



河海大学

并行体系结构

黄倩

huangqian@hhu.edu.cn

勤学楼4203



内容提要

- 基本概念
- 超长指令字处理机
- 多线程与超线程处理机
- 向量处理机
- 多处理机
- 机群系统



基本概念

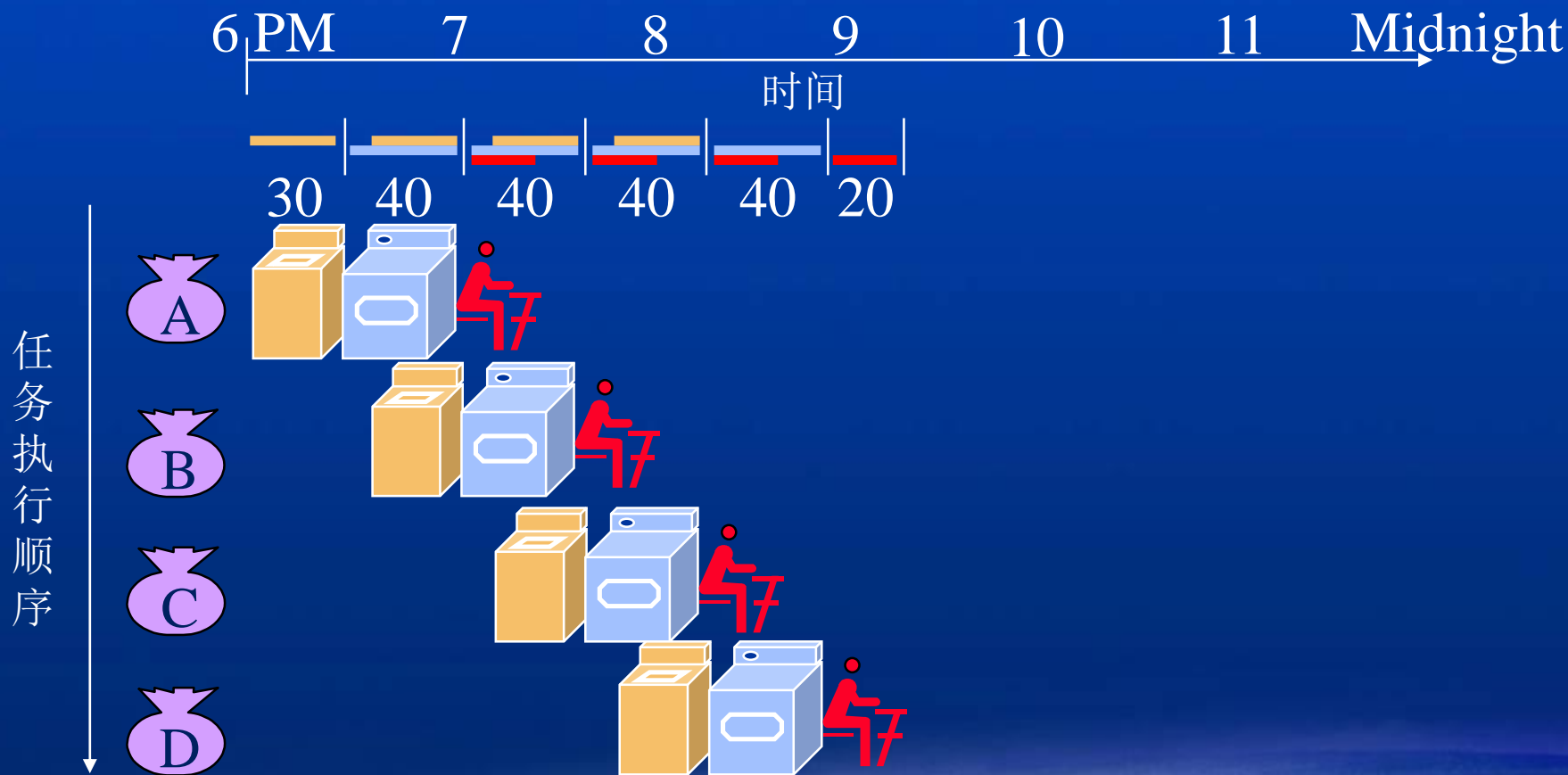
- 并行性 (Parallelism)
 - 并行性是指计算机系统具有可以同时进行运算或操作的特性，它包括并发性与同时性两种含义。
 - 并发性是指两个或两个以上的事件在同一时间间隔发生（如分时交替执行、重叠执行等）。
 - 同时性是指两个或两个以上的事件在同一时刻发生。



基本概念

- 并行性 (Parallelism)

- 第5章洗衣服的例子





基本概念

- 并行性 (Parallelism)

- 指令重叠执行的例子

指令执行过程：取指令、分析指令、执行指令 t

- 一次重叠



若每一小步时间都为 t ，则执行 n 条指令的总时间是 $(2n+1)t$

- 二次重叠



若每一小步时间都为 t ，则执行 n 条指令的总时间是 $(n+2)t$



基本概念

- 并行性等级的划分

- 从数据处理的角度划分（等级由低到高）

- 字串位串：同时只对一个字的一位进行处理。这是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。这里已开始出现并行性。
- 字并位串：同时对许多字的同一位进行处理。这种方式有较高的并行性。
- 全并行：同时对许多字的全部位进行处理。这是最高一级的并行。



基本概念

- 并行性等级的划分

- 从执行程序的角度划分（等级由低到高）

- 指令内部并行：一条指令执行时各微操作之间的并行。
 - 指令级并行：并行执行两条或多条指令。
 - 任务级或过程级并行：并行执行两个以上过程或任务(程序段)。
 - 作业或程序级并行：并行执行两个以上作业或程序。

- 并行处理领域

- 体系结构、硬件、软件、算法、语言等多方面综合研究的领域
 - 并行性等级达到一定程度，可认为进入并行处理领域



基本概念

- 计算机并行处理的主要形式

- 时间重叠

- 也称时间并行。在并行概念中引入时间因素，让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。

- 实现方式：采用流水处理部件。

- 资源重复

- 也称空间并行。在并行概念中引入空间因素，通过重复设置硬件资源，大幅度提高计算机系统的性能。如单处理机系统中的多种多个计算部件、多处理机系统、多计算机系统。



基本概念

- 计算机并行处理的主要形式
 - 时间重叠+资源重复
 - 同时运用时间并行技术和空间并行技术，这种方式在计算机系统中得到广泛应用，成为并行性主流技术。例如，奔腾CPU采用了超标量流水技术，在一个机器周期中同时执行两条指令，因而既具有时间并行性，又具有空间并行性。
 - 资源共享
 - 一种软件方法，使多个任务按一定时间顺序轮流使用同一套硬件设备
 - 既降低了成本，又提高了计算机设备的利用率



基本概念

- 并行性的发展

- 单处理机系统中并行性的发展

- 时间并行

- 起主导作用

- 实现方式：采用流水处理部件。（1）把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（“部件功能专门化”）；（2）把各部分执行过程在时间上重叠起来，使所有部件一次分工完成一组同样的工作。

- 空间并行

- 多体存储器、多操作部件

- 资源共享（软件方法）

- 用单处理机模拟多处理机，形成所谓虚拟机的概念。例如分时系统，在多终端情况下，每个终端上的用户都觉得自己有一台处理机



基本概念

- 并行性的发展
 - 单处理机系统中并行性的发展





基本概念

- 并行性的发展
 - 多处理机系统中并行性的发展
 - 软件工程中的耦合与内聚
 - 耦合：模块之间依赖程度的度量
 - 内聚：模块内部各成分之间相关程度的度量
 - 软件工程一般强调高内聚、低耦合



基本概念

- 并行性的发展

- 多处理机系统中并行性的发展

- 紧耦合系统、松耦合系统

- 耦合度：反映多处理机系统各机器之间物理连接的紧密程度与交互作用能力的强弱。

- 紧耦合系统：又称直接耦合系统，各处理机之间一般通过总线或高速开关实现互连，物理距离较近，具有较高的信息传输速率，可快速并行处理任务、可共享主存。

- 松耦合系统：又称间接耦合系统，各处理机之间一般是通过通道或通信线路实现互连，物理距离较远，相对来说信息传输速率较低，可信息交换、资源（外设）共享、并行/分布处理任务。



基本概念

- 并行性的发展

- 多处理机系统中并行性的发展

- 同构型系统、异构型系统

- 同构型系统：在多处理机系统中，各个处理机具有相同的功能、可完成相同的任务。随着硬件价格的降低，人们追求的目标是通过多处理机的并行处理来提高整个系统的速度。如高性能计算、应用服务器群。为了使并行处理的任务能在处理机之间随机地调度，就必须使各处理机具有同等功能，从而成为同构型多处理机系统。

- 异构型系统：在多处理机系统中，为了实现时间重叠，将处理功能分散给各专用处理机去完成（即处理机功能专用化）。各个处理机之间采用流水线技术协同工作，按时间重叠原理分别完成特定的功能/任务。



基本概念

• 并行性的发展

• 多处理机系统中并行性的发展

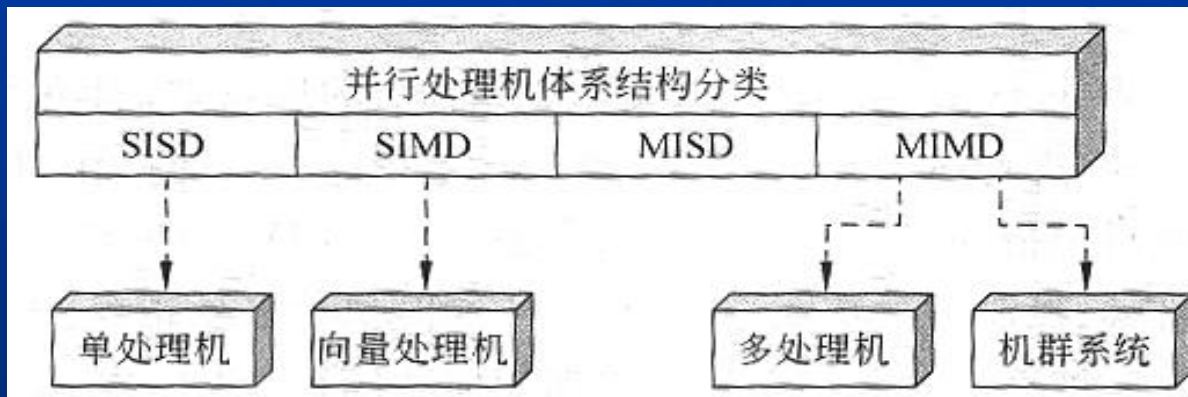
• 我国自行研制的几种并行计算机系统

- 2000: 神威号计算机（每秒3480亿次运算）
- 2004: 曙光4000A（每秒10万亿次运算，全球第10）
- 2010: 天河一号（每秒2.57千万亿次运算，全球第1）
- 2014: 天河二号（每秒33.86千万亿次运算，全球第1）
- 2015.4: 美国商务部禁止英特尔向中国4家国家超级计算机机构出售“至强”（XEON）芯片
- 2016: 神威·太湖之光（每秒93.02千万亿次运算，全球第1）
 - 国产申威处理器，源于美国DEC公司的Alpha 21164



基本概念

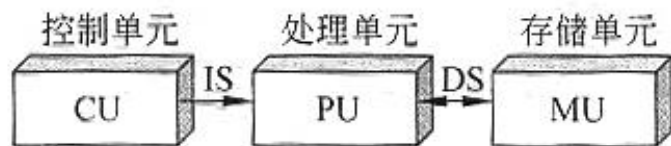
- 并行处理机的体系结构类型（格林分类法）
 - 单指令流单数据流SISD：传统单处理机系统
 - 单指令流多数据流SIMD：向量处理机
 - 多指令流单数据流MISD：没意义
 - 多指令流多数据流MIMD：多处理机系统、机群系统



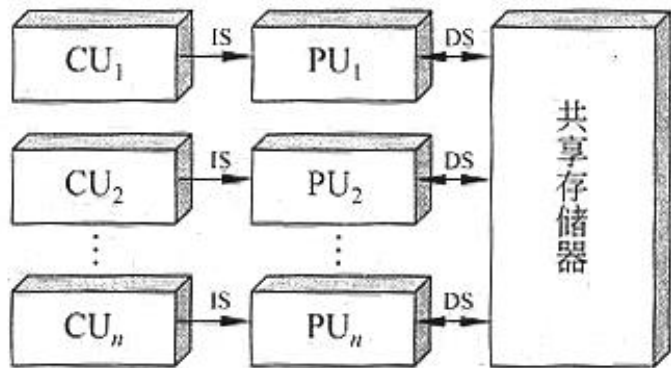


基本概念

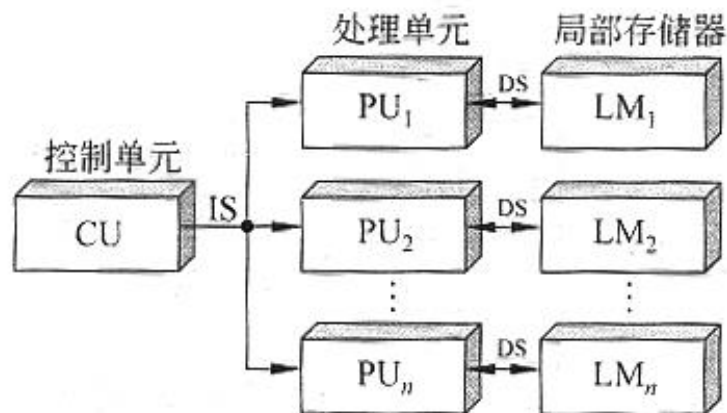
• 并行处理机的体系结构类型（格林分类法）



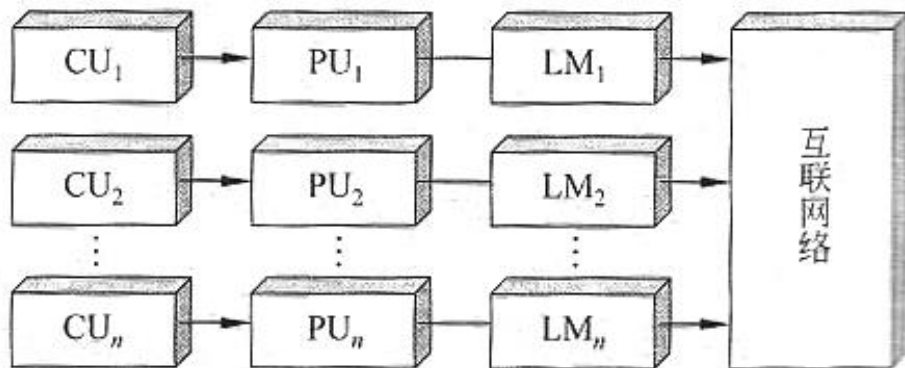
(a) SISD



(c) MIMD(共享存储器)



(b) SIMD(分布式存储器)



(d) MIMD(分布式存储器)



超长指令字处理机

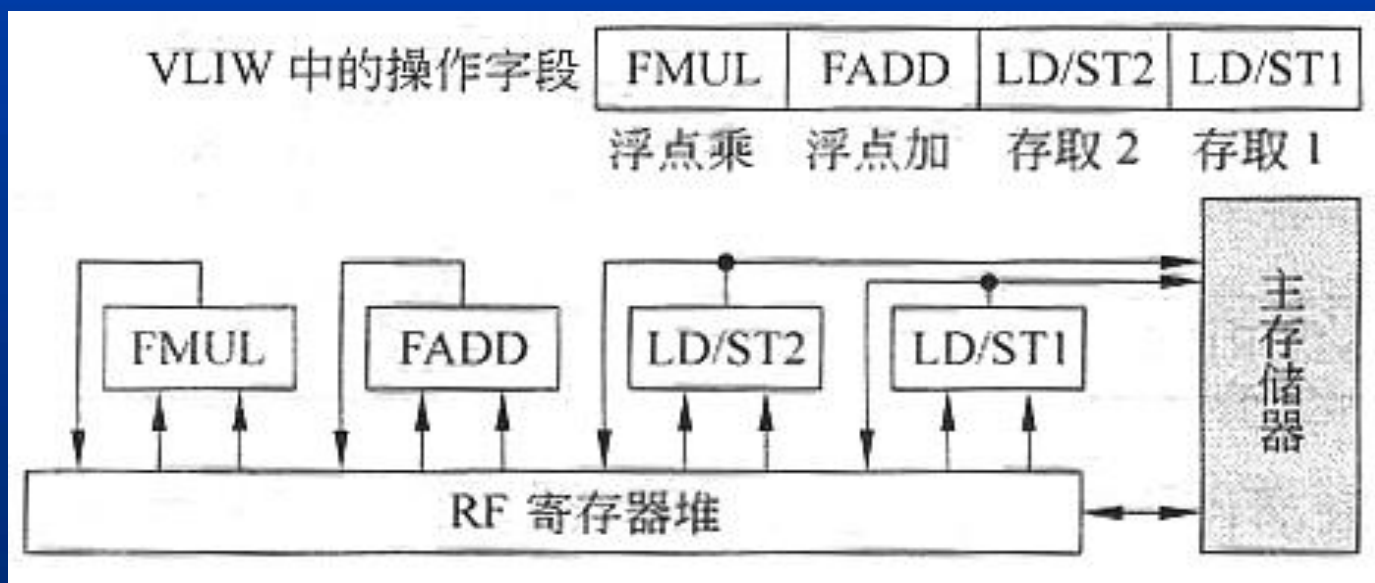
- **超长指令字: Very Long Instruction Word**
 - 由编译程序在编译时找出指令间潜在的并行性，进行适当调度安排，把多个能并行执行的操作组合在一起，成为一条具有多个操作段的超长指令。
 - 由这条超长指令去控制VLIW处理机中多个互相独立工作的功能部件，每个操作段控制一个功能部件，相当于同时执行多条指令。



超长指令字处理机

• 超长指令字处理机结构模型

- LD/ST1、LD/ST2：从主存储器存取数部件
- FADD：浮点加法处理部件
- FMUL：浮点乘法处理部件





超长指令字处理机

[例1] 假设要执行以下程序，数据都是浮点数。

$C = A + B$

$K = I + J$

$L = M - K$

$Q = C \times K$



串行指令序列：共14个周期

源代码	操作性质	所需周期
$C = A + B$	LOAD A	1
	LOAD B	1
	$C = A + B$	1
	STORE C	1
$K = I + J$	LOAD I	1
	LOAD J	1
	$K = I + J$	1
	STROE K	1
$L = M - K$	LOAD M	1
	$L = M - K$	1
	STORE L	1
$Q = C \times K$	$Q = C \times K$	2
	STORE Q	1



超长指令字处理机

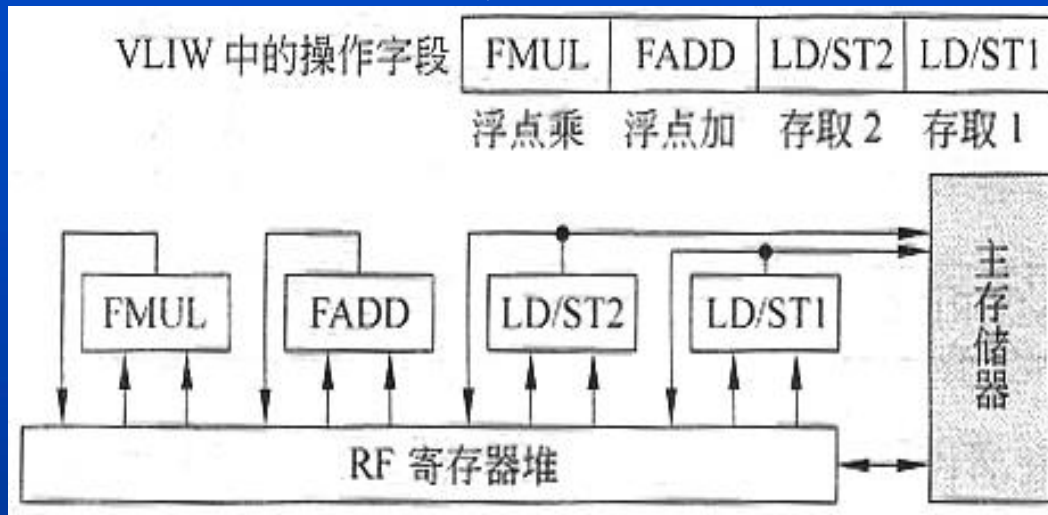
[例1] 假设要执行以下程序，数据都是浮点数。

$$C=A+B$$

$$K=I+J$$

$$L=M-K$$

$$Q=C \times K$$



周期	LD/ST1	LD/ST2	FADD	FMUL
周期1	LOAD A	LOAD B		
周期2	LOAD I	LOAD J	$C=A+B$	
周期3	LOAD M	STORE C	$K=I+J$	
周期4		STORE K	$L=M-K$	$Q=C \times K$
周期5		STORE L		
周期6	STROE Q			



超长指令字处理机

- 超长指令字处理机的主要特点

- 超长指令字的生成是由编译器来完成的，由于将串行的操作序列合并为可并行执行的指令序列，以最大限度实现操作并行性。
- 单一的控制流，只有一个控制器，每个时钟周期启动一条长指令。
- 超长指令字被分成多个控制字段，每个字段直接独立地控制每个功能部件。
- 含有大量的数据通路和功能部件。



多线程与超线程处理机

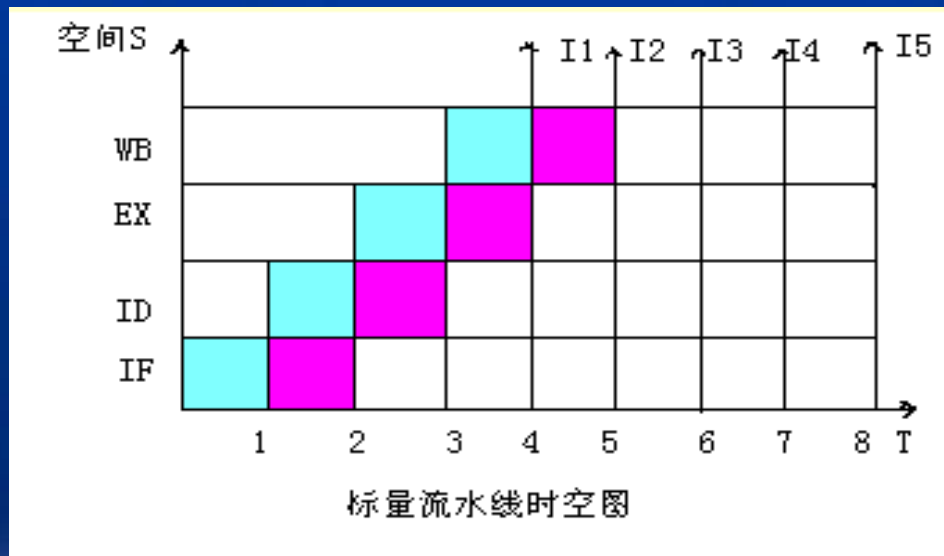
• 指令级并行



一条指令的执行划分为若干过程段，采用流水线实现时间重叠。

由于资源相关、数据相关、控制相关可能会造成空闲等待周期、断流。

计算机组成与体系结构

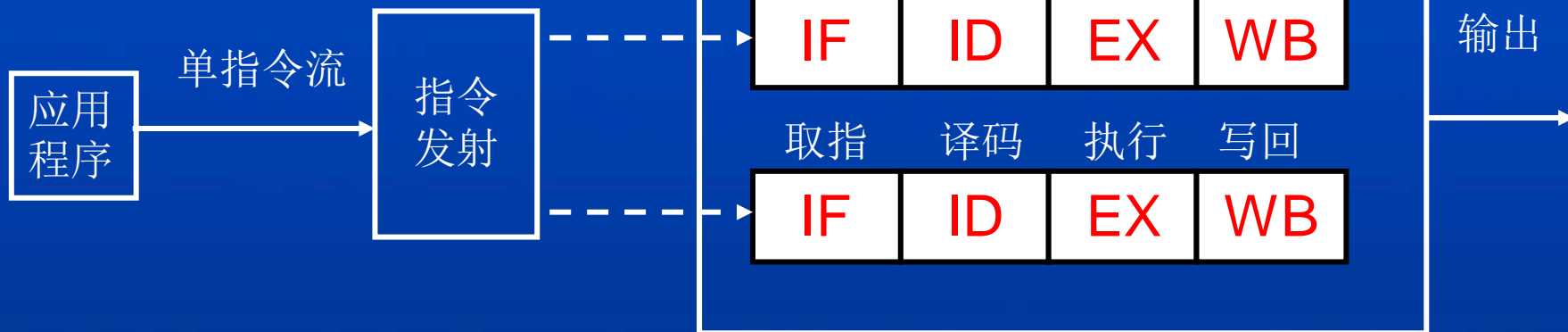


河海大学计算机与信息学院



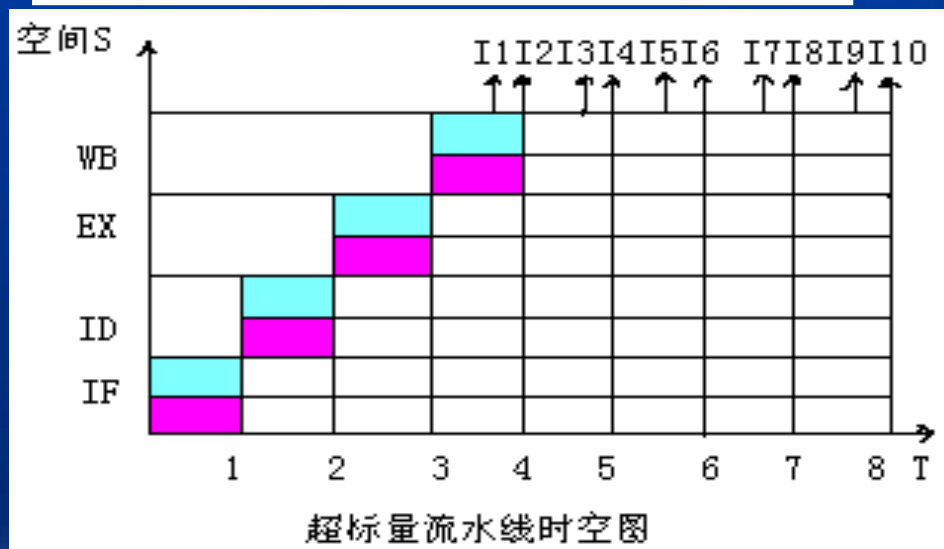
多线程与超线程处理机

• 指令级并行



单指令流向两条流水线发射指令。

由于资源相关、数据相关、控制相关可能会造成空闲等待周期、断流。

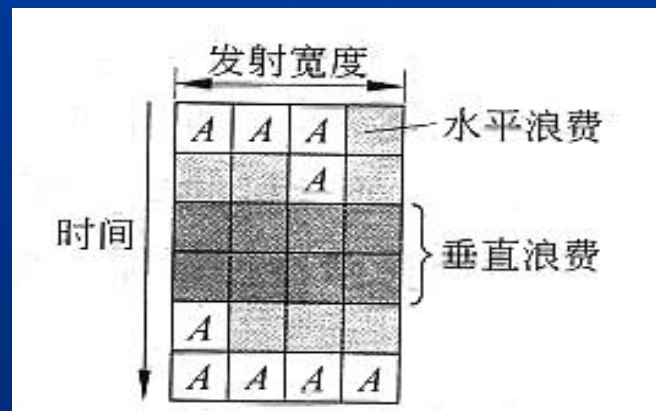
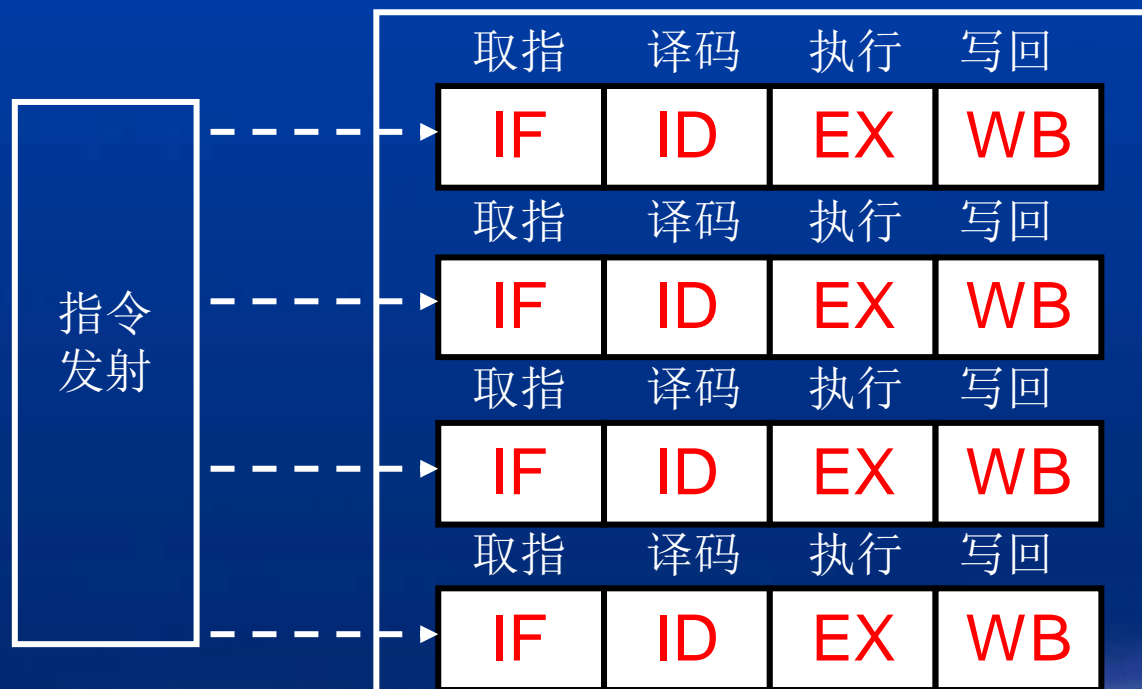




多线程与超线程处理机

• 指令级并行

- 具有4条流水线的超标量流水线处理机
- 图理解：横向看作是针对某个过程段（如取指IF，共4个格子；纵向是指每个流水线时钟周期。

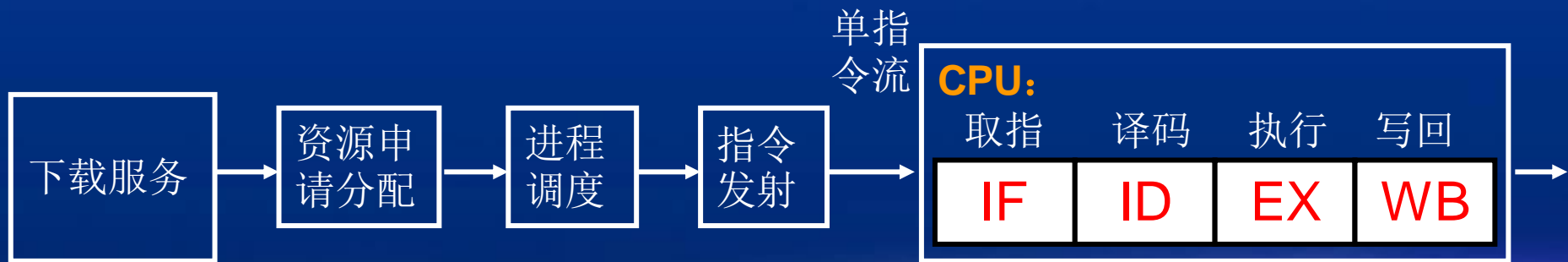




多线程与超线程处理机

• 进程

- 程序的一次执行称为进程。程序是静态概念，进程是动态概念。
- 每个进程拥有私有的资源，如内存/CACHE空间、数据寄存器等。
- 进程的处理过程包括资源申请分配、CPU调度执行两个部分。

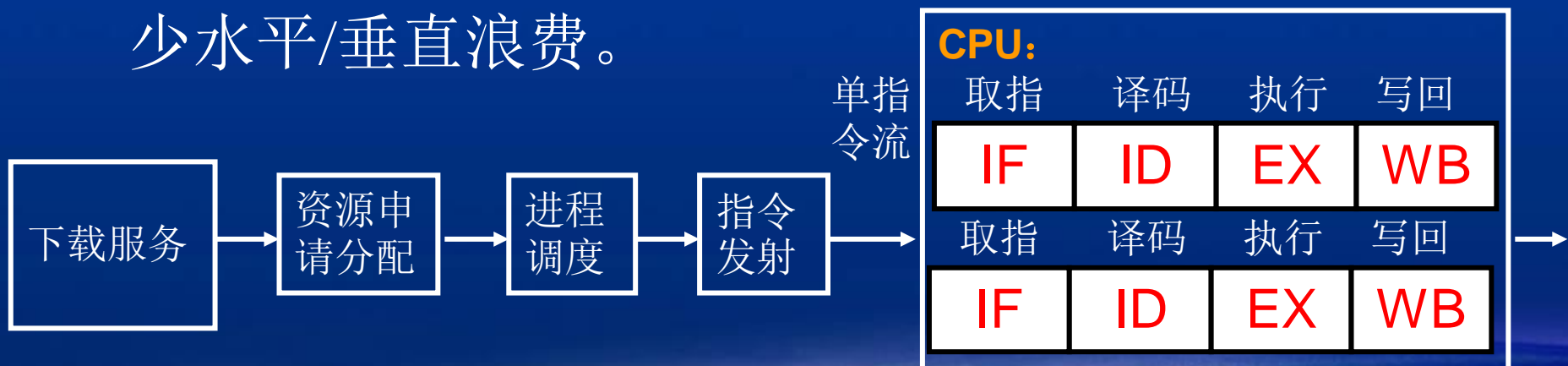




多线程与超线程处理机

• 进程

- 以进程为基本执行单位进行资源分配、调度执行。
- 同时为多进程服务时，系统开销大。假设同时为甲、乙、丙3个用户提供下载服务，则对应应有甲、乙、丙3个进程，都要占用系统资源。
- 在某个进程（如乙进程）被调度执行时，该时间段内只能发射乙的指令，不管有多少条流水线、不管有多少水平/垂直浪费。

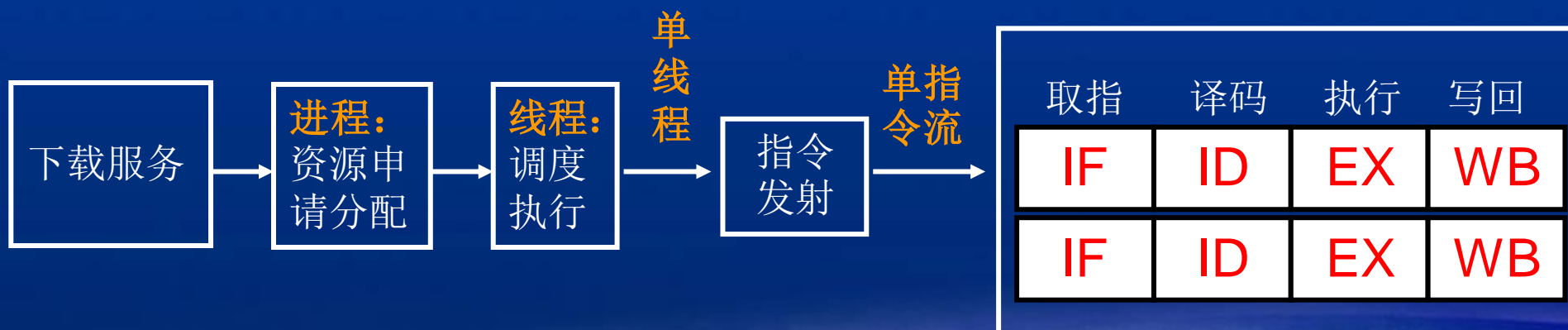




多线程与超线程处理机

• 线程

- 线程是进程概念的延伸，是比进程还小的、基本执行单位，一个进程至少包含一个线程。
- 以进程为单位分配资源，进程下所有的线程共享使用进程申请的资源。
- 以线程为单位调度，每个线程可为一个用户服务。

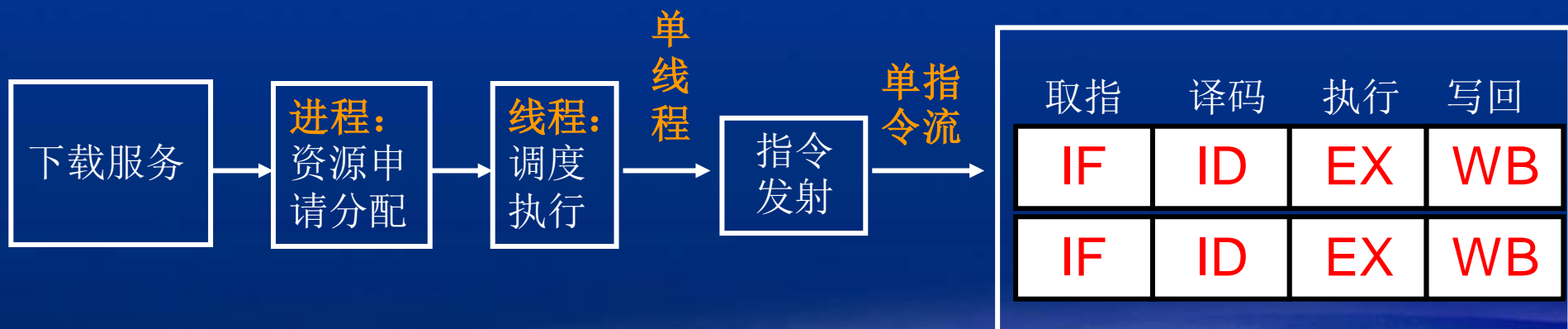




多线程与超线程处理机

• 线程级并行

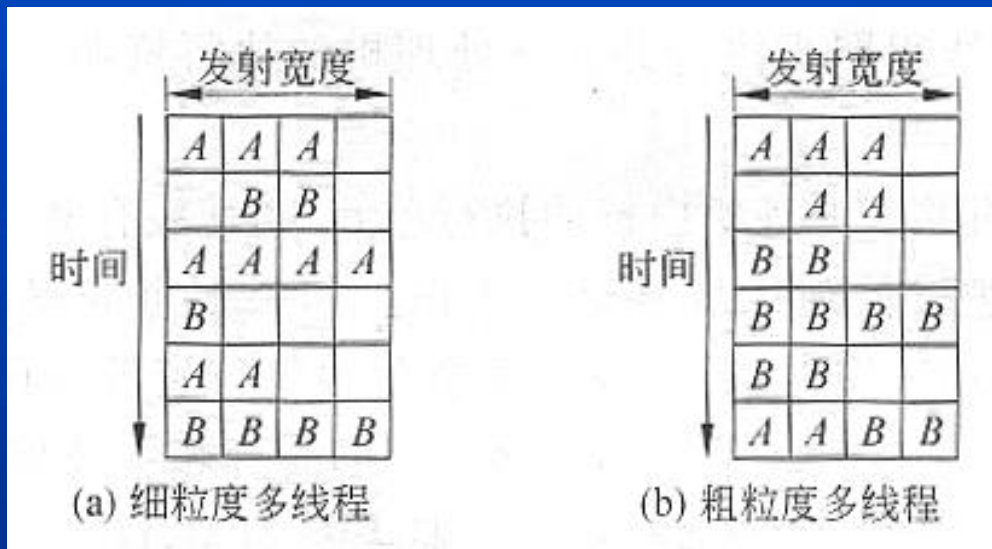
- 当正在被调度执行的线程的指令流发生严重阻塞或断流时，可以调度其他线程执行，而不需要释放资源、重新申请资源，减少系统开销。
- 由于各个线程相互独立，且可交替调度执行，从而有效提高处理机执行单元的利用率。
- 一个线程被调度执行时间内，还是单指令流发射。





多线程与超线程处理机

- 线程级并行



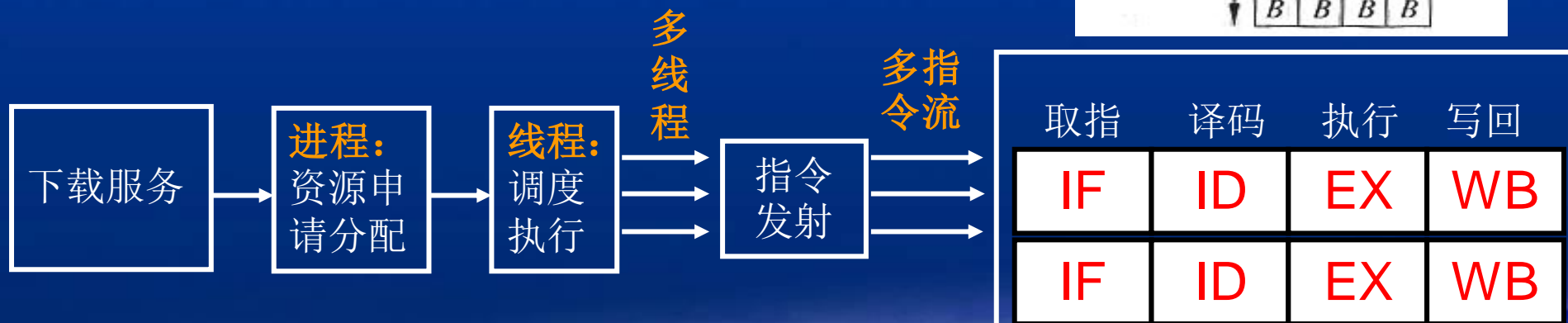
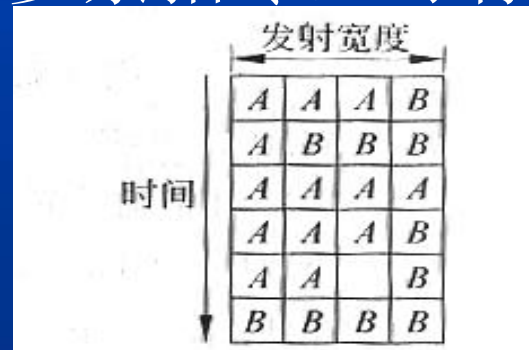
- 细粒度多线程：多线程交替执行。
- 粗粒度多线程：遇到代价较高的流水线阻塞（如长延迟时间、断流）时，才由处理机硬件进行线程切换。



多线程与超线程处理机

- 同时多线程结构（常称超线程技术）

- 同时调度多个线程执行，即多条指令流，按一定的策略往超标量流水线中交替/混合发射指令。流水线处理机可以同时处理来自不同线程的多条指令，可有效避免单指令流中的相关问题





向量处理机

• 向量

- 向量指一组标量。例如，数组 $A = (a_1, a_2, a_3, \dots, a_n)$ ，其中每一个分量元素 a_i 就是一个标量，而 A 称为向量，它由一组标量组成。
- 向量运算举例： $D = A \times (B + C)$ ，其中 A 、 B 、 C 、 D 都是长度为 n 的向量。
- 采用标量处理机计算：一个循环次数为 n 的循环程序，每次循环处理 $k_i = b_i + c_i$ ； $d_i = a_i \times k_i$ 。
- 采用向量处理机计算（使用向量指令）：
 - `ADDV B, C, K`
 - `MULV A, K, D`



向量处理机

• 向量处理方法

横向处理方法:

$$b1+c1 \rightarrow k1$$

$$k1 \times a1 \rightarrow d1$$

$$b2+c2 \rightarrow k2$$

$$k2 \times a2 \rightarrow d2$$

.....

$$b_n+c_n \rightarrow k_n$$

$$k_n \times a_n \rightarrow d_n$$

①数据相关;

②加法、乘法转换n次;

计算机组成与体系结构

纵向处理方法:

$$b1+c1 \rightarrow k1$$

$$b2+c2 \rightarrow k2$$

.....

$$b_n+c_n \rightarrow k_n$$

$$k1 \times a1 \rightarrow d1$$

$$k2 \times a2 \rightarrow d2$$

.....

$$k_n \times a_n \rightarrow d_n$$

①数据无关;

②加法、乘法转换1次;

③适合流水线处理;

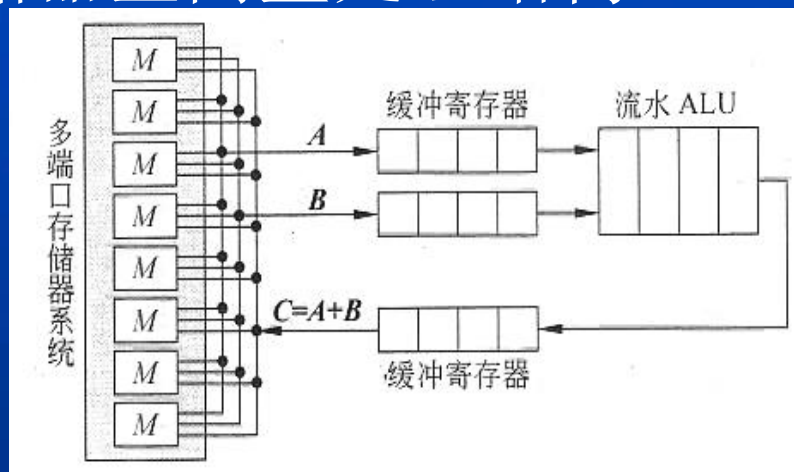
河海大学计算机与信息学院



向量处理器

- 向量处理机的结构

- 存储器-存储器型向量处理结构



- 思想：采用流水线技术实现向量处理，向量的分量源源不断地进入流水线，各个分量的处理时间重叠，整体上提高向量的处理速度。

- 流水线满载以后，每个时钟周期向流水线发射一组分量，每个时钟周期输出一组分量的运算结果

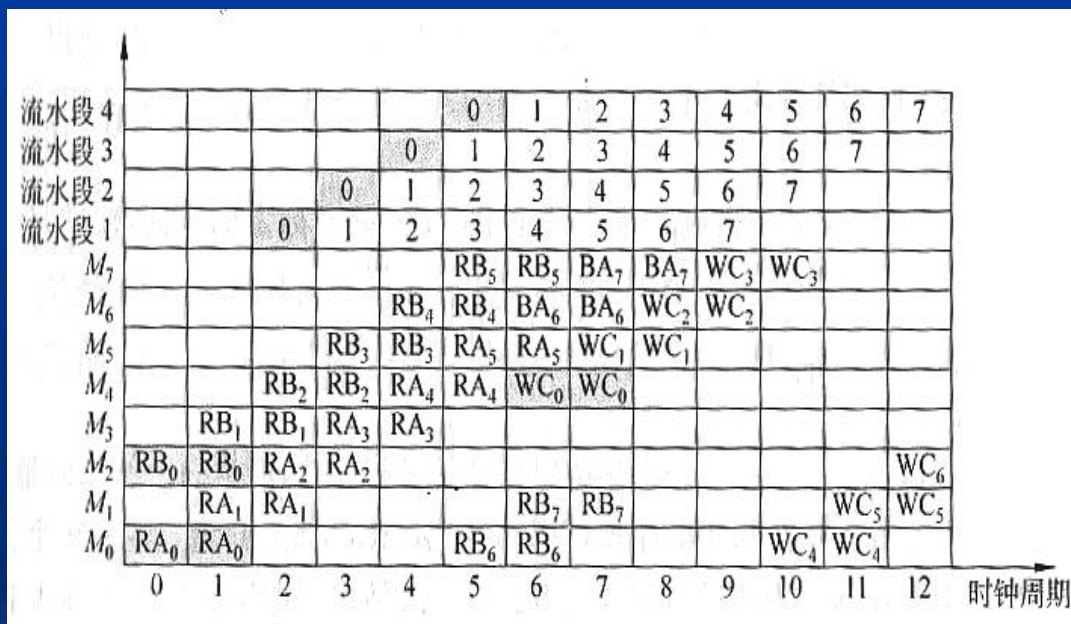
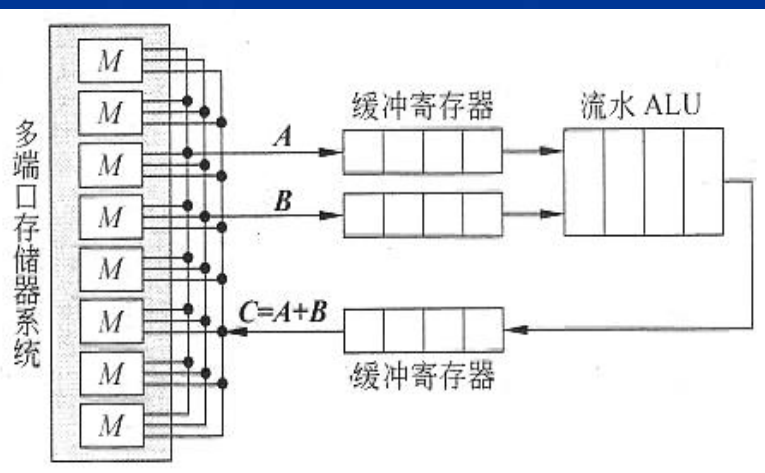


向量处理器

• 向量处理机的结构

• 存储器-存储器型向量处理结构

- 由于每个时钟周期都需要从存储器中取A、B，并保存结果C，因此必须采用多存储体多端口存取，使每个时钟周期的取A、取B、存C针对不同的存储体，且从不同的访问端口进行操作。



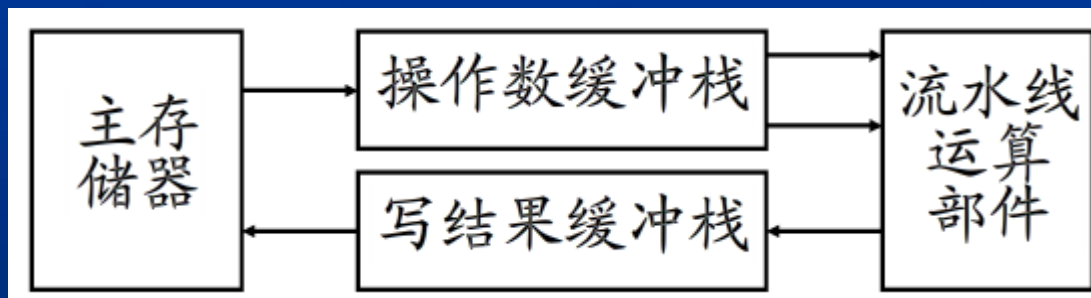


向量处理机

- 向量处理机的结构

- 存储器-存储器型向量处理结构

- 缓冲栈的作用：解决访问存储器冲突。
- 提速方法：多体交叉、并行访问
- 优点：硬件结构简单、造价低
- 缺点：速度相对较慢



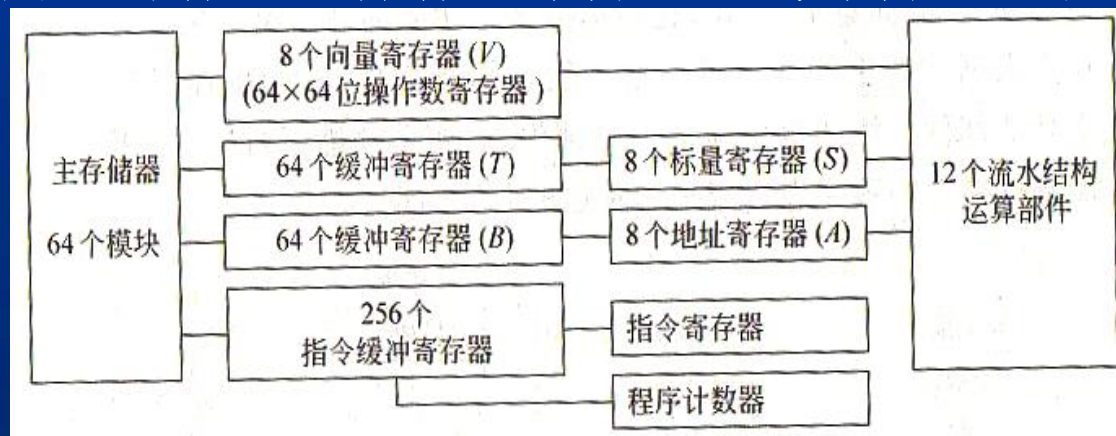


向量处理机

• 向量处理机的结构

• 寄存器-寄存器型向量处理结构

- 思想：使操作数离处理器很近，以保证处理器一直处于忙碌状态。
- 做法：降低对存储系统访问速度的要求，把缓冲栈改为向量寄存器。
- Cray-1：8个向量寄存器，其内容从主存储器成组读取。每个向量寄存器可保存64个分量，每个分量64位。





向量处理机

- 向量处理机的运算

- 向量平衡点 (VBP)

- 定义：为了使向量硬件设备和标量硬件设备的利用率相等，程序中向量代码所占的百分比。

机器型号	向量性能 Mflops	标量性能 Mflops	向量平衡点
Cray IS	85.0	9.8	0.90
Cray 2S	151.5	11.2	0.93
Cray X-MP	143.3	13.1	0.92
Cray Y-MP	201.6	17.0	0.92
Hitachi S820	737.3	17.8	0.98
NEC SX2	424.2	9.5	0.98
Fujitsu VP400	207.1	6.6	0.97



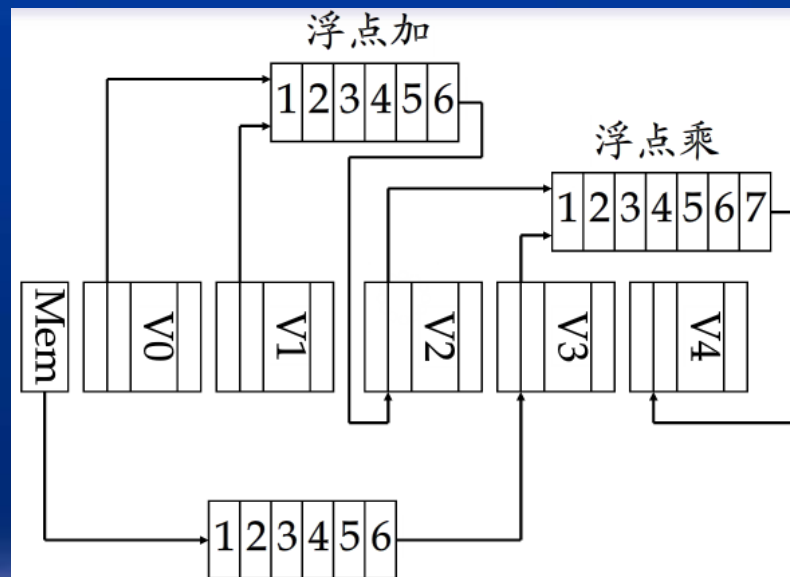
向量处理器

• 向量处理机的运算

• 向量链接技术

- 定义：前一条指令的结果寄存器可作为后继指令的操作数寄存器时，使多条存在数据相关的向量指令并行执行的技术。
- 例：如下三条指令中，1、2可同时执行，3可链接执行。

1: $V3 \leftarrow A$
2: $V2 \leftarrow V0 + V1$
3: $V4 \leftarrow V2 \times V3$





多处理机

- 多处理机系统的分类

- 多处理机系统由多个独立的处理机（CPU/cache；或私有存储器、或使用共享存储器）组成，每个处理机能够独立执行自己的程序、但没有独立的操作系统，都是在集中操作系统统一管理调度下协同完成作业/任务。
- 多处理机系统通常分类
 - 并行向量处理机PVP
 - 对称多处理机SMP（高性能小型机中常用）
 - 分布共享存储器多处理机DSM
 - 大规模并行处理机MPP



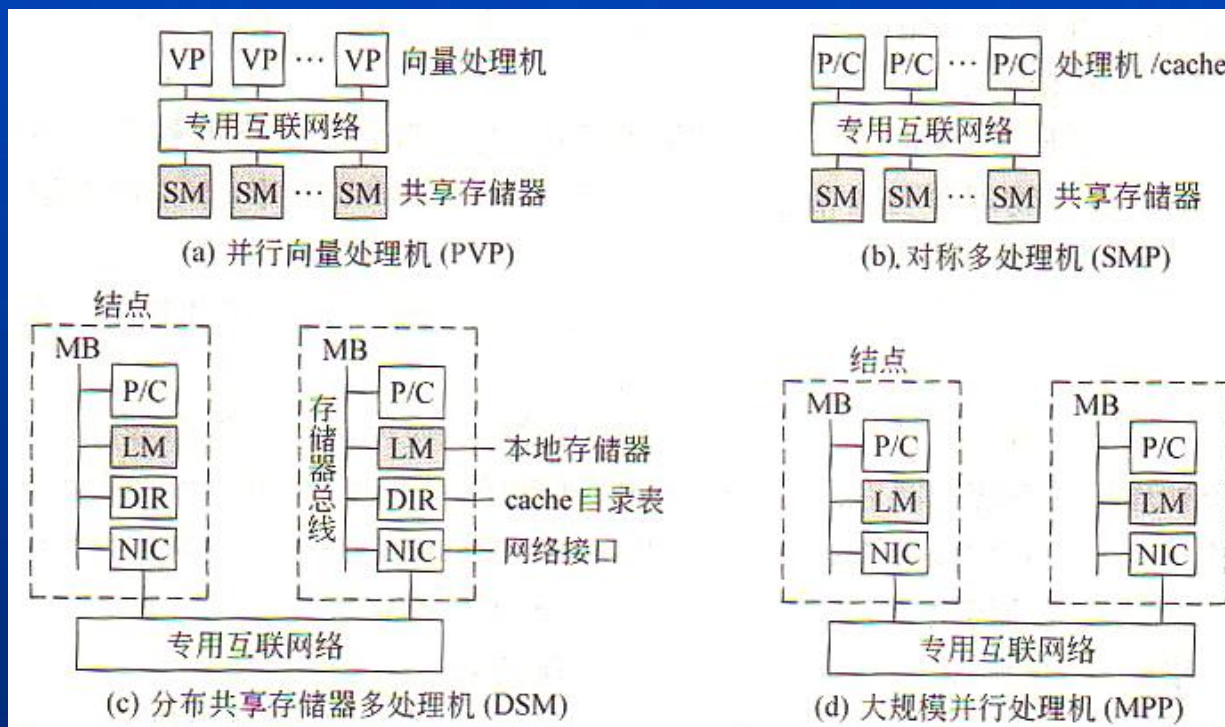
多处理机

• 多处理机系统的分类

共享存储器方式

共享存储器方式

每台处理机功能完全相同、同构系统



物理分布、逻辑共享存储器方式

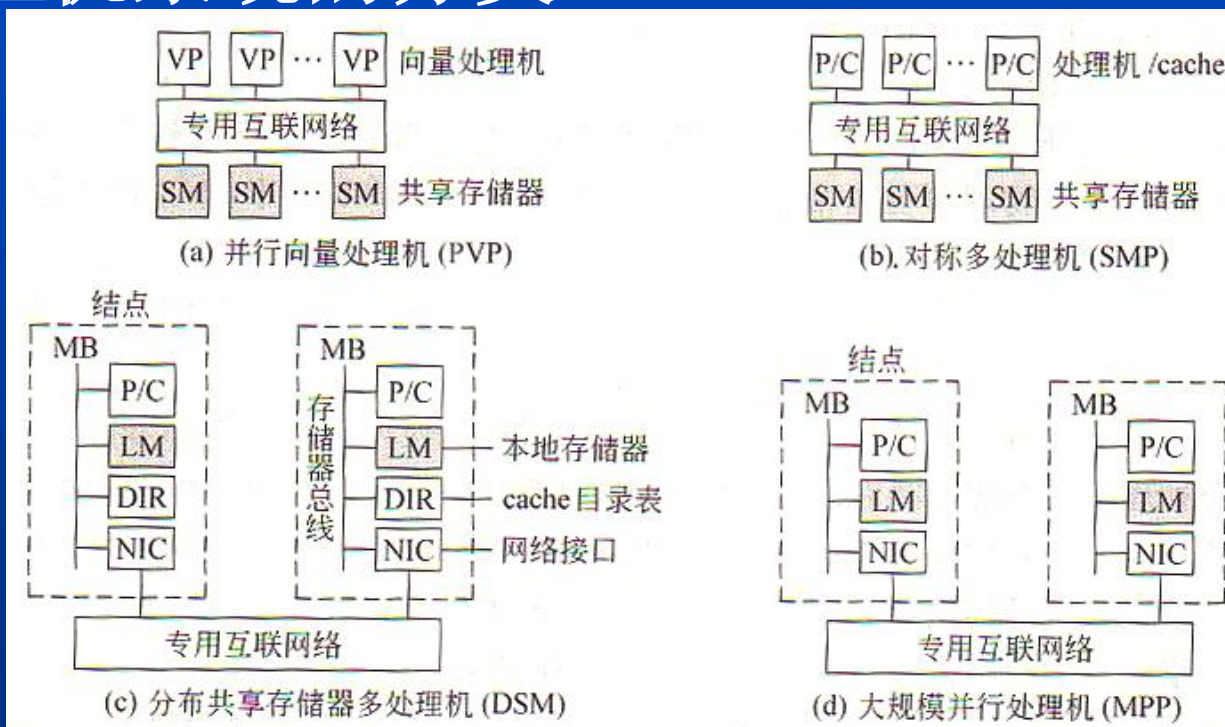
分布存储器方式

每台处理机功能可不同、异构系统



多处理机

• 多处理机系统的分类



- 专用互连网络：总线、高速开关、专用网络
 - 由集中操作系统统一管理调度处理机，协同完成作业/任务
 - 一个计算机系统的范畴；不同于计算机网络系统、机群系统
- 计算机组成与体系结构



多处理机

• SMP的一般结构

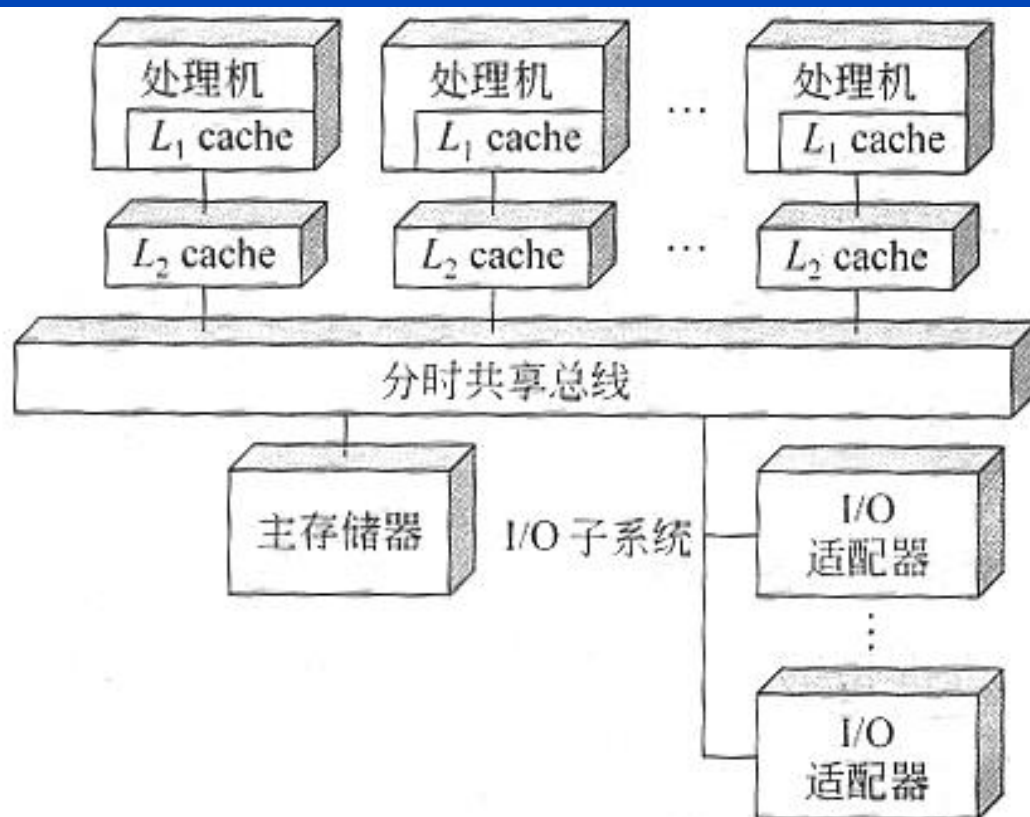


图 11.17 对称多处理机(SMP)的一般结构

- 1、高性能小型机中典型采用SMP结构
- 2、处理机数目可以根据需要动态扩展



多处理机

- **SMP的特点**

- 对称多处理机，即各个处理机结构与功能相相同，可完成相同的任务。
- 多处理机共享主存和I/O设施，以总线或者其它专用网络连接。
- 系统由集中操作系统OS统一管理，提供各处理机的作业/任务调度、以及处理机间的同步，使得多个处理机的存在对用户是透明的，感觉就是一个处理机。



多处理机

- **SMP的概念**

- 对称多处理机SMP是指结构与功能相同的多处理机通过总线或专用网络连接在一起，共享主存和I/O设施，由集中操作系统统一管理，提供各处理机的作业/任务调度、以及处理机间的同步，使得多个处理机的存在对用户是透明的，感觉就是一个处理机。



多处理机

• SMP与单处理机系统的比较

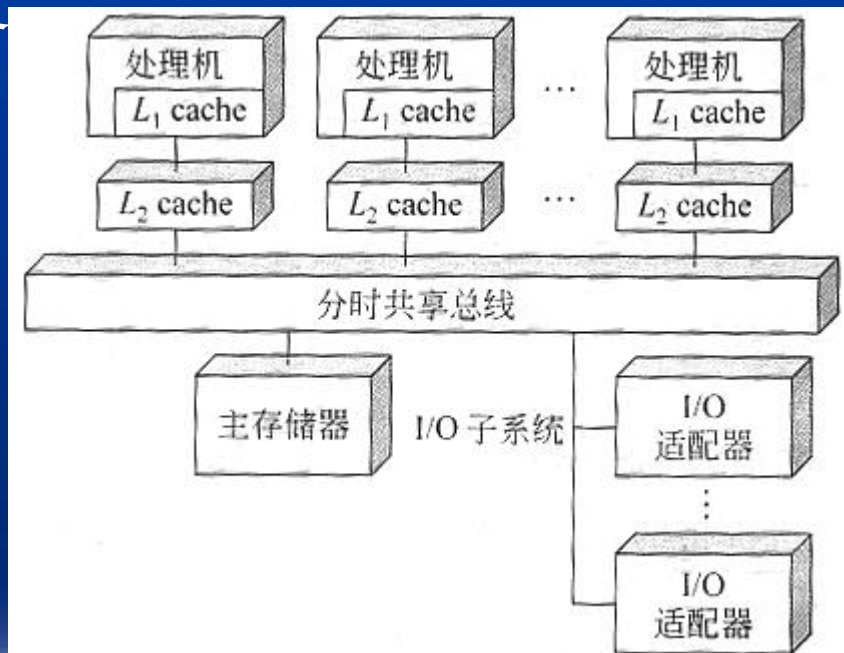
- 性能：如果可以对一台计算机完成的工作进行组织，使得某些工作部分能够并行完成；则具有多个处理机的系统与具有同样类型的单处理机系统相比，将产生更高的性能。
- 可用性：在一个对称多处理机系统中，所有处理机都能完成同样的功能，故单个处理机的故障不会造成系统的停机，系统可在性能降低情况下继续运行。
- 增量式增长：用户可以在系统中添加处理机来提高系统性能。
- 可扩展：厂商能提供一个产品范围，它们基于系统中配置的处理机数目不同而有不同的价格和性能特征。



多处理机

- **SMP的主要技术问题**

- 保证多cache的一致性
- 集中操作系统统一管理调度多处理机、以及处理机间同步问题
- 多核处理机芯片的管理调度





机群系统

- 机群系统的定义

- 机群系统由一组完整的计算机（指离开机群系统仍能独立运行自己任务，有自己的操作系统、内存、I/O设施，一般称为节点），通过标准通信协议、商品化网络（易于连接应用）连接在一起，在机群软件（中间件）统一管理调度下，可作为一个统一的计算资源协同工作，使用户感觉就是一个计算机系统。
 - （1）机群是由完整的计算机（节点）互连而成
 - 机群的节点是一台可独立工作的计算机：有自己的操作系统、内存、I/O设施，既可独立完成任务，也可协同完成任务。
 - 机群节点之间采用标准通信协议、商品化网络实现互连。



机群系统

• 机群系统的定义

- 机群系统由一组完整的计算机（指离开机群系统仍能独立运行自己任务，有自己的操作系统、内存、I/O设施，一般称为节点），通过标准通信协议、商品化网络（易于连接应用）连接在一起，在机群软件（中间件）统一管理调度下，可作为一个统一的计算资源协同工作，使用户感觉就是一个计算机系统。
 - （2）机群应能作为一个统一的计算资源使用
 - 在机群软件的统一管理调度下，各节点可协同工作、形成一个统一的计算资源。
 - 即从用户角度来看，整个机群就像一个计算机系统，可以很方便地来使用这个机群，而不必关心提供服务的设备在什么地方。



机群系统

• 机群系统的定义

- 机群系统由一组完整的计算机（指离开机群系统仍能独立运行自己任务，有自己的操作系统、内存、I/O设施，一般称为节点），通过标准通信协议、商品化网络（易于连接应用）连接在一起，在机群软件（中间件）统一管理调度下，可作为一个统一的计算资源协同工作，使用户感觉就是一个计算机系统。
 - (3) 机群不同于局域网
 - 在局域网中各台计算机基本上都是各自独立地工作，他们只是通过局域网共享资源，没有单一系统形象。
 - 机群中的各计算机既可以单独使用，又是多台计算机连接的一个整体，因而机群可以充分利用机器资源，充分利用通用的计算机产品，达到高并行性和高可靠性的要求。



机群系统

• 机群系统的定义

- 机群系统由一组完整的计算机（指离开机群系统仍能独立运行自己任务，有自己的操作系统、内存、I/O设施，一般称为节点），通过标准通信协议、商品化网络（易于连接应用）连接在一起，在机群软件（中间件）统一管理调度下，可作为一个统一的计算资源协同工作，使用户感觉就是一个计算机系统。

- (4) 机群不同于MPP

- MPP节点上采用的处理机往往较简单、无独立操作系统，利用总线或专用网络连接，必须由集中OS统一管理调度。
- 机群中的节点是完整的计算机系统，包括自己的存储和I/O设备资源，之间采用标准通信协议、商品化网络互连；其并行性在操作系统级之上（即操作系统之上的机群软件支撑）



机群系统

- 机群系统的特点

- 使用方便

- 机群中的节点一般是中低端的、传统的典型的成熟的平台，用户可在熟悉的环境中开发、管理、运行程序。
 - 易于传统软件移植。
 - 实际上是利用传统的中低端计算机，组成机群系统，获取中高端计算机的性能。

- 可靠性好

- 机群中每个节点都是完整的计算机系统，一个节点故障或崩溃，其它节点仍能工作。

- 可伸缩性好

- 机群计算能力随节点增加而增加，增加节点方便。



机群系统

- 机群系统的特点

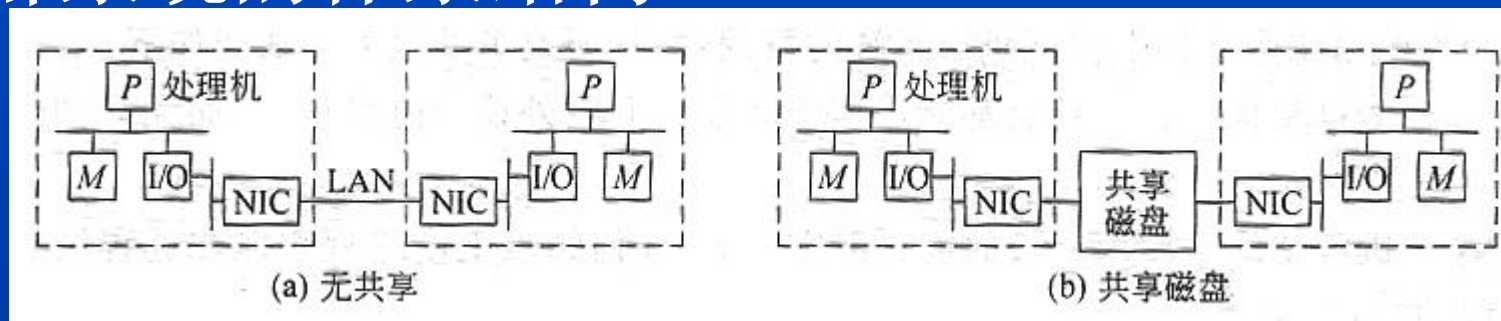
- 性能价格比好

- 机群的节点和互连网络都是商品化的通用的计算机产品，大批量生产，成本降低，因而性能价格比好。

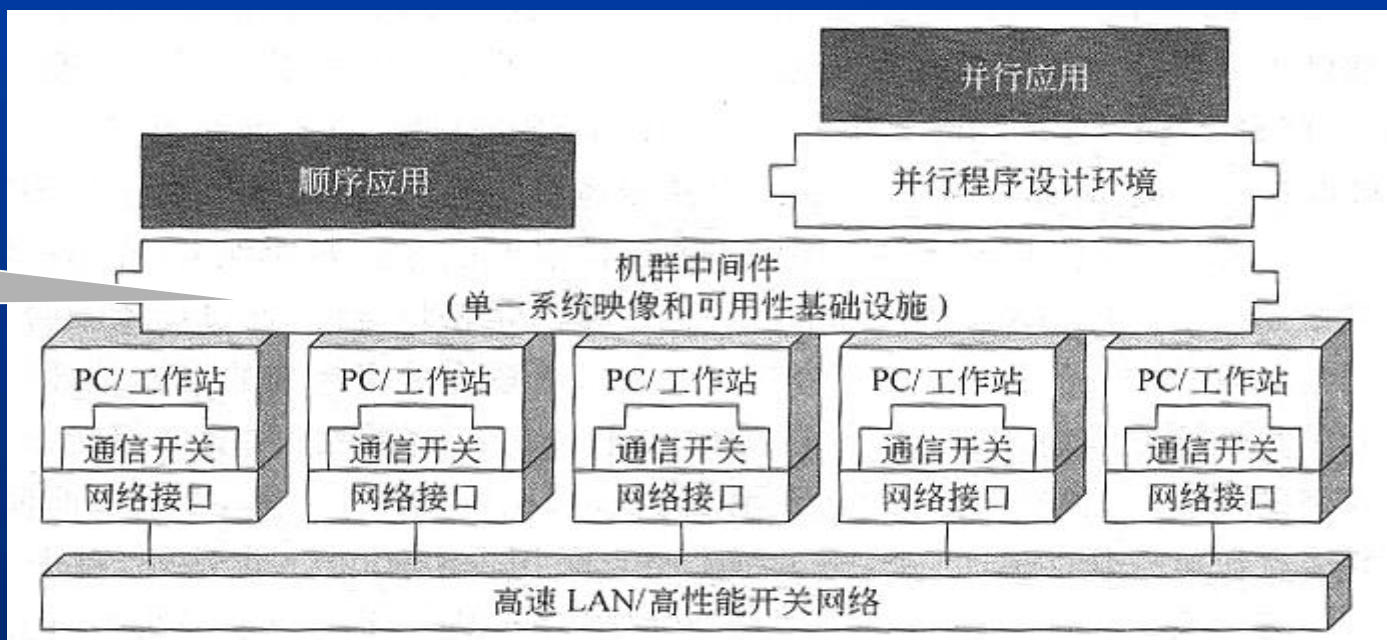


机群系统

• 机群系统的体系结构



机群
软件





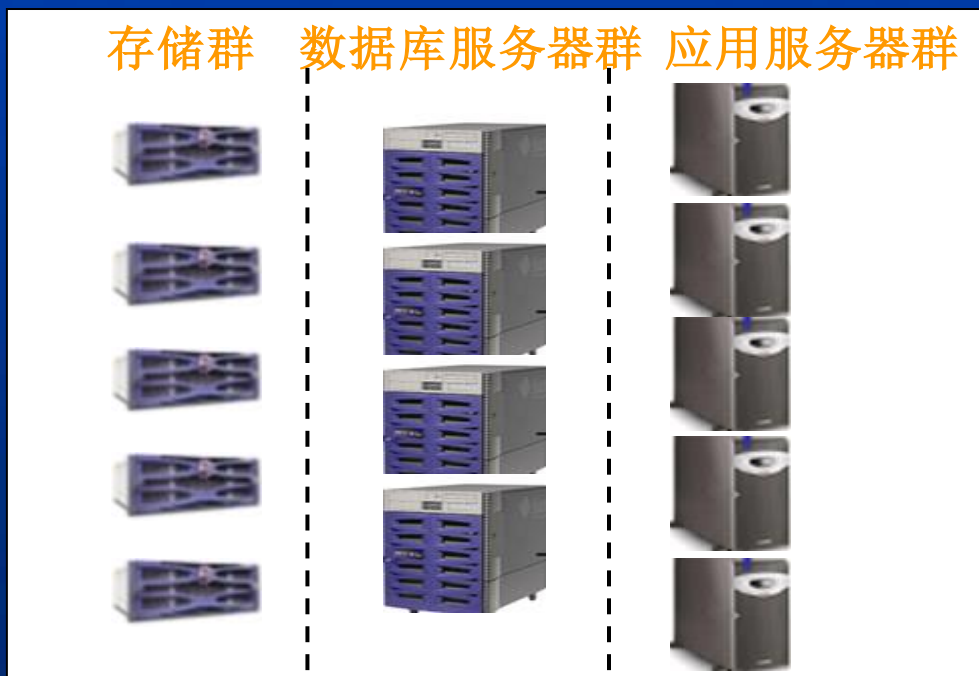
机群系统

- 机群系统的典型应用

- CLUSTER系统

- 利用cluster构建高可用性的数据库服务器、应用服务器群，SUN、IBM都有相应的cluster产品。

数据中心





机群系统

- 机群系统的典型应用

- 超级刀片系统

- 面向高性能计算、商业应用机群。
 - 主机柜：利用总线、高速开关互连的，可热插拔的系统架构；刀片：一个计算机系统单元、独立操作系统，即一台计算机，可单独使用，也可机群使用。





重点、难点

- 基本概念
 - 并行性、VLIW处理机、超线程技术、向量处理机、对称多处理机SMP、机群系统
- 体系结构的分类
 - 格林分类法



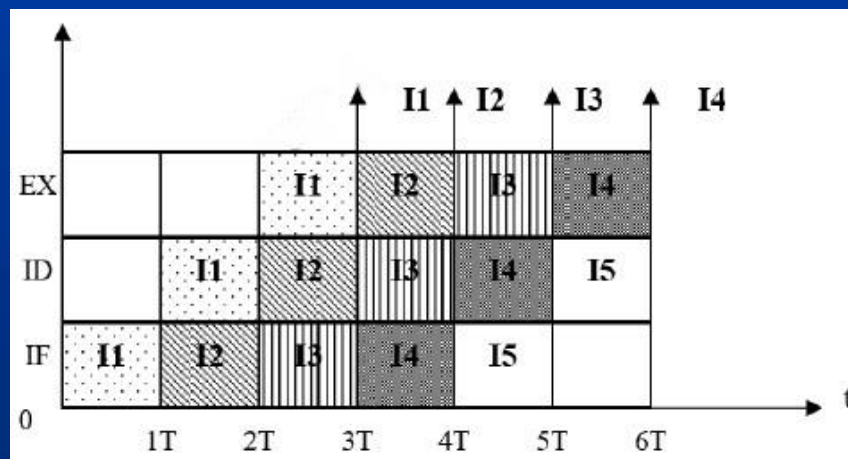
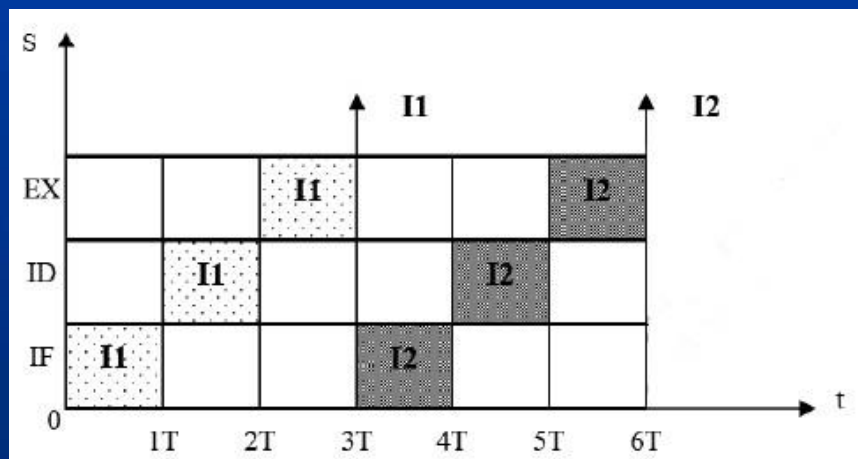
习题



习题

2. 如果一条指令的执行过程分为取指令、指令译码、指令执行三个子过程，每个子过程时间都为100 ns。

- 请分别画出指令顺序执行和流水执行方式的时空图。
- 计算两种情况下执行 $n=1000$ 条指令所需的时间。
- 流水方式比顺序方式执行指令的速度提高了几倍？



时间: $3 \times 100 \times n = 3 \times 10^5$ vs.

$3 \times 100 \times n - 2 \times 100 \times (n-1) = 100200$

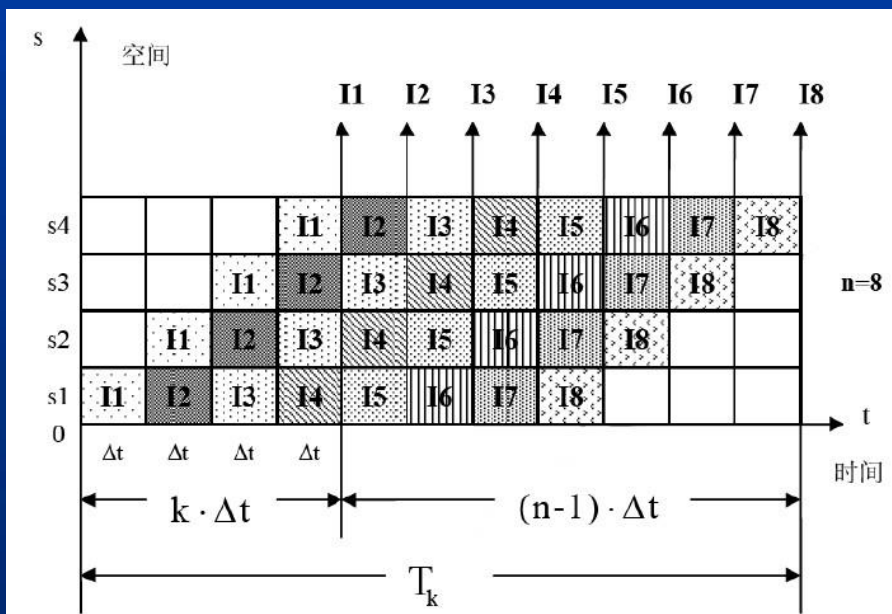
加速比 = $3 \times 10^5 / 100200 \approx 2.994$



习题

3. 设有 $k=4$ 段指令流水线，它们是取指令、译码、执行、存结果，各流水段持续时间均为 Δt 。

- 连续输入 $n=8$ 条指令，请画出指令流水线时空图。
- 计算流水线的吞吐率（单位时间中输出的指令数）。
- 计算流水线的加速比。



完成 n 条指令所需时间:

$$T_k = (k+n-1)\Delta t$$

吞吐率: n/T_k

加速比: $kn\Delta t/T_k$



习题

6. 试推导VLIW处理机对SISD处理机的加速比公式。

- 解：假设指令的执行分为 k 个过程，需要执行的指令有 n 条，VLIW处理机能并行执行 n 次运算。那么
 - SISD处理机需要的时间为 $(k+n-1)T$
 - VLIW处理机需要的时间为 kT
 - 加速比为 $(k+n-1)/k$



习题

7. 向量算术运算表达式为

$$D(I) = A(I) + B(I) \times C(I), \quad 0 \leq I \leq 63$$

现用如下向量指令来完成上述计算（M是存储器）：

LOAD	R1, B(I)	;	R1 ← M(K1+I)	K1是B(I)的起始地址
LOAD	R2, C(I)	;	R2 ← M(K2+I)	K2是C(I)的起始地址
MUI	R1, R2	;	R1 ← (R1) × (R2)	向量乘
LOAD	R3, A(I)	;	R3 ← M(K3+I)	K3是A(I)的起始地址
ADD	R3, R1	;	R3 ← (R3) + (R1)	向量加
STORE	D(I), R3	;	M(K4+I) ← (R3)	K4是D(I)的起始地址

（1）假设在SISD标量机中每次取数和存数需4个时钟周期，加法需2个时钟周期，乘法需8个时钟周期，循环结束判断需2个时钟周期。计算在SISD标量机上依次重复执行上述代码段64次，需要多少时钟周期？

$$64 * (4 + 4 + 8 + 4 + 2 + 4) + 64(2 + 2) = 1664 + 256 = 1920$$



习题

7. 向量算术运算表达式为

$$D(I) = A(I) + B(I) \times C(I), \quad 0 \leq I \leq 63$$

现用如下向量指令来完成上述计算（M是存储器）：

LOAD	R1, B(I)	;	R1 ← M(K1+I)	K1是B(I)的起始地址
LOAD	R2, C(I)	;	R2 ← M(K2+I)	K2是C(I)的起始地址
MUI	R1, R2	;	R1 ← (R1) × (R2)	向量乘
LOAD	R3, A(I)	;	R3 ← M(K3+I)	K3是A(I)的起始地址
ADD	R3, R1	;	R3 ← (R3) + (R1)	向量加
STORE	D(I), R3	;	M(K4+I) ← (R3)	K4是D(I)的起始地址

（2）若用一台SIMD向量处理机，以上述6条向量指令对64组向量数据执行向量操作，所用时钟速率与上面一样，计算向量处理机的总执行时间？**需4+4+8+4+2+4=26个时钟周期，加速比1920/26=73.85**

（3）向量处理机与SISD标量机相比，加速比是多少？



习题

8. 设 F 为多处理机系统中 n 台处理机可以同时执行的程序代码的百分比，其余代码必须用单台处理机顺序执行。每台处理机的执行速率为 x MIPS（每秒百万条指令），并假设所有处理机的处理能力相同。试用参数 n 、 F 、 x 推导出系统专门执行该程序时的有效MIPS速率表达式。

$$\frac{nx}{F + n(1 - F)}$$

9. 利用上题表达式，假设 $n=32$, $x=8$ MIPS，要求得到的系统性能为64 MIPS，试求 F 的值。

90%



习题

10. 假设使用100台多处理系统获得加速比80，求原计算机程序中串行部分所占的比例是多少？

解：设程序总量为 Y ，所求比例为 x ，则串行程序总量为 $Y \cdot x$ ，并行程序总量为 $Y \cdot (1-x)$ 。

设单机运算速度为 s ，则串行程序运行的时间为 $Y \cdot x / s$ ，并行程序运行的时间为 $Y \cdot (1-x) / 100s$ 。

加速比为：

$$80 = \frac{Y / s}{Y \cdot x / s + Y \cdot (1-x) / 100s} = \frac{1}{x + \frac{1-x}{100}}$$

计算得 $x \approx 0.25\%$ 。

13. 与第10题类似，答案：（1）加速比为3，（2）并行代码百分比75%。



习题

14. 现有如下程序，其并行化版本将在一个由32台计算机组成的机群系统上运行：

```
L1:   for (i=1; i <= 1024; i++){  
L2:       S[i]=0;  
L3:       for(j=1; j <= i; j++)           //从1加到i，共i次加法  
L4:           S[i] = S[i] + j;  
L5:   }
```

假设L2、L4两行分别需要2个机器周期（包括所有的处理器和存储器存取动作），不计其他代码的开销及资源冲突。计算：

（1）程序在单个计算机上总的执行时间是多少？

$$1024*2 + (1+...+1024)*2 = 1051648 \text{ 个机器周期}$$



习题

14. 现有如下程序，其并行化版本将在一个由32台计算机组成的机群系统上运行：

```
L1:   for (i=1; i <= 1024; i++){  
L2:       Si=0;  
L3:       for(j=1; j <= i; j++)  
L4:           Si = Si + i;  
L5:   }
```

假设L2、L4两行分别需要2个机器周期（包括所有的处理器和存储器存取动作），不计其他代码的开销及资源冲突。计算：

（2）提出并行化方式，使计算负载在32台计算机上平衡并行执行。负载平衡意味着对i、j两个循环，都要将相等数量的加法运算指派到各计算机上。

（3）机群系统并行执行产生的最小执行时间是多少？与单个计算机相比，速度提高多少？

先分为512组：（1+1024+2），（2+1023+2），.....

再细分为32组，每组 $1025 \times 16 = 16400$ 次加法，32次赋值。执行时间32864个周期。



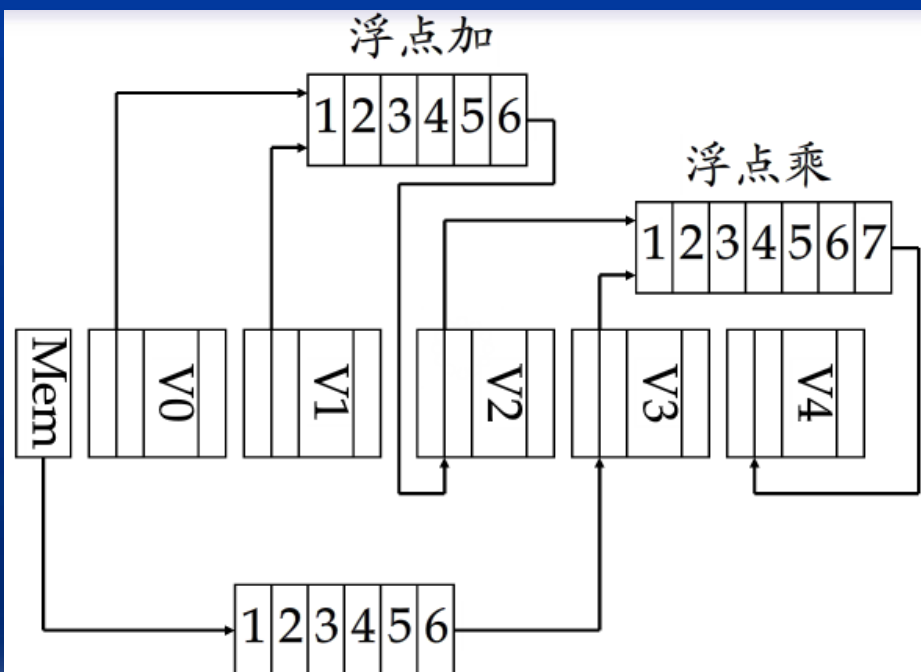
习题

补充题：向量链接技术

如下三条指令中，1、2可同时执行，3可链接执行。假设数据流入、流出功能部件都需要1个时钟周期，浮点加法和访存需要6个时钟周期，浮点乘法需要7个时钟周期。计算如下情况的运算时间：（1）三条指令串行；（2）前两条指令并行，然后与第三条指令串行；（3）向量链接。

```

1:  V3 ← A
2:  V2 ← V0 + V1
3:  V4 ← V2 × V3
    
```



串行: $(1+6+1+N-1) * 2 + (1+7+1+N-1) = 3N+22$

二并一串: $(1+6+1+N-1) + (1+7+1+N-1) = 2N+15$

向量链接: $(1+6+1) + (1+7+1)+N-1=N+16$