

计算机体系结构--

# 基本概念

[zhaofangbupt@163.com](mailto:zhaofangbupt@163.com)



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

# 主要内容

- ◆ 1. 计算机系统结构的概念.....●
- ◆ 2. 计算机系统结构的发展.....●
- ◆ 3. 影响体系结构的因素.....●
- ◆ 4. 定量分析技术基础.....●
- ◆ 5. 体系结构中并行性的发展.....●

# 并行性的概念

□ **并行性** (Parallelism)：计算机系统在**同一时刻**或者**同一时间间隔内**进行多种运算或操作

- 只要在时间上相互重叠，就存在并行性
- 目的：提高效率

## □ 两重含义

- **同时性** (Simultaneity)：两个或两个以上的事件在同一时刻发生
- **并发性** (Concurrency)：两个或两个以上的事件在同一时间间隔内发生

# 并行性等级

□ 从**处理数据**的角度来看，从低到高可分为：

➤ **字串位串**：每次只对一个字的一位进行处理

- 最基本的串行处理方式，不存在并行性

➤ **字串位并**：同时对一个字的全部位进行处理，不同字之间是串行的

- 开始出现并行性，又称为**位级并行**
- 上个世纪70年代-80年代，处理器芯片上的位级并行占主导地位

# 并行性等级

- **字并位串**：同时对许多字的同一位（称为位片）进行处理
  - 具有较高的并行性
  - 开始进入并行处理领域
- **全并行**(字并位并)：同时对许多字的全部位或部分位进行处理
  - 最高一级的并行

# 并行性等级

□ 从计算机信息加工的各个步骤和阶段分为：

- 存储器操作并行——用单体多字、多体单字或多体多字方式在一个存储周期内访问多个字
  - 并行存储器系统、相联处理机(相联存储器为核心)
- 处理器操作步骤并行——指令的取指、分析、执行，浮点加法的求阶差、对阶、尾加、舍入、规格化等操作，执行步骤在时间上重叠流水地进行
  - 流水线处理机

# 并行性等级

- **处理器操作并行**——为支持向量、数组运算，通过重复设置大量处理单元，在同一控制器的控制下，按照同一条指令的要求对多个数据组同时操作
  - 阵列处理机
- **指令、任务、作业并行**——是较高级的并行，指令级以上的并行是多个处理机同时对多条指令及有关的多数据组进行处理，构成多指令流多数据流计算机
  - 多处理机

# 并行性等级

□ 从**执行程序**的角度来看，从低到高可分为：

➤ **指令内部并行**：单条指令中各微操作之间的并行

➤ **指令级并行** (Instruction Level Parallelism, ILP)：

并行或并发地执行两条或两条以上的指令

- 流水线技术使多条指令重叠执行
- RISC方法，计算机达到一条指令/每个时钟周期
- 超标量方法，计算机达到启动多条指令/每个时钟周期，产生多个运算结果/每个时钟周期



# 并行性等级

➤ **线程级并行** (Thread Level Parallelism, TLP): 并行执行两个或两个以上的线程

- 通常是以一个进程内派生的多个线程为调度单位
- 仅拥有运行过程中必不可少的一点资源
- 能够提高程序的并发和并行程度，与多核技术相结合，能达到很好地开发并行性的效果
- 2000以来，开发线程级并行的计算机越来越多

# 并行性等级

➤ **任务级或过程级并行**：并行执行两个或两个以上的过程或任务（程序段）

- 以子程序或进程为调度单元

➤ **作业或程序级并行**：并行执行两个或两个以上的作业或程序

□ **高级并行性**通常包含**低级并行性**，低级并行性蕴涵于高级并行性

# 并行性实现方式

- **单处理机体系**中，并行性升到某一级别后，需要通过软件来实现
- **多处理机体系**中，各任务或作业通常由各处理机完成，其并行性多由硬件实现
- 计算机中可采取多种并行性措施：
  - 可以有**数据处理**方面的并行性，可以有**执行程序**方面并行性
- **并行处理**是体系结构、硬件、软件、算法、语言等多方面综合研究的领域

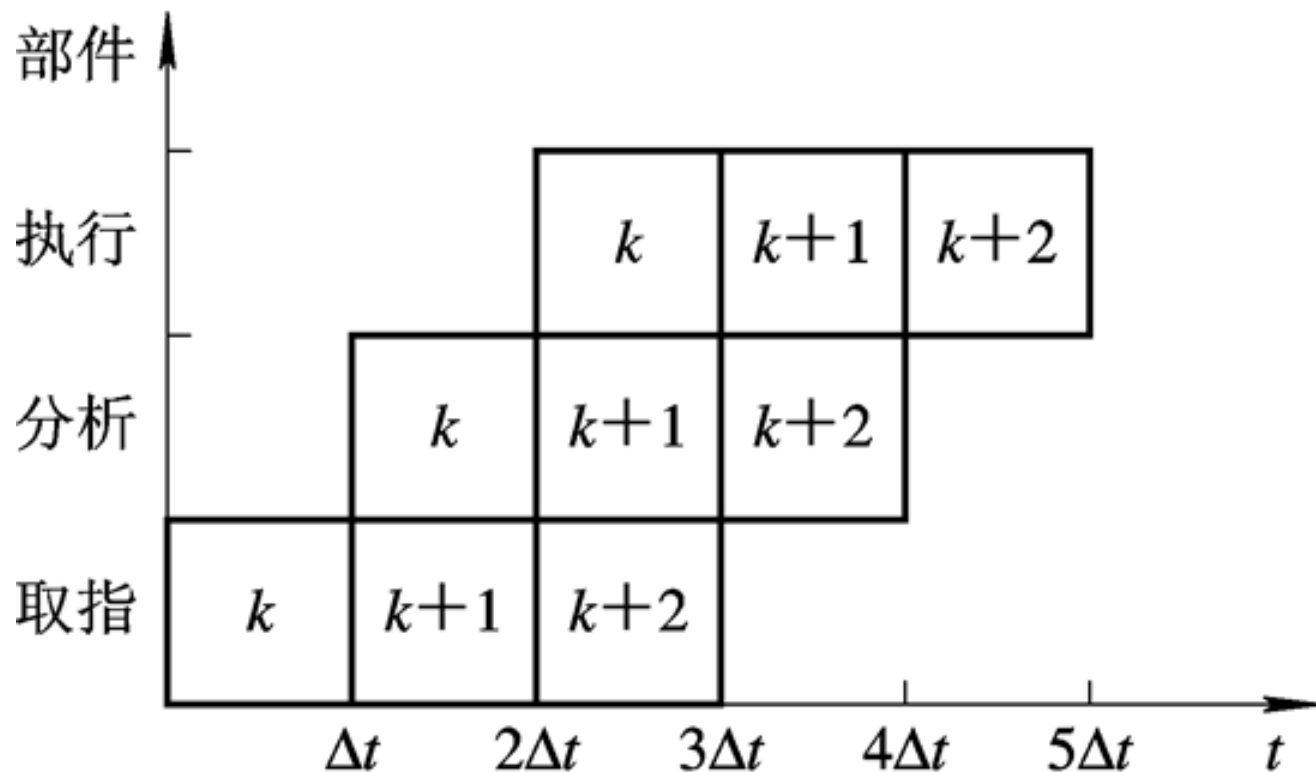
# 提高并行性的技术途径

## □ 时间重叠 (time-interleaving)

- 在并行性概念中引入时间因素，即多个处理过程在时间上相互错开，**轮流重叠**地使用同一套硬件设备的各个部分，以加快硬件周转而赢得速度
- 并行性中的并发性
- 优点：不需要增加硬件
  - 例如：指令的执行分为取指、分析、执行三个步骤



# 提高并行性的技术途径

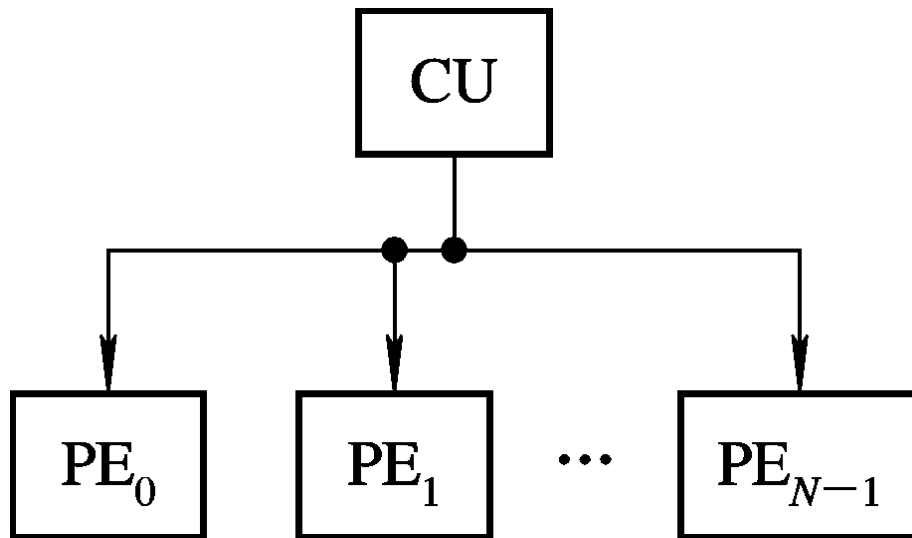


# 提高并行性的技术途径

## □ 资源重复 (resource-replication)

- 在并行性概念中引入空间因素，通过重复设置硬件资源，可以大幅度提高并行处理能力
- 并行性的同时性
- 随着集成电路的发展，硬件价格和体积不断下降，资源复制已经成为提高系统性能最有效的手段
  - 例如：多处理器/机和多计算机

# 提高并行性的技术途径



□ 在现代的计算机系统中，经常是**同时运用**时间重叠和资源重复

# 提高并行性的技术途径

## □ 资源共享 (Resource Sharing)

- 这是一种**软件方法**，它使多个任务按一定时间顺序轮流使用同一套硬件设备
- 并行性的并发性
- 多道程序分时系统、计算机网络和分布处理系统等
- 既降低了成本，又提高了计算机设备的利用率

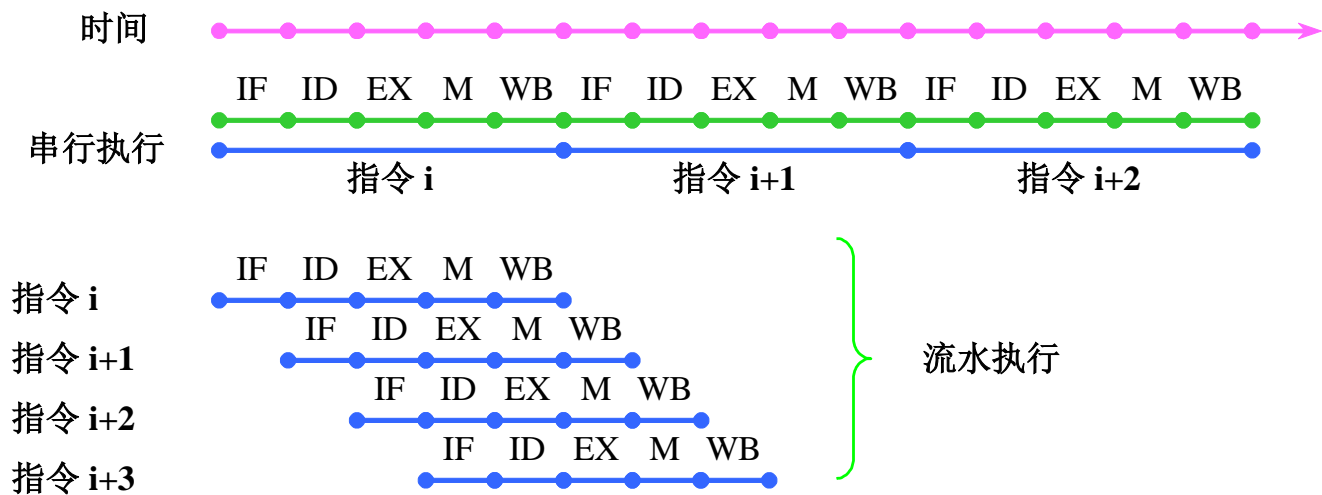


# 单机系统中并行性的发展

- 在发展高性能单处理机过程中，起主导作用的是  
**时间重叠原理**
- 实现时间重叠的基础：**部件功能专用化**
  - 把一件工作按功能分割为若干相互联系的部分
  - 把每一部分指定给专门的部件完成
  - 按时间重叠原理把各部分的执行过程在时间上重叠起来，使所有部件依次分工完成一组同样的工作

# 单机系统中并行性的发展

- 例如：如果把一条指令解释指令的5个过程，就分别需要5个**专用**的部件：取指令部件(IF)、指令译码部件(ID)、指令执行部件(EX)、访问存储器部件(M)和写结果部件(WB)



# 单机系统中并行性的发展

□ 在单处理机中，资源重复原理的运用也已经十分普遍

➤ 多体存储器

➤ 多操作部件

- 通用部件被分解成若干个专用部件

- 实现了指令级并行

➤ 阵列处理机（并行处理机）

□ 在单处理机中，资源共享的概念形成虚拟机概念

➤ 分时系统

# 多机系统中并行性的发展

□ 遵循时间重叠、资源重复、资源共享原理，发展为3种不同的多处理机：

- 同构型多处理机
- 异构型多处理机
- 分布式系统

□ 耦合度：

- 反映多机系统中各机器之间物理连接的紧密程度和交互作用能力的强弱
- 紧密耦合和松散耦合

# 多机系统中并行性的发展

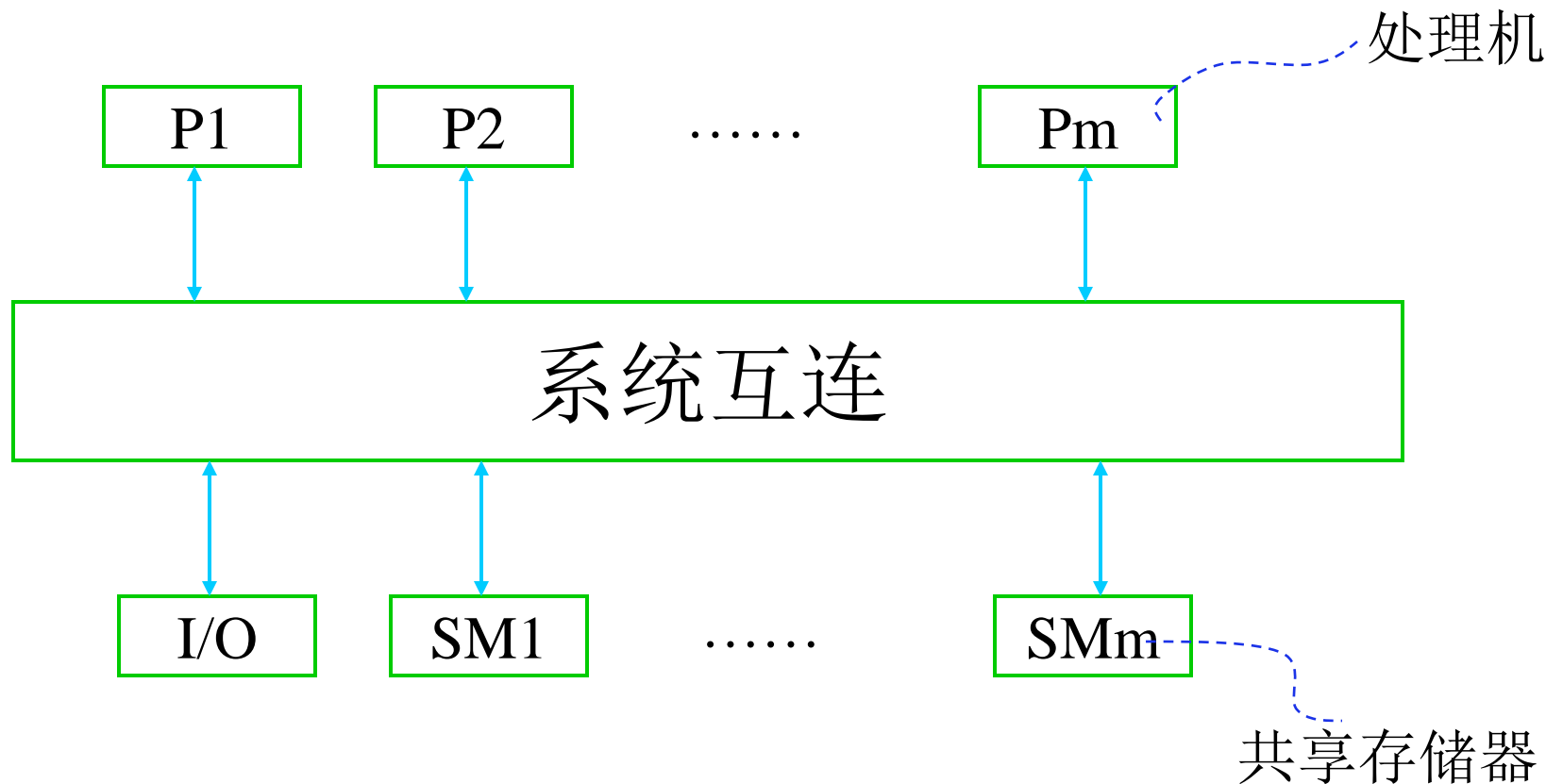
## □ 紧密耦合系统（又称直接耦合系统）

- 计算机之间的物理连接的频带较高，通过总线或高速开关互连，可以共享主存

## □ 松散耦合系统（又称间接耦合系统）

- 通过通道或通信线路实现计算机间互连，共享外存设备。机器间相互作用是在文件或数据集一级上进行
  - 多台计算机和共享的外存设备连接，不同机器间实现功能专用化
  - 计算机网络，实现更大范围资源共享

# 多机系统中并行性的发展



# 多机系统中并行性的发展

## □ 功能专用化（实现时间重叠）

- 输入/输出功能的分离→专用外围处理机
- 许多主要功能（数组运算、高级语言翻译等）分离出来→专用处理机

## ➤ 异构型多处理机系统

- 由多个不同类型、至少担负不同功能的处理机组成，它们按照作业要求的顺序，利用时间重叠原理，依次对多个任务进行加工，各自完成规定的功能动作

# 多机系统中并行性的发展

## □ 机间互连机制

- 容错系统：设置多台相同类型计算机构成的容错系统
- 可重构系统：一旦某处理机发生故障，系统就会重新组织，降低档次继续运行，直到排除故障为止
- 目标：通过多处理机并行处理提高整个系统的速度

## □ 同构型多处理机系统

- 由多个同类型或至少担负同等功能的处理机组成，它们同时处理同一作业中能并行执行的多个任务



# 计算机系统结构的分类

## □ 从并行性的角度对计算机系统结构分类

- Flynn(弗林)分类法

- Kuck(库克)分类法

## □ 从系统的最大并行度对计算机系统结构分类

- 冯氏分类法

## □ 从并行度和流水线角度对计算机系统结构分类

- Handler分类法

## □ 按照计算机的性能和应用特征分类

- 桌面计算机、服务器型计算机和嵌入式计算机

# 计算机系统结构的分类

## □ Flynn分类法

- 按照指令流和数据流的多倍性进行分类
- **指令流**：计算机执行的指令序列
- **数据流**：由指令流调用的数据序列
- **多倍性**：在系统受限的部件上，同时处于同一执行阶段的指令或数据的最大数目
- 4类计算机的基本结构
  - IS：指令流，DS：数据流，CS：控制流
  - CU：控制部件，PU：处理部件，MM和SM：存储器

# 计算机系统结构的分类

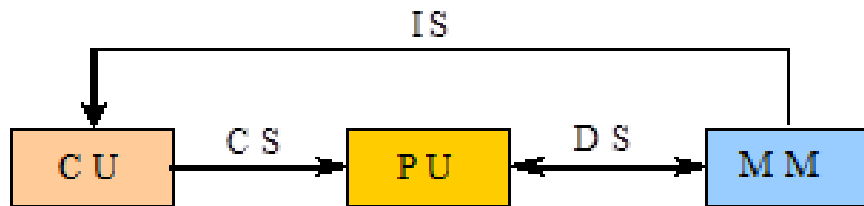
□ Flynn分类法把计算机系统的结构分为4类

- 单指令流单数据流(SISD, Single Instruction Stream Single Datastream)
- 单指令流多数据流(SIMD, Single Instruction Stream Multiple Datastream)
- 多指令流单数据流(MISD , Multiple Instruction Stream Single Datastream)
- 多指令流多数据流(MIMD, Multiple Instruction Stream Multiple Datastream)

# 计算机系统结构的分类

□ 单指令流单数据流SISD：传统的顺序处理计算机

- 指令部件一次只对一条指令译码并只对一个执行部件分配数据
- 可以有多个存储体和多个执行部件
- 可以是流水线，可以有多个功能部件，均由一个控制部件管理

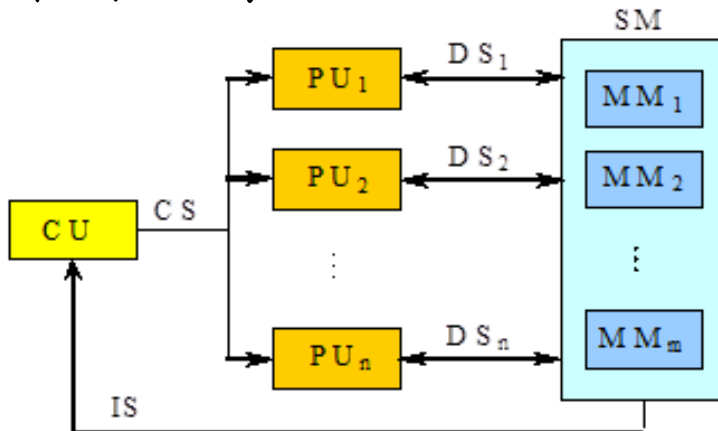


(a) SISD 计算机

# 计算机系统结构的分类

## □ 单指令流多数据流SIMD

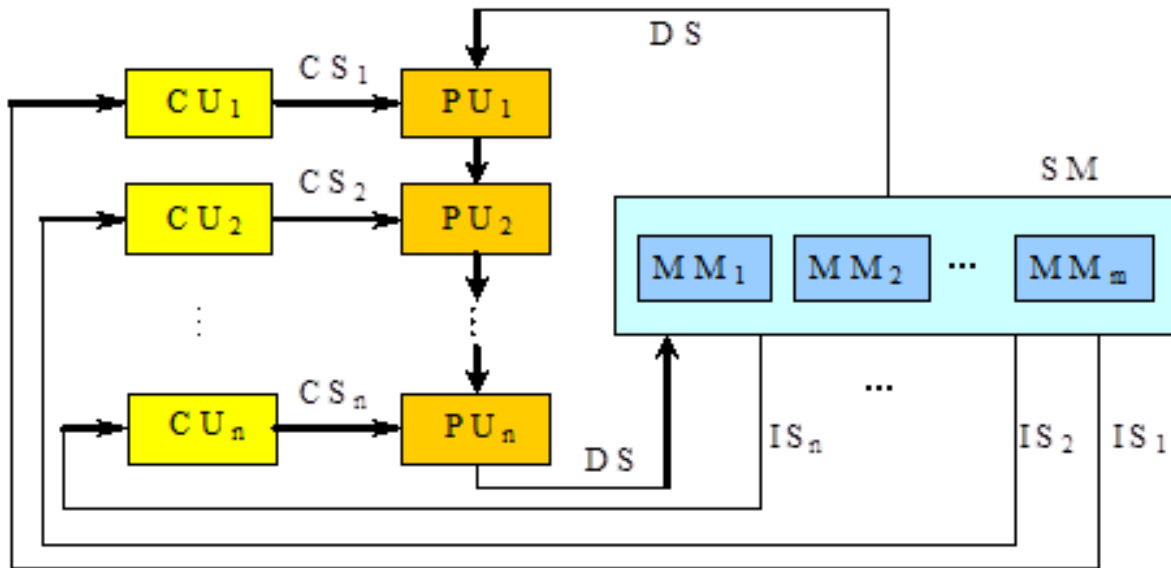
- 以并行处理机（阵列处理机）为代表
- 有多个处理单位PU, 均接受同一个控制部件发出的同一条指令管理控制
- 操作对象是来自不同数据流的数据组



# 计算机系统结构的分类

## □ 多指令流单数据流MISD

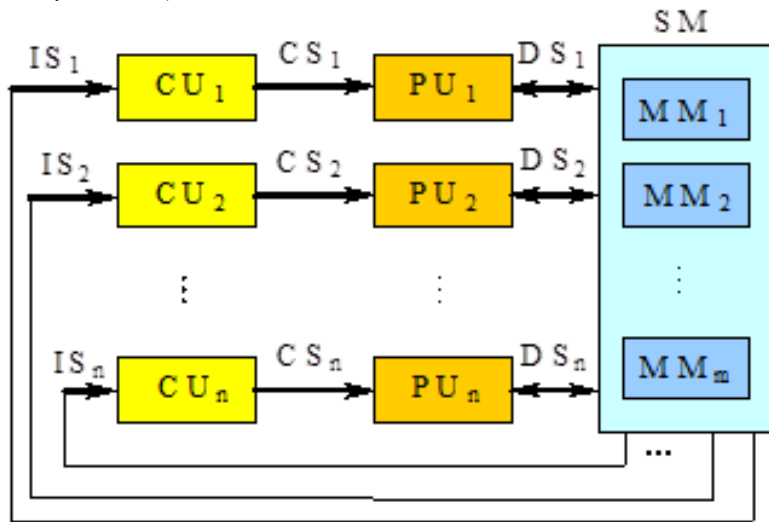
➤ MISD是一种人为的划分，目前没有实际的机器



# 计算机系统结构的分类

## □ 多指令流多数据流MIMD

- 大多数多处理机系统和多计算机系统可以归为该类
- 多处理机之间有相互作用，所有来自不同处理机的数据可能共享同一个空间



# 计算机系统结构的分类

□ 流水线处理机该如何划分？

➤ 把标量流水线处理机划入SISD

➤ 把向量流水线处理机划入SIMD

□ 优点：反映大多数并行方式和结构特点

□ 缺点：不能反映工作原理上的差别



# 计算机系统结构的分类

## □ Kuck (库克) 分类法

- 用指令流和执行流 (Execution Stream) 及其多倍性来描述计算机系统总控制器的结构特征
- 单指令流单执行流 (SISE) —— 典型的单处理机系统
- 单指令流多执行流 (SIME) —— 带多操作部件的处理机
- 多指令流单执行流 (MISE) —— 带指令级多道程序的单处理机
- 多指令流多执行流 (MIME) —— 典型的多处理机系统

# 计算机系统结构的分类

## □ 冯氏分类法

- 冯泽云提出用系统最大并行度对计算机系统进行分类
- 最大并行度 $P_m$ ：计算机系统在单位时间内能够处理的最大二进制位数
- 若设每一个时钟周期 $\Delta t_i$ 内能同时处理的二进制位数为 $P_i$ ，则 $T$ 个时钟周期内平均并行度为：

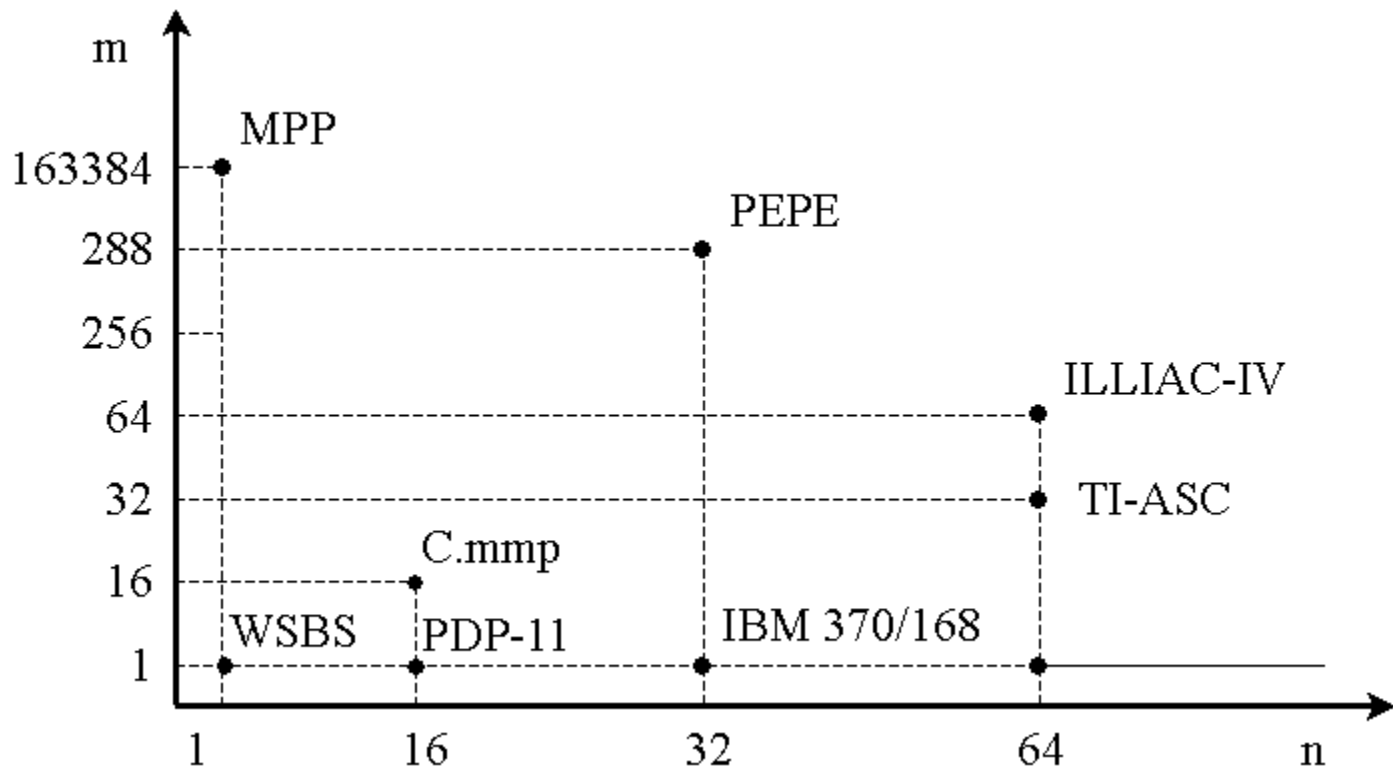
$$P_a = (\sum P_i) / T$$

- $T$ 个时钟周期内的平均利用率：

$$\mu = P_a / P_m = (\sum P_i) / (T \times P_m)$$

# 计算机系统结构的分类

□ 用最大并行度对计算机系统结构进行分类



# 计算机系统结构的分类

□ 按计算机对数据处理方式，则Pm值有下列4种类型：

- 字串位串 (WSBS) --  $n=1, m=1$ ，如位串行计算机
- 字串位并 (WSBP) --  $n>1, m=1$ ，如大多数传统的单处理机
- 字并位串 (WPBS) --  $n=1, m>1$ ，如某些相联处理机及阵列处理机
- 字并位并 (WPBP) --  $n>1, m>1$ ，如某些相联处理机，大多数阵列处理机及多处理机

# 计算机系统结构的分类

## □ Handler 分类法

- 根据并行度和流水线提出了另一种分类法
- 把计算机硬件结构分成三个层次
  - 程序控制部件(PCU)的概述 $k$
  - 算术逻辑部件(ALU)或处理部件(PE)的个数 $d$
  - 每个算术逻辑部件所包含基本逻辑线路(BLC)套数 $w$
- 一个计算机系统可描述为:
  - $t(\text{体系型号}) = \langle k, d, w \rangle$

# 计算机系统结构的分类

➤ 分别考虑各层并行度和流水处理程度：

- $T(\text{体系型号}) = \langle k \times k', d \times d', w \times w' \rangle$

- $k$  —— 处理控制器PCU的数目；

- $k'$  —— 可组成流水线的PCU数目；

- $d$  —— 每个PCU所控制的ALU(或PE)数目；

- $d'$  —— 可组成流水线的ALU数目；

- $w$  —— ALU或PE的字长；

- $w'$  —— 在所有ALU或一个PE中的流水段数目

# 计算机系统结构的分类

□ 【例】 CRAY-1计算机有1个CPU，12个相当于ALU或PE的处理部件，最多可以实现8级流水线。字长为64位，可以实现1~14位流水线处理。所以CRAY-1的体系结构可表示为：

$$T(\text{CRAY-1}) = \langle 1, 12 \times 8, 64 \times (1 \sim 14) \rangle$$

# 计算机系统结构的分类

□ 计算机系统结构领域中，有两个突出变化：

- 大中型计算机系统结构技术已用在个人计算机/工作站
- 一些高性能计算机是由多个微处理器构成的并行机器

□ 按照计算机的**性能**和**应用特征**，计算机分为：

- 桌面计算机
  - 个人计算机
  - 工作站
- 服务器型计算机
- 嵌入式计算机



# 计算机系统结构的分类

## □ 桌面计算机

- 个人计算机：为一个用户提供良好的计算性能和较低成本的工作环境
- 工作站：完整人机交互界面、图形处理和较高计算性能，大容量内存和硬盘，I/O和网络功能完善，使用多任务多用户操作系统的小型通用个人化计算机系统
- 桌面计算机典型的系统结构：
  - 以总线形式互连，构成计算机系统
  - SISD、MIMD(SMP)

# 计算机系统结构的分类

## □ 服务器型计算机

- 为网上客户机在网络环境下提供共享资源的高性能计算机
- 具有高可靠性、高性能、高吞吐能力、大内存容量
- 具备强大的网络功能和友好的人机界面
- 服务器是网络中的核心设备，其可靠性是关键
- 服务器需要具备容错能力和安全保护能力
  - 采用ECC内存、RAID技术、热插拔技术、冗余电源、冗余风扇、机箱锁、口令保护等方法

# 计算机系统结构的分类

- 服务器应具有**可扩展性**
- 能够在计算能力、存储器容量、存储系统以及I / O带宽等方面进行升级是至关重要的
- 网络吞吐率的整体性能以每分钟处理的事务数或每秒所提供的页面数来衡量
  - 用单位时间处理的请求数目表示整体效率和成本效率对大多数服务器来说更为关键
- 服务器通常采用多处理器/机结构，极大地提高了计算能力

# 计算机系统结构的分类

## ➤ 服务器按规模可分为：

- 大型服务器（企业级或计算中心级）
- 中型服务器（部门级）
- 小型服务器（基层工作组级）
- 入门级服务器

## ➤ 按服务器的外形与结构来分：

- 塔式服务器：价廉，用作入门、工作组服务器
- 机柜式服务器：大、中型服务器

# 计算机系统结构的分类

## □ 嵌入式计算机

➤ 计算机作为应用产品的核心控制部件，隐藏在各种装置、设备和系统中

- 硬件：嵌入式微处理器、存储器、I/O接口和外设部件

- 软件：实时和多任务操作系统软件和应用软件

➤ 核心是嵌入式微处理器

# 并行计算机简介

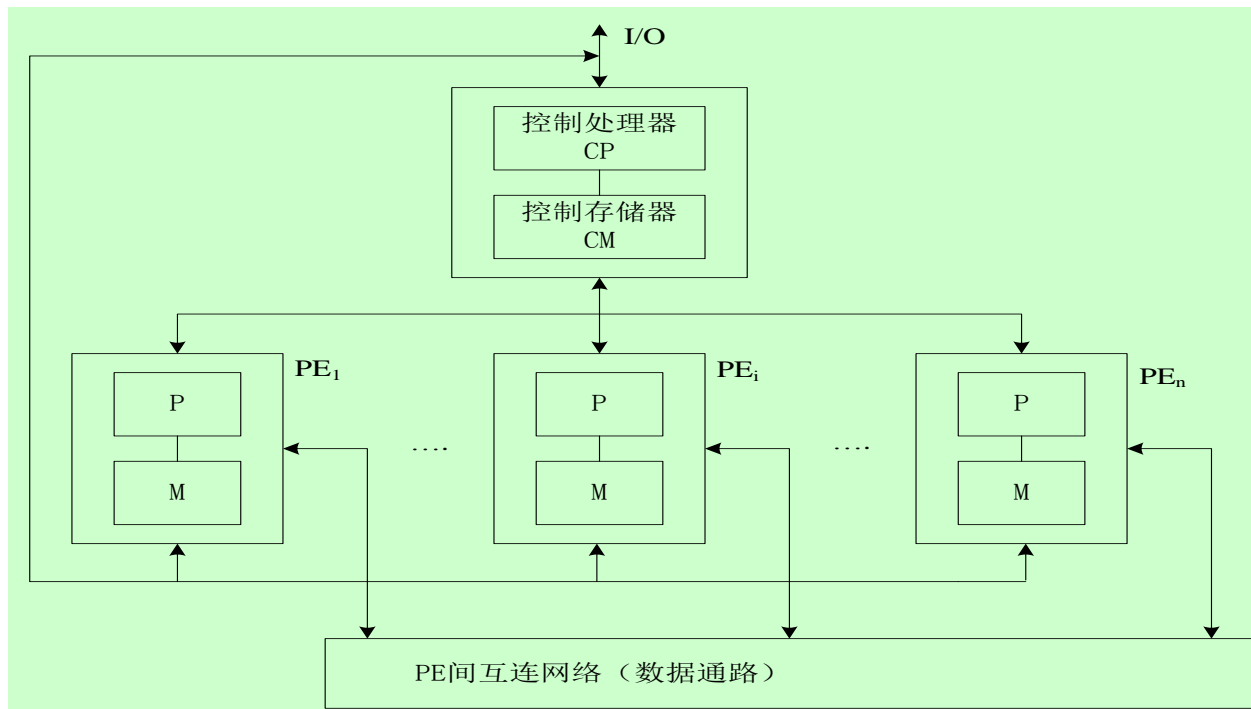
□ 按照并行计算机的系统结构，可分为：

- 流水线向量处理机
- 阵列计算机
- 多处理器/机系统
- 多计算机系统
  - 大规模并行处理系统MPP
  - 集群

# 并行计算机简介

## □ 阵列计算机系统

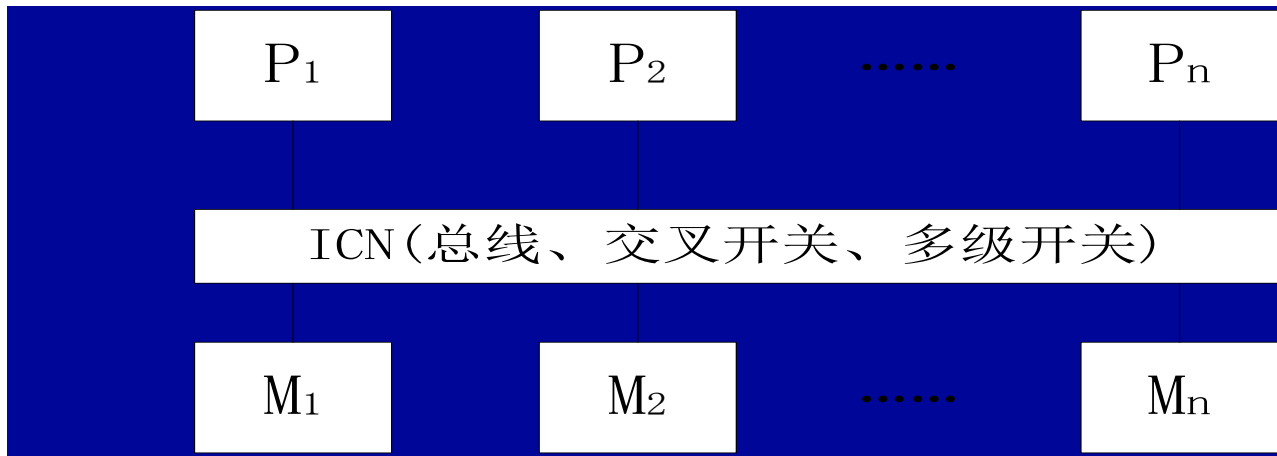
➤ 适合于处理向量类运算的多处理器/机



# 并行计算机简介

## □ 多处理器/机系统

- 由多个处理器/机及存储器模块构成的并行计算机
- 最普遍的多处理器组织方式是**对称多处理器SMP**



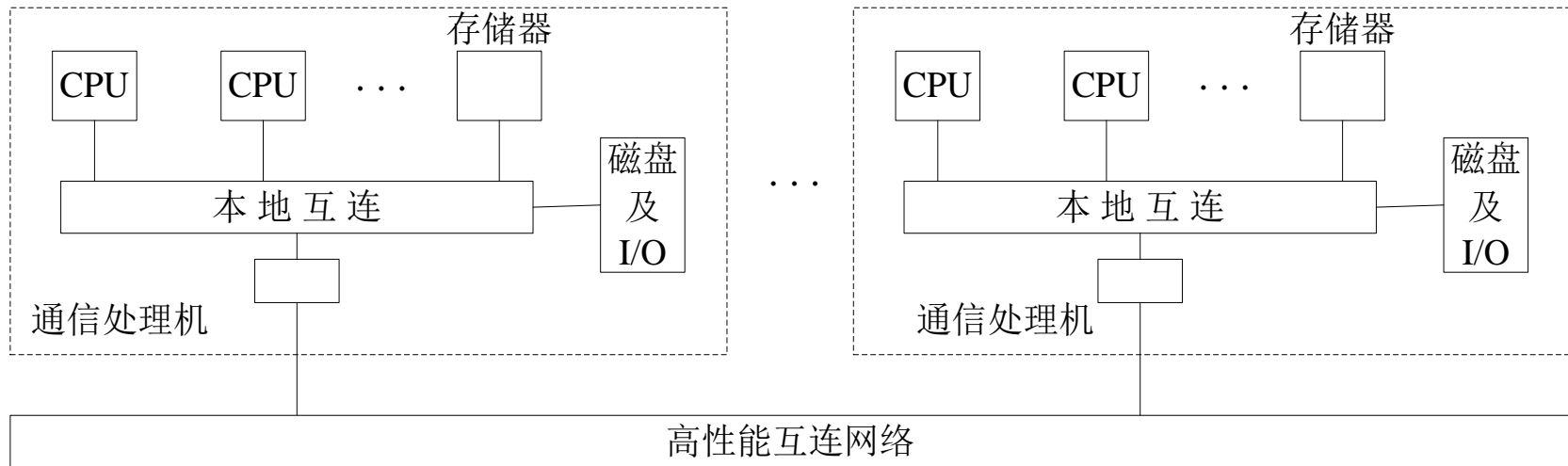


# 并行计算机简介

## □ 多计算机系统

➤ 由多个计算机构成的并行计算机

- 每个计算机只能访问自己内部的私有存储器
- 计算机之间的通信通过消息传递方式进行



# 并行计算机简介

- 大规模并行处理系统MPP (Massively Parallel Processing) 属于多计算机系统
  - 由成百上千个结点以高带宽低时延的专有网络互联构成大规模计算机系统
- 集群系统 (cluster) 也是多计算机系统
  - 由一组完整的计算机通过专用网络或局域网互连而成，作为统一的计算资源一起工作，并能产生像一台机器在工作的印象

# 内容小结

## □ 系统结构中并行性的发展

- 并行性的概念及等级（重点）
- 实现并行性的方式、途径及发展
- 系统结构的分类及并行计算机简介

## □ 知识要点

并行性	时间重叠	资源重复
资源共享	耦合度	紧密耦合
松散耦合	异构型多处理机	同构型多处理机

# 练习题

- 1. 分别从执行程序的角度和处理数据的角度来看，计算机系统中并行性等级从低到高可分为哪几级？
- 2. 提高计算机系统并行性的方法有哪些？

# Thank You !

[zhaofangbupt@163.com](mailto:zhaofangbupt@163.com)



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS