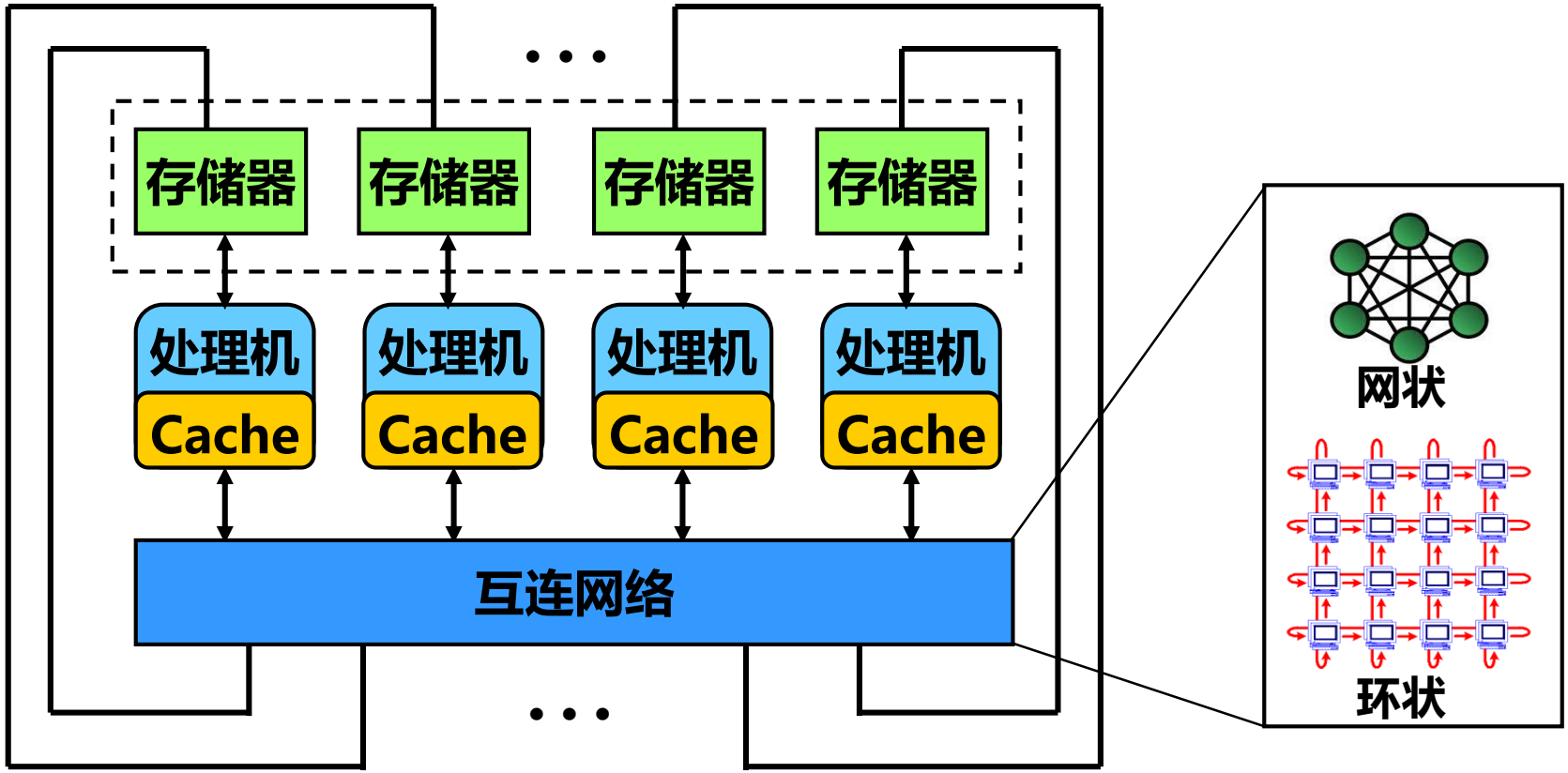
2. NUMA结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- NUMA = Non Uniform Memory Access
- 非均衡存储器访问结构
- 也称为**分布式共享存储器** (DSM=Distributed Shared-Memory) 结构
- 特点:
  - 分布于各个处理机的存储器被统一编址，可由所有处理机共享;
  - 根据存储器位置的不同，各处理机对存储器的访问时间不相等。处理机访问本地存储器的速度较快，通过互连网络访问其他处理机上的远地存储器相对较慢。

2. NUMA结构

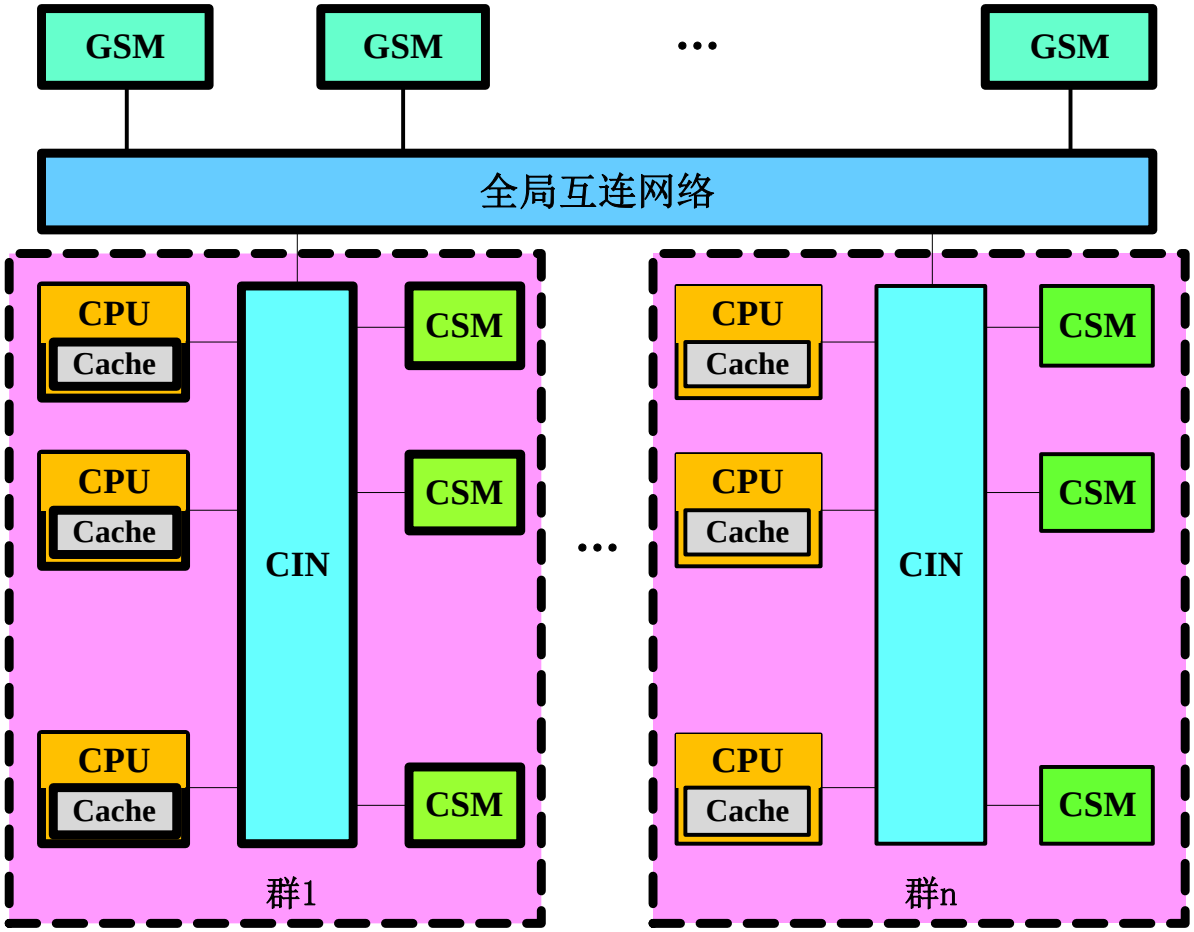
- ☑ 各处理机拥有自己的本地存储器，可以独立工作；
- ☑ 各处理机借助互连网络、通过消息传递机制相互通信；



NUMA结构的多处理机系统 – 共享本地存储器

2. NUMA结构

允许各处理机拥有自己独立结构，甚至可以是SMP。



NUMA结构的多处理机系统 – 层次式机群模型

2. NUMA结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- **优点：**
  - 比SMP扩展性好，并行程度更高，性能更好；
  - 采用与SMP相同的编程模型，为SMP编写的程序仅需少量修改即可移植运行；
  - 每个处理机都可以访问较大的存储空间，因此可以更高效地运行大程序；
  - 实现数据共享时不需要移动数据；
  - 传递包含指针的数据结构比较容易；
  - 系统构建成本较低，利用成熟技术搭建系统。

- **缺点：**
  - 如果过多地访问远程存储器，则性能会下降；
  - 对存储器的访问不透明，需要处理分页（例如哪个页面在哪个存储器中）、进程分配等，对软件设计要求较高

3. ccNUMA

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- ccNUMA = Cache-coherent Non Uniform Memory Access
- 高速缓存一致性非均匀存储访问
- 在NUMA多处理机中，逻辑上共享的存储器在物理上是分布的。如果各处理机Cache内容一致，则将这种NUMA称为ccNUMA
- 绝大多数商用ccNUMA多处理机系统使用基于目录的高速缓存一致性协议。

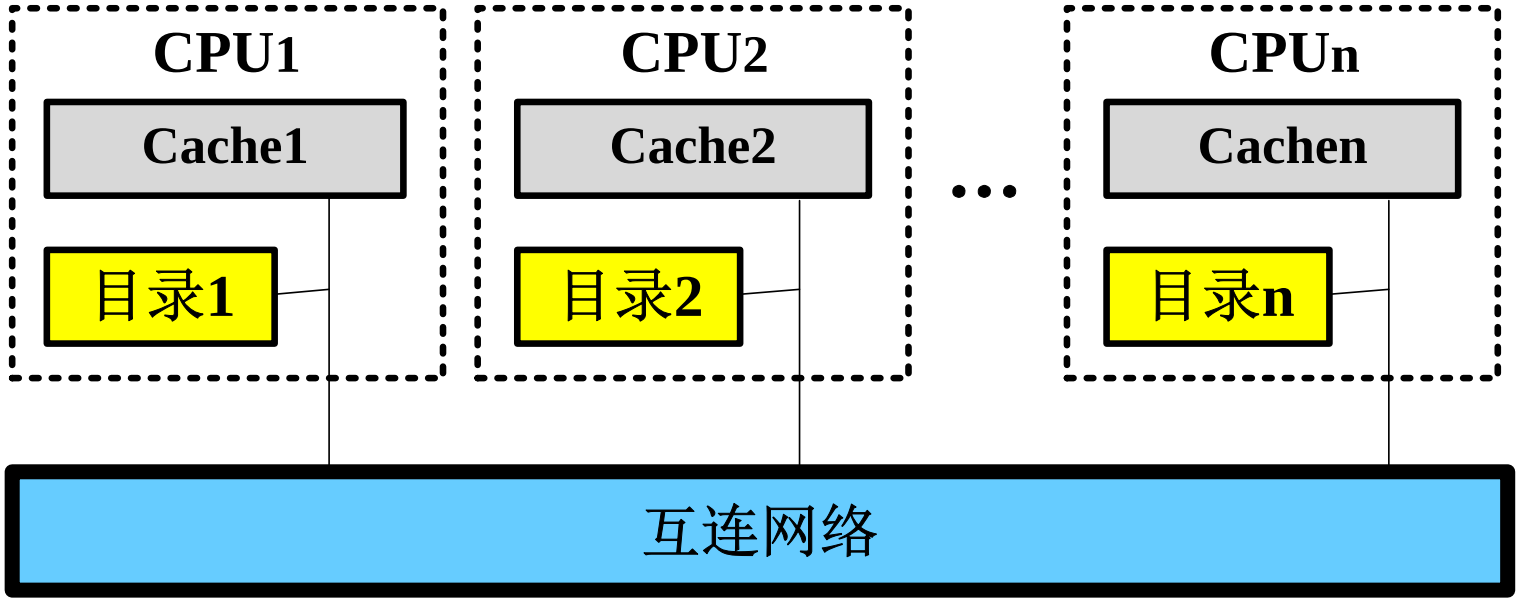
4. COMA

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

- COMA = Cache-Only Memory Architecture
- 仅用高速缓存存储器结构
- NUMA的一个特例，只是将NUMA中的分布存储器换成了Cache
  - 各处理机结点上没有主存储器，没有存储层次结构，仅有Cache
  - 所有的高速缓存构成了全局地址空间，全部Cache组成了全局虚拟地址空间
  - 对远程Cache的访问通过分布式Cache目录进行

4. COMA

- 各处理机结点上没有主存储器，没有存储层次结构，仅有Cache
- 所有的高速缓存构成了全局地址空间，全部Cache组成了全局虚拟地址空间
- 对远程Cache的访问通过分布式Cache目录进行



COMA多处理机结构

7.2.2 分布式存储器结构

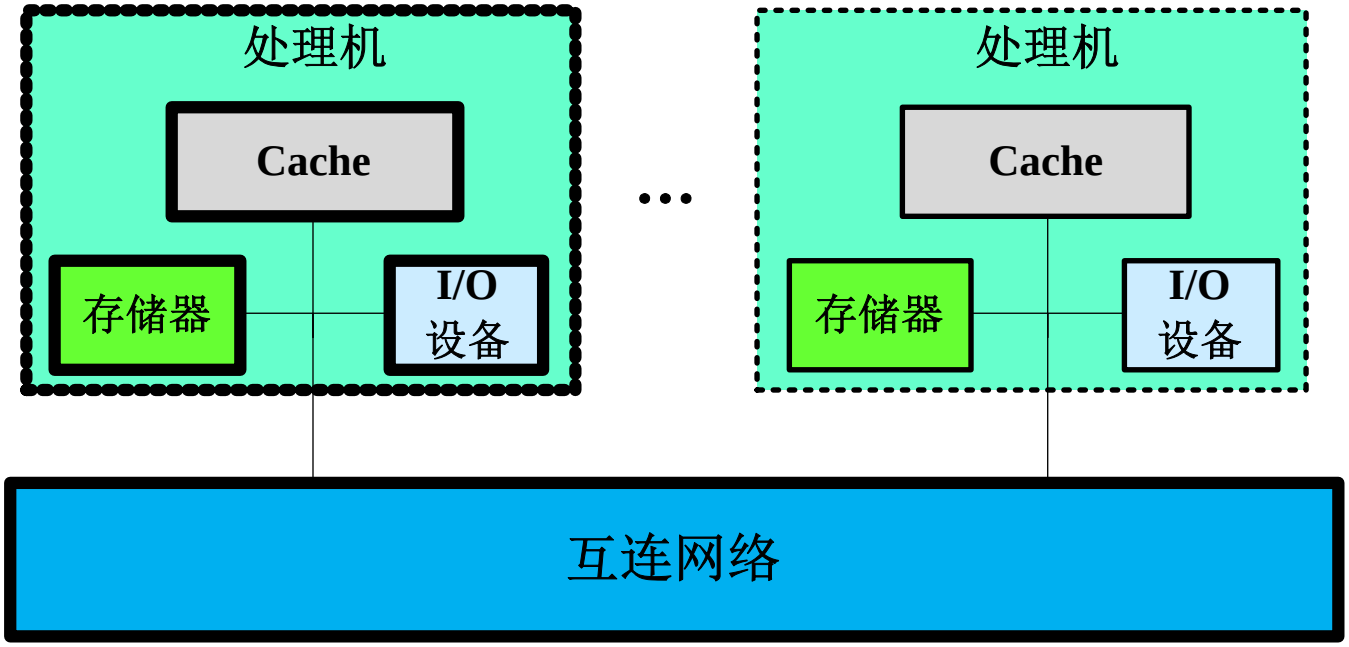
北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 也被称为非远程存储访问（NORMA，No-Remote Memory Access）模型
- 各处理机拥有自己的本地存储器，在本地操作系统控制下独立工作。
- 各处理机的本地存储器是私有的，不能被其他处理机访问。
- 各处理机借助互连网络、通过消息传递机制相互通信，实现数据共享。
- 大规模并行处理机（MPP）、机群（cluster）等采用了这种结构。

7.2.2 分布式存储器结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 各处理机借助互连网络、通过消息传递机制相互通信，实现数据共享



分布式存储器多处理机

7.2.2 分布式存储器结构

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- **优点：**
  - 结构灵活，扩展性较好
- **缺点：**
  - 任务传输以及任务分配算法复杂，通常要设计专有算法
  - 处理机之间的访问延迟较大
  - 需要高带宽的互连

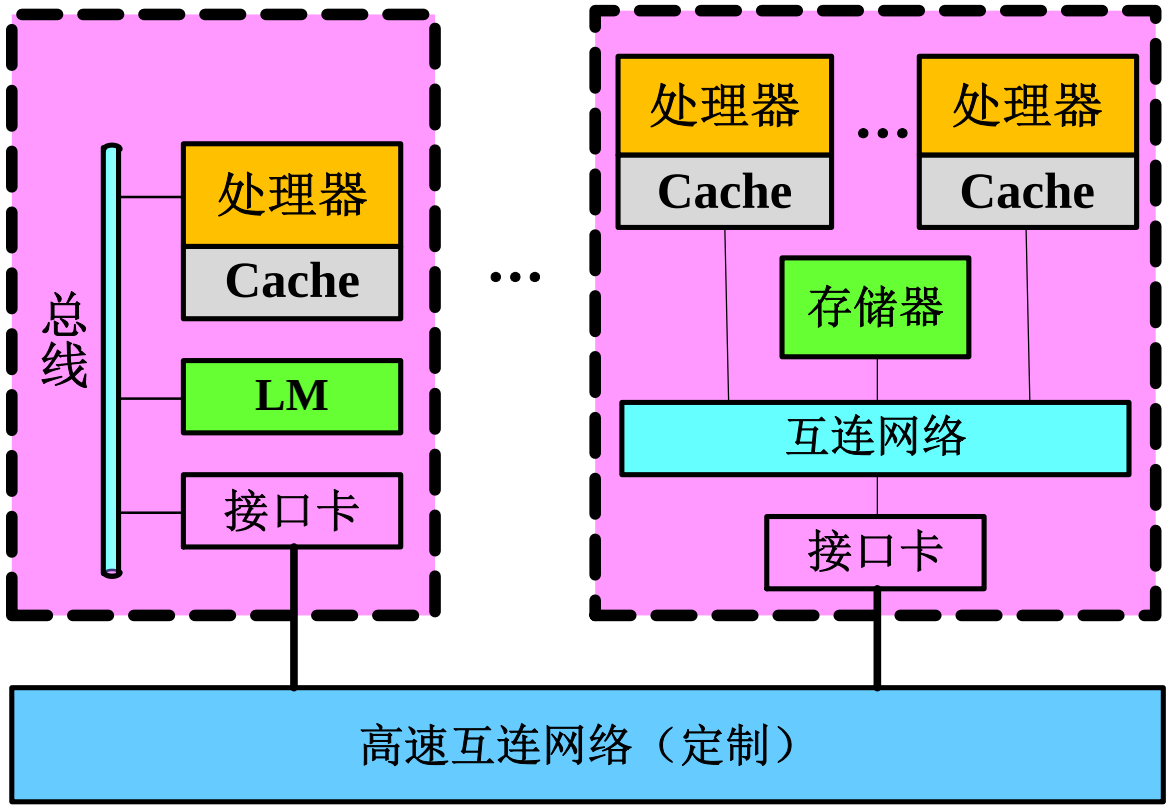
7.2.3 大规模并行处理机

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- MPP = Massively Parallel Processor
- 由几百或几千台高性能、低成本处理机组成的大规模并行计算机系统
  - 每个处理机都有自己的私有资源，如内存、网络接口等
  - 每个处理器能直接访问的只有本地存储器，但不能直接访问其它处理器的存储器
  - 处理机之间以**定制的高带宽、低延迟**的高速互连网络互连

7.2.3 大规模并行处理机

- MPP系统大多采用分布式存储结构。所有存储器在物理上是分布的，而且都是私有的。每个处理器能直接访问的只有本地存储器，不能直接访问其它处理器的存储器。



MPP结构

7.2.3 大规模并行处理机

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

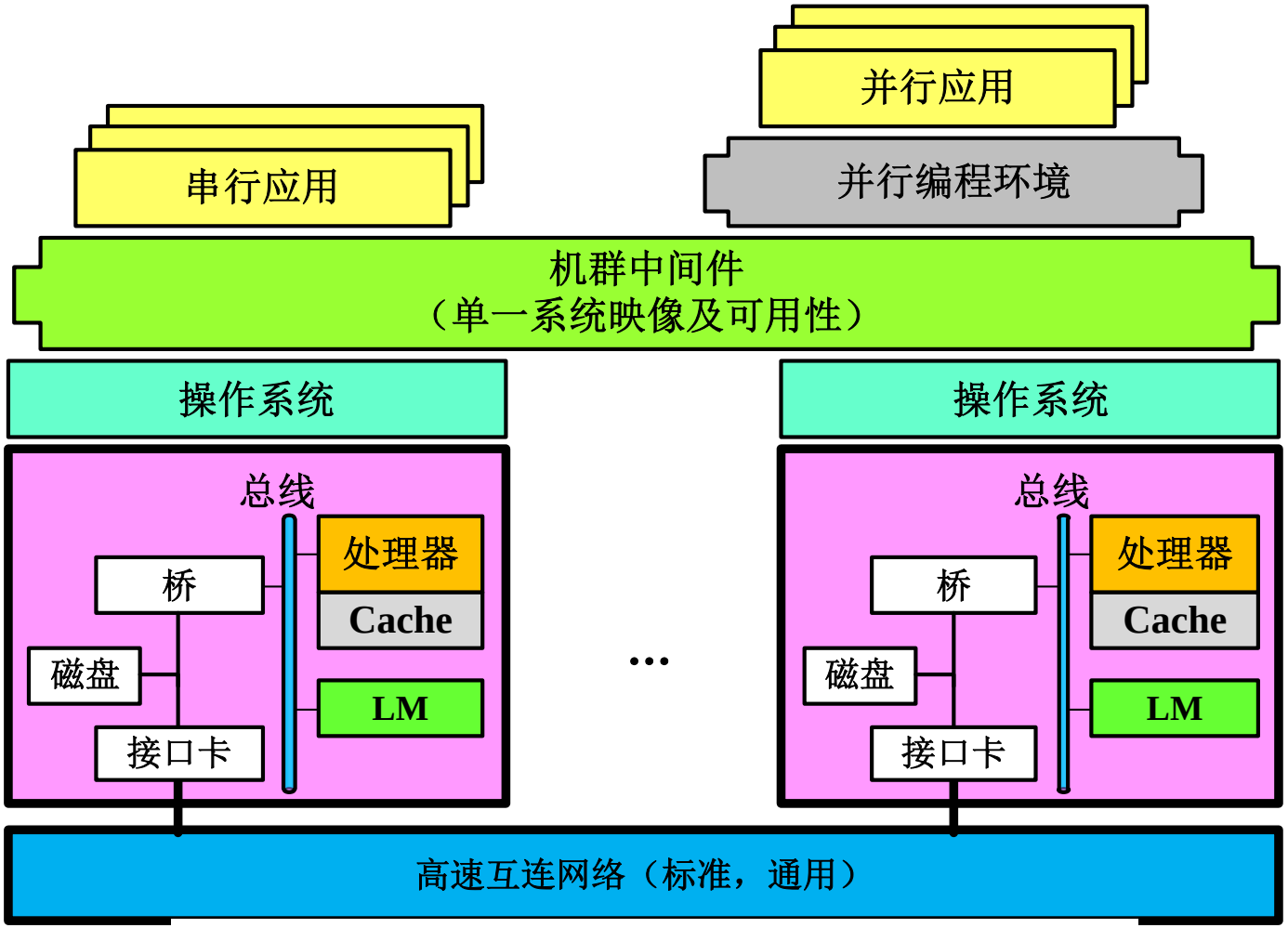
- 特点：
  - 处理结点采用商用处理器
  - 系统中有物理上的分布式存储器
  - 采用高通信带宽和低延迟的互联网络（专门设计和定制的）
  - 能扩放至成百上千乃至上万个处理器
  - 它是一种异步的MIMD机器
    - 程序系由多个进程组成，每个都有其私有地址空间，进程间采用传递消息相互作用

7.2.4 机群

- 也被称为集群
- COW = Cluster of Workstations
- NOW = Network of Workstations
- 通过一组松散耦合的计算机软件和/或硬件连接起来、高度紧密地协作完成计算工作的计算机系统
  - **结点**：单独运行的商品化计算机
  - **网络**：高速通用网络，如局域网或互连网络连接
  - **操作系统**：并行编程环境，系统资源管理和相互协作
  - **集群中间件**：屏蔽差异，在异构系统上提供统一运行环境和服务

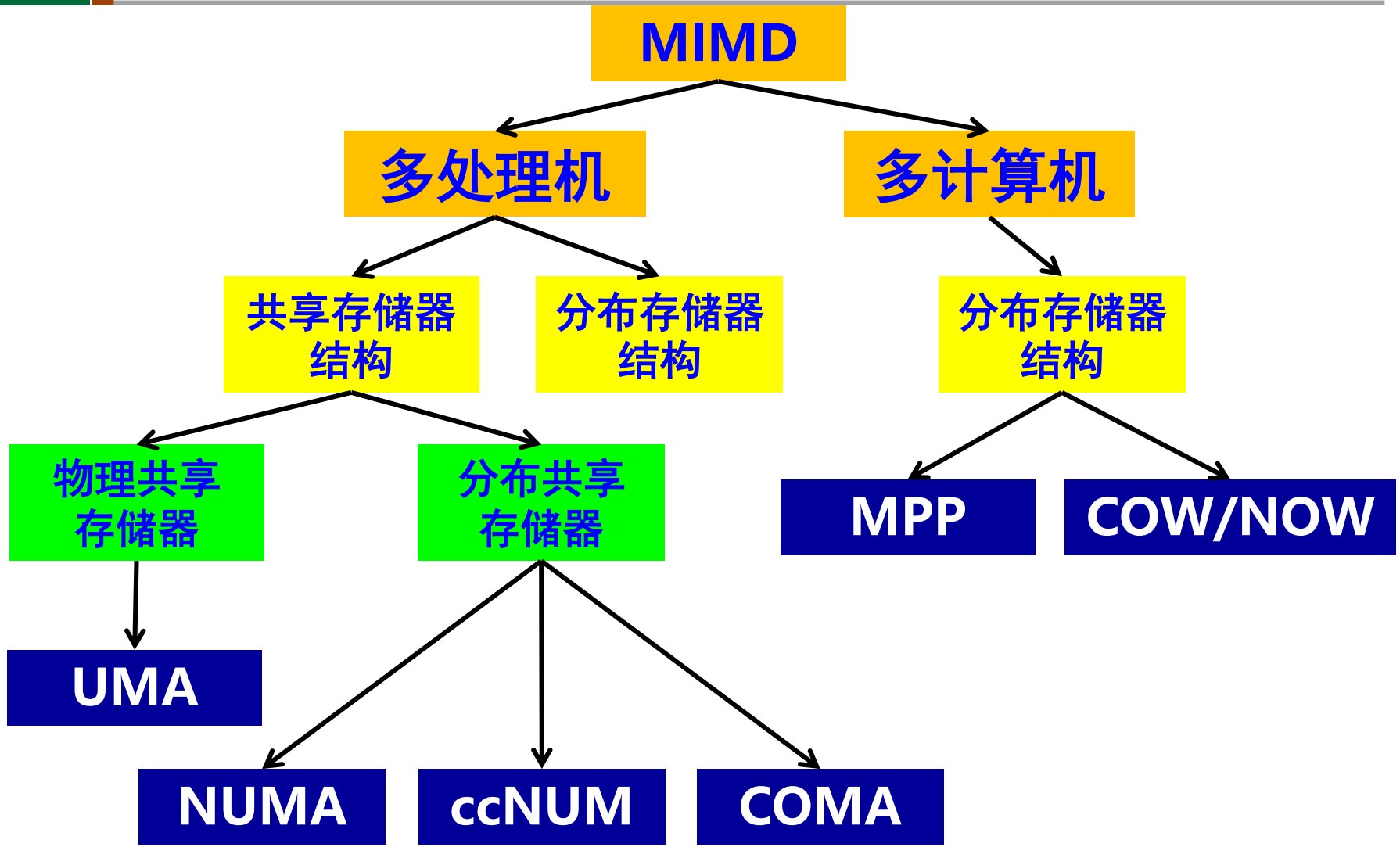
7.2.4 机群

从结构和结点之间的通信方式来看，机群属于非均匀存储访问的MIMD型的分布式存储并行计算机



机群系统结构

MIMD计算机分类


MIMD
多处理机
多计算机
共享存储器结构
分布存储器结构
分布存储器结构
物理共享存储器
分布共享存储器
MPP
COW/NOW
UMA
NUMA
ccNUM
COMA

7.1 多处理机概念

7.2 多处理机结构

7.3 多核处理器

7.4 多处理机的多Cache一致性

7.3 多核处理器

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Multi-core Processor
- 片上多处理机 (CMP，Chip Multi-Processors)
- 片上多处理机系统 (MPSoC，Multiprocessor System-on-Chip)
- 是多处理机的一种特殊形式： SMP on a single chip
- 1996年，美国斯坦福大学首次提出了片上多处理器(CMP)思想
- 2000年，IBM于发布了世界上第一个双核处理器POWER4
- 2005年，Intel和AMD多核处理器开始大规模应用

微处理器发展（回顾）

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- 问题：晶体管数量 ≠ 能力
- 如何利用晶体管资源?
  - 更好的核
    - 超流水，超标量，…
    - 更好的向量处理（SIMD）
    - 问题：存储器墙 = 存储器速度不能满足CPU速度要求
  - 更大的Cache
    - 改善访存速度
  - 更多的核
    - 问题：存储器墙

微处理器发展（回顾）

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- **制造厂商目标**
  - 尽可能多得收益：卖处理器 …
  - 但客户仅在运行的更快时才购买
  - 提升CPU功能
- **如何提升CPU功能?**
  - 更高的时钟频率
  - 更多的并行性
    - 指令级并行 (ILP = Instruction Level Parallelism)
    - **线程级并行** (TLP = Thread Level Parallelism)

线程级并行 (TLP)

- 硬件线程 (例如 SMT：超线程)
- 多核

- **进程定义：资源分配单位**
  - 一个独立的进程空间，可装入进程映像；
  - 进程关联的执行文件；
  - 进程所用系统其它资源(如设备、文件等）；
  - 一个或多个线程。 进程在创建时一般同时创建好第一个线程，其它线程按需要由用户程序请求创建。
- **线程定义：CPU分配单位（执行单位）**

硬件多线程（hardware multithreading）

- 在传统的CPU中，线程的切换包含了一系列开销，频繁切换回极大影响系统性能。在处理器中多开辟几份状态，为每个线程提供单独的通用寄存器组、单独的程序计数器等，线程切换时，处理器只需要激活选中的寄存器，达到快速切换线程的目的，这种方式叫做**硬件多线程（Hardware Multithreading）**。

线程切换实现方式

- 1）细粒度多线程
- 处理器每个时钟周期轮流发射不同线程的指令。
- 要求多个线程之间的指令无相关性，可以乱序并行执行。该方式下，处理器能在每个时钟周期切换线程。
- 2）粗粒度多线程
- 仅在一个线程出现较大开销的阻塞时，才切换线程，如Cache缺失。当发生流水线阻塞时，必须清除被阻塞的流水线，新线程的指令开始执行前需要重载流水线，开销较上一种较大。
- 3）同时多线程（SMT，Simultaneous Multi Threading）
- 在超标量处理器中，处理器能一次发射多条指令，如果这多条指令来自于不同的线程，即多个线程的指令同时被发射，那么这种工作方式就叫同时多线程。SMT是前两种方式的变体，在指令级并行的同时，实现线程级的并行，即在同一时钟周期内，发射不同线程中的多条指令执行。

多核处理器结构设计主要考虑以下因素(1/2)：

- (1) 同构还是异构
  - 同构 CMP：集成多个相同的处理器核，同一个任务可以分配给任意一个核处理，简化了任务分配。
  - 异构 CMP：包含了不同结构的处理器核，用不同类型的处理器核处理不同的任务，是异构体系结构处理器的优势所在。
- (2) 核的数量
  - 双核，四核，八核，…

多核处理器结构设计主要考虑以下因素(2/2)：

- (3) Cache的设置与访问
  - 分布式存储器结构
  - 共享存储器结构，共享访问的Cache多大
  - Cache层次
- (4) 核间通信技术
  - 通过基于总线共享的Cache
  - 通过片上互连网络

7.1 多处理机概念

7.2 多处理机结构

7.3 多核处理器

7.4 多处理机的多Cache一致性

7.4.1 存储器一致性定义（自学）

如果一道程序的任何运行结果都与按假定序列（也就是说，将所有进程发出的读写操作排成一个全序）执行的结果一样，则称多处理机存储器是一致性的。

存储器一致性是指一个系统正确执行存储器操作的能力。各处理机看到存储器的值是一致的。

强调的是顺序一致性

顺序一致性

- **三种次序：**
  - 程序次序
  - 执行次序
  - 存储器访问次序
- **顺序一致性：** 存储器访问次序同程序执行次序一致
- **顺序不一致性：** 存储器访问次序同程序执行次序不一致

多机系统中的难点

- 在多机系统中，只检测多处理机的内部相关性是不够的，**顺序一致性比较难维持**，这是因为：
  - （1）多个指令流的执行次序并不一定一样，如无同步，大量指令交叉执行可能发生。
  - （2）如想改善性能，每台处理机中指令执行的次序与程序次序不同，则会发生更多的指令交叉执行。
  - （3）由于互连网络对不同的存储器访问而引起不一样且无法预测的延迟（如冲突），可能存储器访问命令按程序次序发生，但到达存储器的次序不一样了。
  - （4）存储器访问对Cache-based的系统就**不是原子访问**，如并非所有的Cache副本同时修改或失效，则可按程序发出访问命令并到达Memory，但完成不一定按程序次序，不同处理机观察到的将是不同的交叉。此时，一个程序可能会产生多种执行情况。

存储器访问的原子性

- 如果所有处理机同时看到一份数据拷贝，则系统中的存储器访问是原子性的。
- 如果存操作所写的值对所有处理机同时可读，则此写操作是原子操作。
- 如果存储器更新时为所有处理机所知道，那么这就是原子共享存储器访问。

所以原子存储器成为顺序一致的充分必要条件是：要所有存储器访问的执行都必须保持所有各个程序的次序。

多处理机的存储器行为

- （1）所有处理机都保持程序次序，并且有相同的执行次序。
- （2）所有处理机都可以不按程序次序执行，但有相同的执行次序。
- （3）不同处理机都可以不按程序次序执行，并且有不同的执行次序。

所以原子存储器成为顺序一致的充分必要条件是：要所有存储器访问的执行都必须保持所有各个程序的次序。

四种存储器一致性模型

- 顺序一致性模型
- 弱一致性模型
- 处理机一致性
- 释放一致性

有兴趣同学请自查资料。