北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

9 并行体系结构

提纲

北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

9.1 体系结构中的并行性

提纲

9.1.1 并行性的概念

- 所谓并行性，是指计算机系统具有可以同时进行运算或操作的特性，它包括同时性与并发性两种含义：
  - 同时性：两个或两个以上的事件在同一时刻发生
  - 并发性：两个或两个以上的事件在同一时间间隔发生

9.1.1 并行性的概念

- 计算机系统中的并行性有不同的等级
- 从处理数据的角度看，并行性等级从低到高可分为：
  - 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

9.1.1 并行性的概念

- 从执行程序的角度看，并行性等级从低到高可分为：
  - 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

9.1.2 提高并行性的技术途径

- 时间重叠：即时间并行
  - 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
- 资源重复：即空间并行
  - 通过重复设置硬件资源，大幅度提高计算机系统的性能
- 时间重叠+资源重复——主流技术
- 资源共享
  - 用软件方法实现多个任务按一定时间顺序轮流使用同一套硬件设备，例如：多道程序

9.1.3 单处理机系统中并行性的发展

- 在发展高性能单处理机过程中，起着主导作用的是时间并行（流水线）技术
- 空间并行技术的运用也已经十分普遍
- 单处理机中，资源共享的概念实质上是用单处理机模拟多处理机的功能，形成即所谓虚拟机的概念
- 单处理机并行性发展的代表作有:
  - 奔腾系列机
  - 安腾系列机

9.1.4 多处理机系统中并行性的发展

- 多处理机系统也遵循时间重叠、资源重复、资源共享原理，向不同体系结构的多处理机方向发展
- 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
- 多处理机系统，分为：
  - 紧耦合系统（又称直接耦合系统）
    - 通过总线或高速开关实现互连，可以共享主存
    - 处理机之间物理连接具有相对较高的信息传输率
  - 松耦合系统（又称间接耦合系统）
    - 通过通道或通信线路互连，共享外存设备（磁带、磁盘）
    - 处理机之间的作用是在文件或数据集一级上进行

9.1.4 多处理机系统中并行性的发展

- 技术路线
  - 异构型多处理机系统：许多主要功能交由专用处理机完成
  - 同构型多处理机系统：为了使并行处理的任务能在处理机之间随机地进行调度，就必须使各处理机具有同等的功能
- 发展状况
  - 20世纪70年代以来，各类并行计算机系统问世
  - 20世纪80年代，我国研制了向量处理机YH-1/2和757

9.1.4 多处理机系统中并行性的发展

- 进入90年代以来，我国又研制了多种类型的并行计算机系统

表11.1 20世纪90年代以来我国自行研制的几种并行机

机器型号	完成时间	研制单位	CPU芯片	CPU数	机器类型
曙光1号	1993	中科院计算所 国家智能中心	M88000	4-16	SMP
曙光1000	1995	中科院计算所 国家智能中心	I860/xr	36	MPP
YH-3	1997	国防科大	MIPS R4000	128	Cluster
神威1	1999	国家并行计算机 工程技术中心	64-bit Alpha Processor	384	Cluster
深腾6800	2003	联想公司	Itanium2,1.3GH	1024	Cluster
曙光4000A*	2004	中科院计算所、 曙光公司	AMD Opteron 2.2GHz 64位	2560	Cluster
超级刀片系统	2004	深圳星盈科技公司	Xeon#EM64T	按刀片数 扩充	Cluster
KD-50-1	2007	中国科技大学	龙芯2F*	330	Cluster
天河1号	2009	国防科大	Xeon E5540 ATI RADEON	6144 5120	Cluster

9.1.4 多处理机系统中并行性的发展

- 天河1号（一期）
  - 2009年居TOP500世界第五位
  - 峰值速度 4700万亿次
    - 运算1小时，相当于全国13亿人同时计算340年以上的时间
    - 运算1天，相当于1台双核的高档桌面电脑运算620年以上的时间
    - 563.1万亿次的Linpack（线性系统软件包）实测性能
  - 存储容量为2PB，也就是2千万亿个字节
    - 做个换算对比，1个汉字平均为2个字节，“天河一号”即可在线存储1000万亿个汉字，相当于存储100万字的书籍10亿册

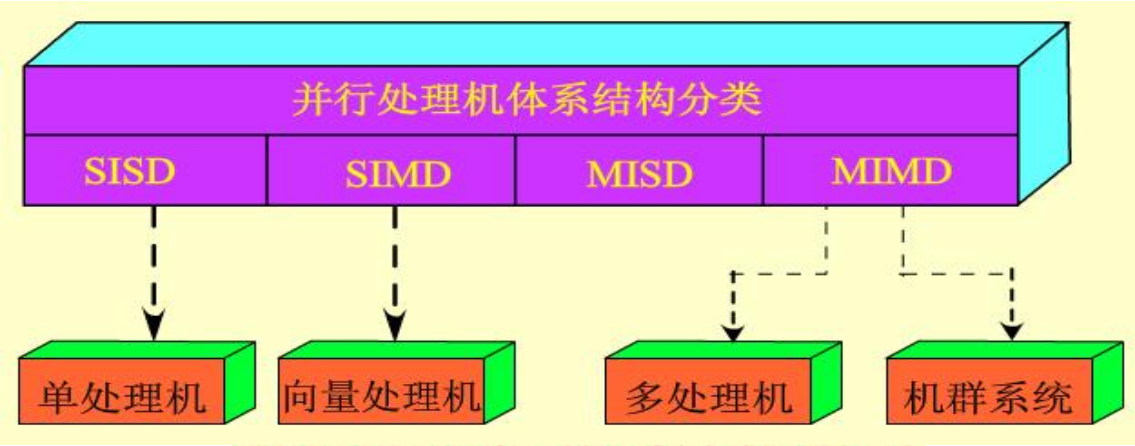
9.1.4 多处理机系统中并行性的发展

- 24小时功耗10万千瓦时（10万度电/天：低功耗）
  - 满负荷运行的总功耗是4.04兆瓦，也就是每小时耗电4040千瓦时，24小时满负荷工作耗电接近10万千瓦时。这个数字令人惊叹，但实际上“天河一号”在超级计算机当中是一台相对节能的、绿色的超级计算机。经过测算，它的能效值仅低于目前能效排名世界第一的IBM蓝色基因系统
- 总重量160吨
  - 由140个机柜组成，占地约700平方米，总重量约160吨
  - 大家站在“天河一号”前，会觉得它气势宏伟、震撼人心。但实际上，“天河一号”在世界上已有的千万亿次超级计算机中（多数是近千平方米的占地），算是一个身材苗条的小个子

9.1.4 多处理机系统中并行性的发展

9.1.5 并行处理机的体系结构类型

- 从计算机体系结构的并行性能出发，按照指令流和数据流的不同组织方式，分为：
  - 单指令流单数据流（SISD），其代表机型是单处理机
  - 单指令流多数据流（SIMD），其代表机型是向量处理机
  - 多指令流单数据流（MISD），这种结构从来没有实现过
  - 多指令流多数据流（MIMD），其代表机型是多处理机（为紧耦合系统）和机群系统（为松耦合系统）



并行处理机体系结构类型

9.1.5 并行处理机的体系结构类型

- CU控制单元，PU：处理单元，MU：存储单元
- IS：单一指令流
- DS：单一数据流

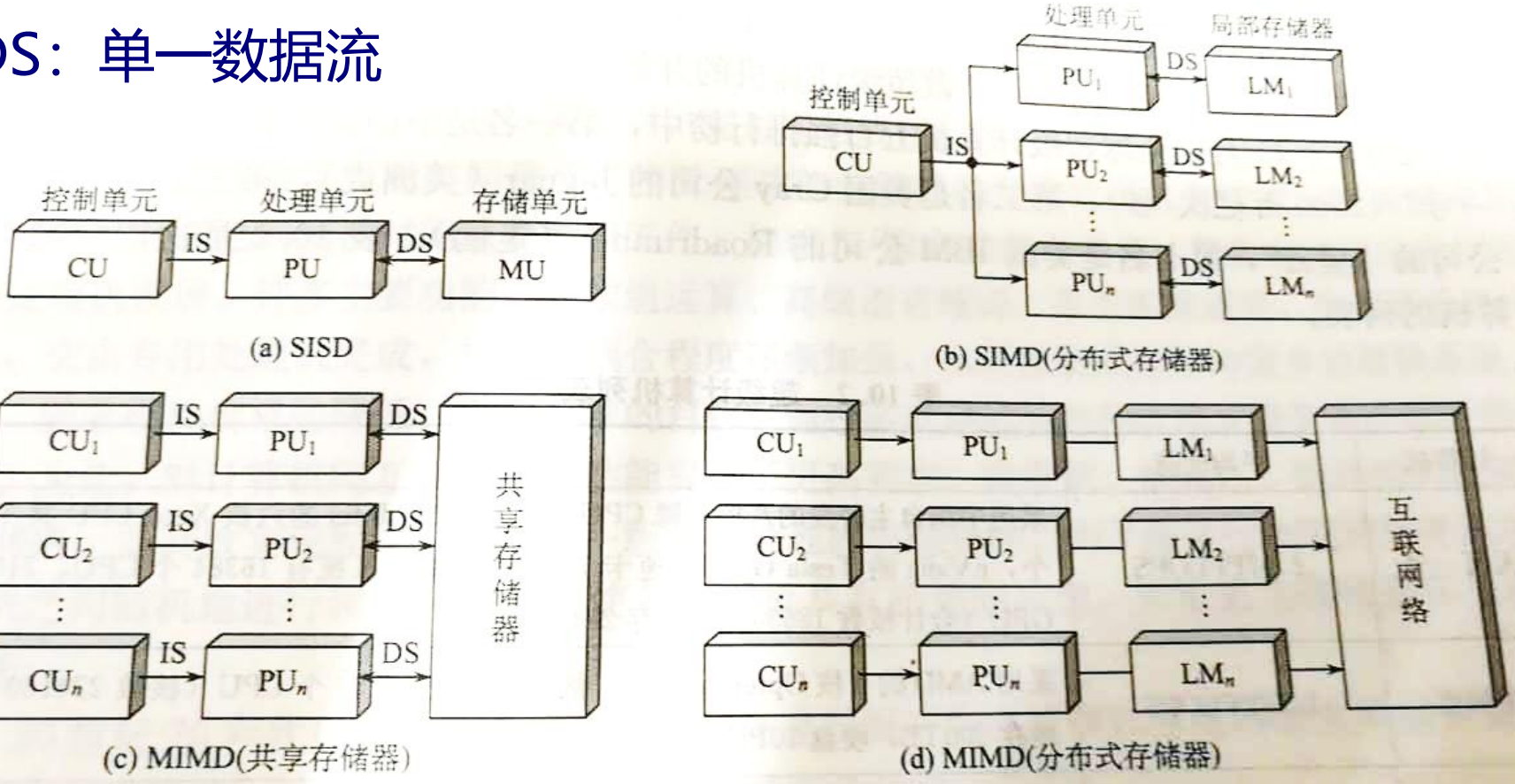


图 10.2　并行处理机的组成

9.2 超长指令字处理机

提纲

9.2.1 VLIW处理机的特点

- VLIW (very long instruction word）超长指令字
  - 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
- VLIW处理机
  - 是一种单指令多操作码多数据的体系结构（指令字长度约在100-1000位之间）
  - 用一条长指令实现多个操作的并行执行，并行操作主要在流水的执行阶段进行的
  - 右图执行阶段可并行执行3个操作

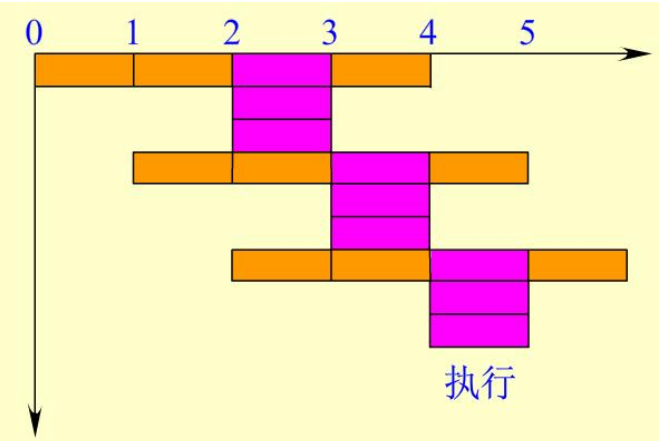
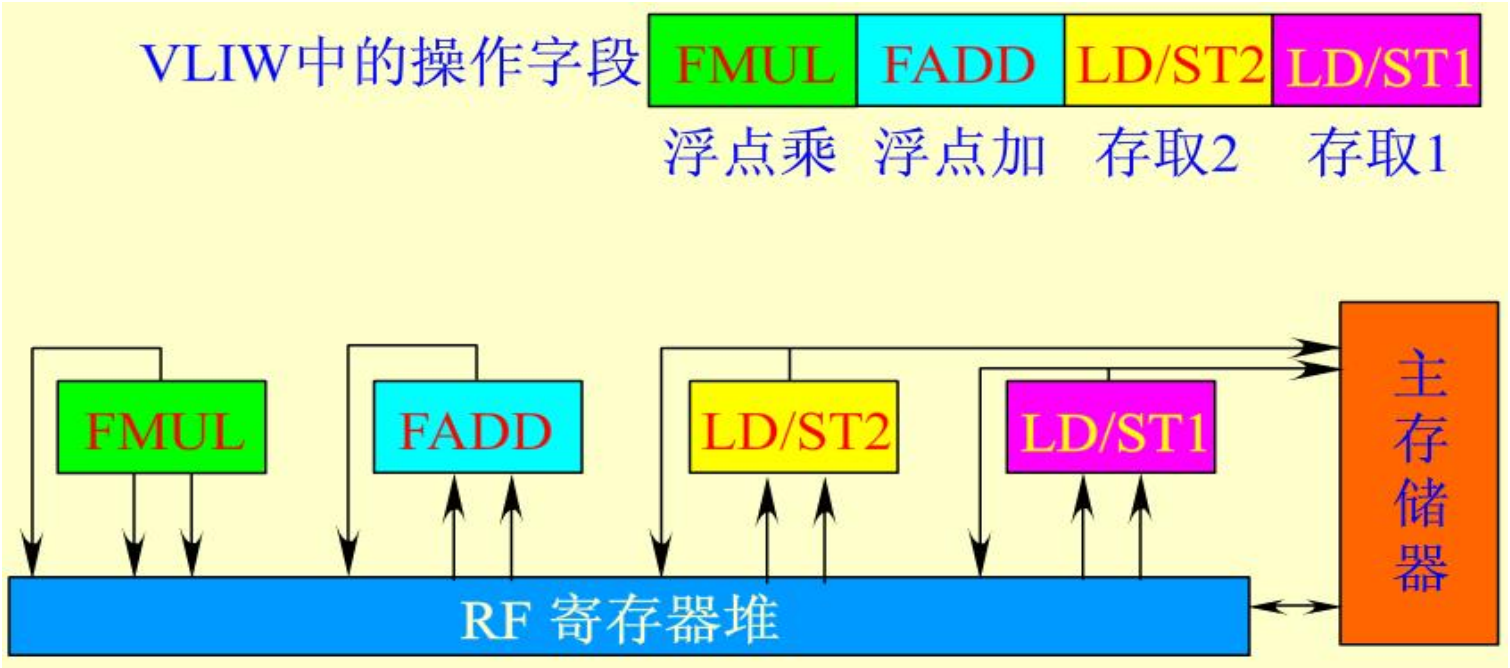


图11.3 超长指令字处理机的时空图

9.2.1 VLIW处理机的特点

- VLIW处理机的主要特点
  - 超长指令字的生成由编译器完成，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。由于编译时已解决可能出现的数据相关和资源冲突，故控制硬件比较简单

9.2.2 VLIW处理机的结构模型



- 两个访问主存储器的存取部件（LD/ST）
- 一个浮点加部件（FADD）
- 一个浮点乘部件（FMUL）

9.2 超长指令字处理机

- 例：假设要执行以下赋值语句：
  - C=A+B
  - K=I+J
  - L=M-K
  - Q=C×K

- 假设取数指令LOAD、存数指令STORE、浮点加法指令FADD操作要一个周期完成
- 浮点乘法指令FMUL操作需两个周期完成

表11.2 串行操作指令序列

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

9.2 超长指令字处理机

- 如果在VLIW机器中，采用表调度的编译方式，可以将串行的13条指令序列压缩为6条长字指令，仅需6个周期就能完成同样的操作

表 10.4 VLIW 操作的表调度

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×K
周期 5		STROE L		
周期 6	STORE Q			

9.3 线程与超线程处理机

提纲

9.3.1 指令级并行与线程级并行

- 提高处理机性能的传统方法：
  - 提高处理机的时钟频率，增大Cache容量（受到半导体工艺技术和功耗的限制）
  - 超标量和超长指令字的方式，设置多条并行指令的指令流水线，实现指令级并行（ILP）（会造成垂直和水平浪费）
  - 指令级并行向线程级并行发展
- 硬件多线程技术是提高处理机并行度的有效手段
- 2002年，英特尔公司推出的采用超线程(Hype Threading)技术的Pentium 4处理机，就是同时多线程技术的具体实现，原有的单个物理内核经过简单扩展后被模拟成两个逻辑内核

9.3.1 指令级并行与线程级并行

- 垂直浪费
  - 资源冲突会导致不能继续执行新指令
- 水平浪费
  - 指令相关导致多条流水线中部分流水线被闲置

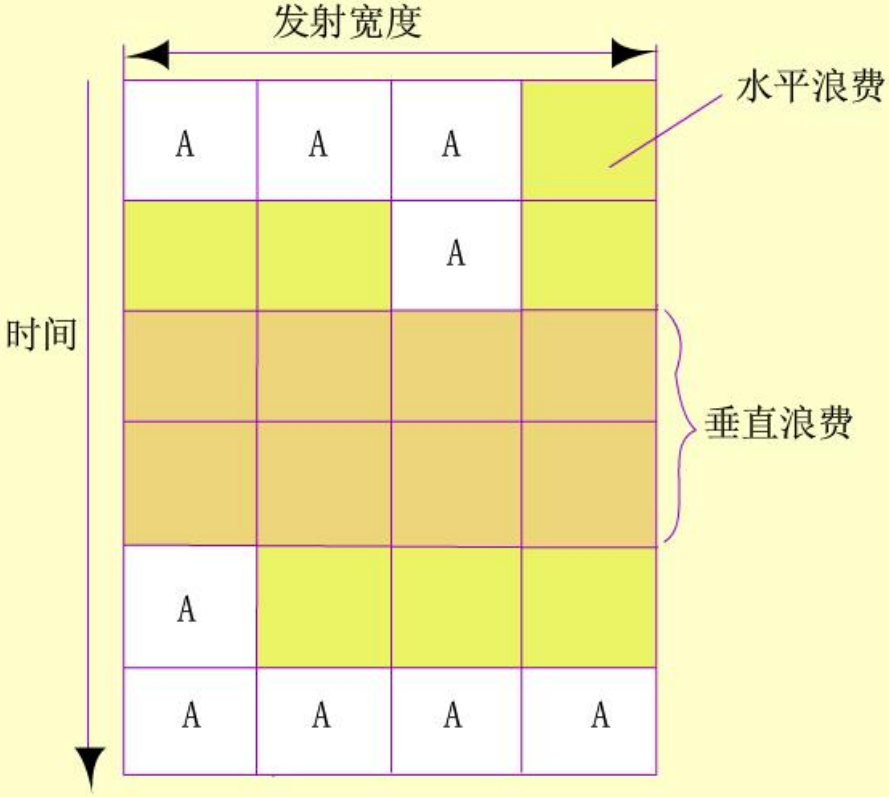


图11.7　超标量处理机的水平浪费和垂直浪费

- 如何减少处理机执行部件的空闲时间成为提升处理机性能的关键，线程级并行技术正是针对这一问题引入的

9.3.1 指令级并行与线程级并行

- 线程：操作系统中能被独立执行的程序代码的基本单位
- 进程调度的缺点：系统资源的分配与回收、现场的保存与恢复等操作频繁，时空开销大
- 解决办法：以线程作为调度和执行的基本单位，每个进程拥有若干线程。
- 线程与属于同一个进程的其他线程共享进程所拥有的全部资源，调度时不进行资源的分配与回收操作，线程切换的时空开销小。

9.3.1 指令级并行与线程级并行

多线程处理机

- 在处理机设计中引入硬件线程的概念
  - 硬件线程用来描述一个独立的指令流，而多个指令流能共享同一个支持多线程的处理机。当一个指令流暂时不执行时，可以转向执行另一个线程的指令流
  - 并行的概念就从指令级并行扩展至线程级并行
- 多线程处理机的具体的实现方法又可分为：
  - 细粒度多线程（交错多线程）处理机（每个时钟周期都进行线程切换）
  - 粗粒度多线程（阻塞多线程）处理机（遇到代价较高的长延时操作时进行线程切换）

9.3.1 指令级并行与线程级并行

- 持两个线程的多处理机，每个时钟周期所有的流水线都用于执行同一个线程的指令，但在下一个时钟周期则可以选择执行另一个线程的指令并执行
- 可有效减少垂直浪费。但是，由于每个时钟周期执行的指令必须来自同一个线程，因而不能有效消除水平浪费

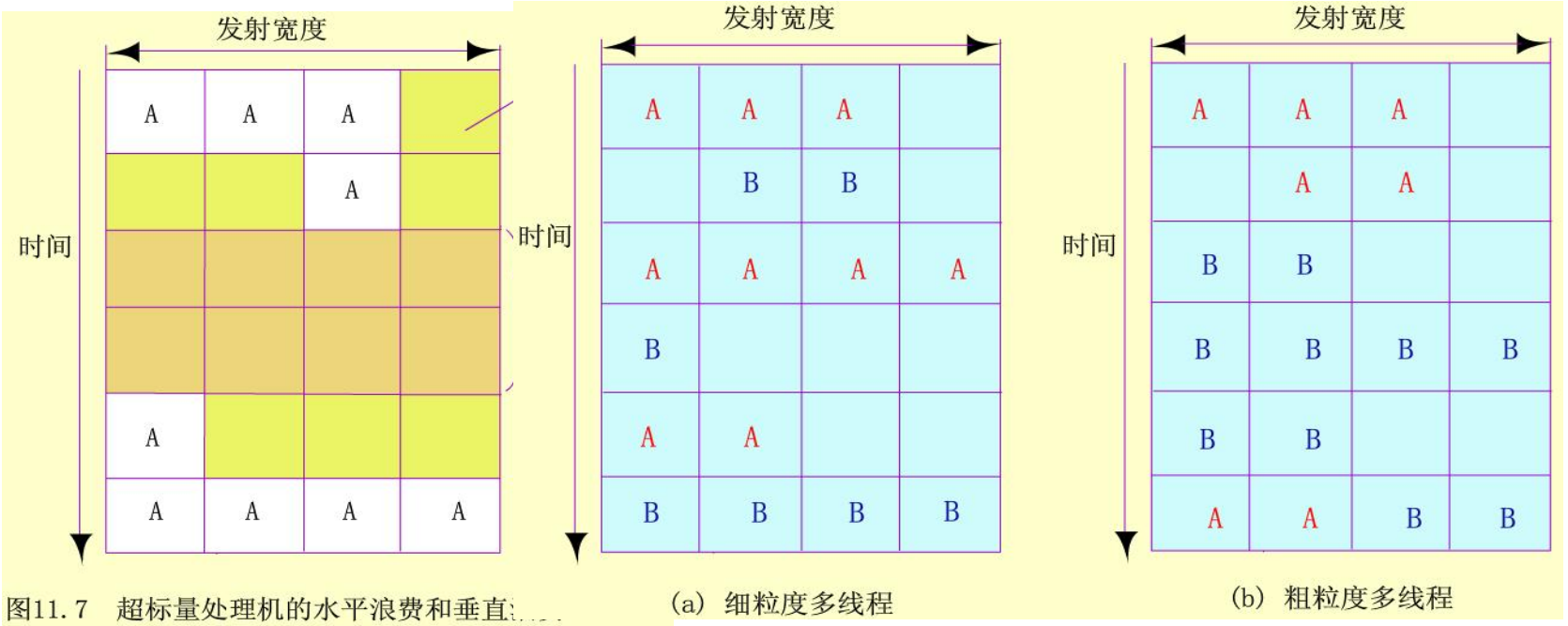


图11.7 超标量处理机的水平浪费和垂直

(a) 细粒度多线程

(b) 粗粒度多线程

9.3.2 同时多线程（SMT）

- 结合了超标量技术和细粒度多线程技术的优点。允许在一个时钟周期内发射多个线程的多条指令，可以同时减少垂直浪费和水平浪费
- 当线程B由于长延迟操作或资源冲突没有指令可以执行时，线程A甚至能够使用所有的指令发射时间

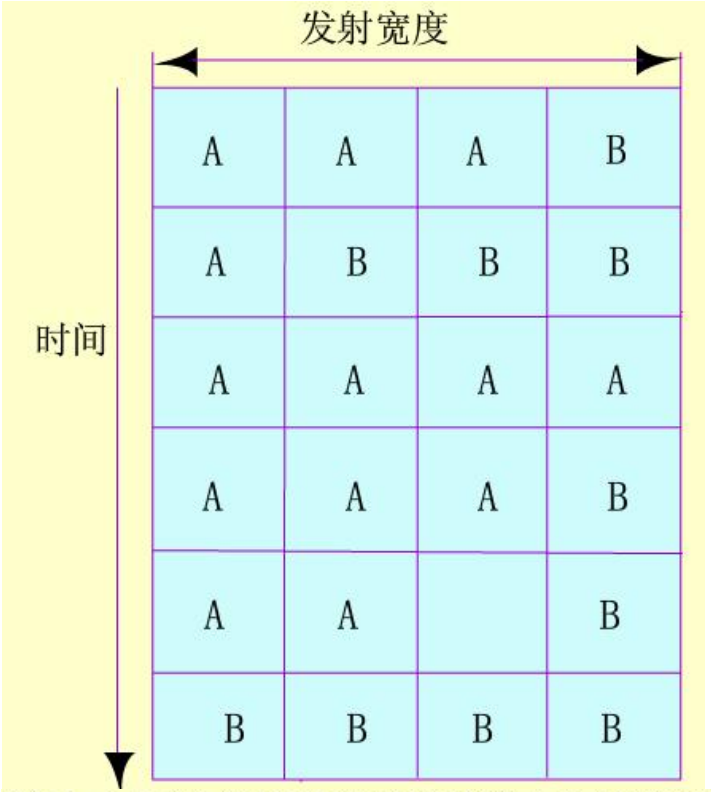


图11.9 同时多线程处理机的指令执行实例

9.3.2 同时多线程（SMT）

- 同时多线程在原有的单线程处理机内部为多个线程提供各自的程序计数器和相关寄存器以及其他运行状态信息，一个“物理”处理机被模拟成多个“逻辑处理机”，以便多个线程同步执行并共享处理机的执行资源。
- 应用程序无须做任何修改就能使用多个逻辑处理机。
- 如果多个线程同时需要某一个共享资源，只有一个线程能够使用该资源，其他线程要暂停并等待资源空闲时才能继续。
- 因此，同时多线程技术就性能而言，远不如多个相同时钟频率处理机组合而成的多核处理机（CMP）。所以，现在又设计了多核处理器计算机

9.3.3 超线程处理机结构

- 超线程技术是同时多线程技术在英特尔系列处理机产品中的具体实现
- 超线程处理机结构：为了支持两个硬件线程同时运行，需要对流水线进行改造。改造的方式是让每级流水线中的资源按以下三种方式之一复用于两个线程。
  - 复制：为处理机多个线程（2个）设置独立的部件。包括处理机状态、指令指针寄存器、寄存器重命名部件和TLB表等
  - 分区：将用于单线程的独立资源分割为两部分，分别供两个线程使用。主要有各种缓冲区和队列，如重排缓冲区、存/取数缓冲区等（每个线程使用的资源的容量减半，处理机成本没有增加）
  - 共享：处理机在执行指令的过程中根据使用资源的需要在两个线程之间动态分享资源。如乱序执行部件和Cache
- 代价：作业调度策略、取指和发射策略、寄存器回收机制、存储系统层次设计将变得非常复杂

9.4 向量处理机

提纲

9.4.1 向量处理的基本概念

- 标量：单个量
- 向量：一组标量
- 向量处理机
  - 具有向量数据表示和相应向量指令的流水线处理机
- 标量处理机
  - 不具有向量数据表示和相应向量指令的处理机
- 把N个相互独立的数叫做“向量”。对这样一组数的运算叫做“向量处理”
- 一条向量指令可以处理N个或N对操作数

9.4.1 向量处理的基本概念

- 例：计算表达式如下：

$$C_i = a_{(i+5)} + b_i \qquad i=10,11,12,\ldots,1000$$

- ① 用高级语言写出此表达式的循环部分；
- ② 用一条向量加法指令描述此表达式。
- 解 ① 使用C语言所写的循环部分为：
  - FOR (I=10, I <= 1000; I++)
    C (I) =A (I+5) +B (I)
- ② 这种FOR语句，在具有向量数据表示的机器中可用如下一条向量加法指令来实现，即
  - C (10 : 1000) =A(10+5:1000+5)+B(10:1000)

9.4.1 向量处理的基本概念

- 对向量的运算可以采用3种不同的处理方法：横向处理方法、纵向处理方法、纵横处理方法

- 例：计算D = A×（B + C ）

  A、B、C、D —— 长度为 N 的向量

向量A、B、C、D

$$A=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \quad B=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \quad C=\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \quad D=\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

9.4.1 向量处理的基本概念

1、水平（横向）处理方式

- 向量计算是按行的方式从左至右横向进行

$$K_i = b_i + c_i$$

$$D_i = K_i \times a_i$$

- 相关：N 次（每次K_i都会发生数据相关）
- 功能切换：2N次（2次乘和加功能的转移）
- 横向处理方法不适合于向量流水处理

9.4.1 向量处理的基本概念

2、垂直（纵向）处理方式

- 向量计算是按列的方式自上而下纵向进行

$$K = B + C$$

$$D = K \times A$$

- 相关：１次（数据相关只有一次）
- 功能切换：１次（加乘切换只需一次）
- 可获得较高的吞吐率，适合于在向量处理机中应用
- 存储器-存储器工作方式的向量处理机都采用纵向处理方法

9.4.1 向量处理的基本概念

3、分组（纵横）处理方式

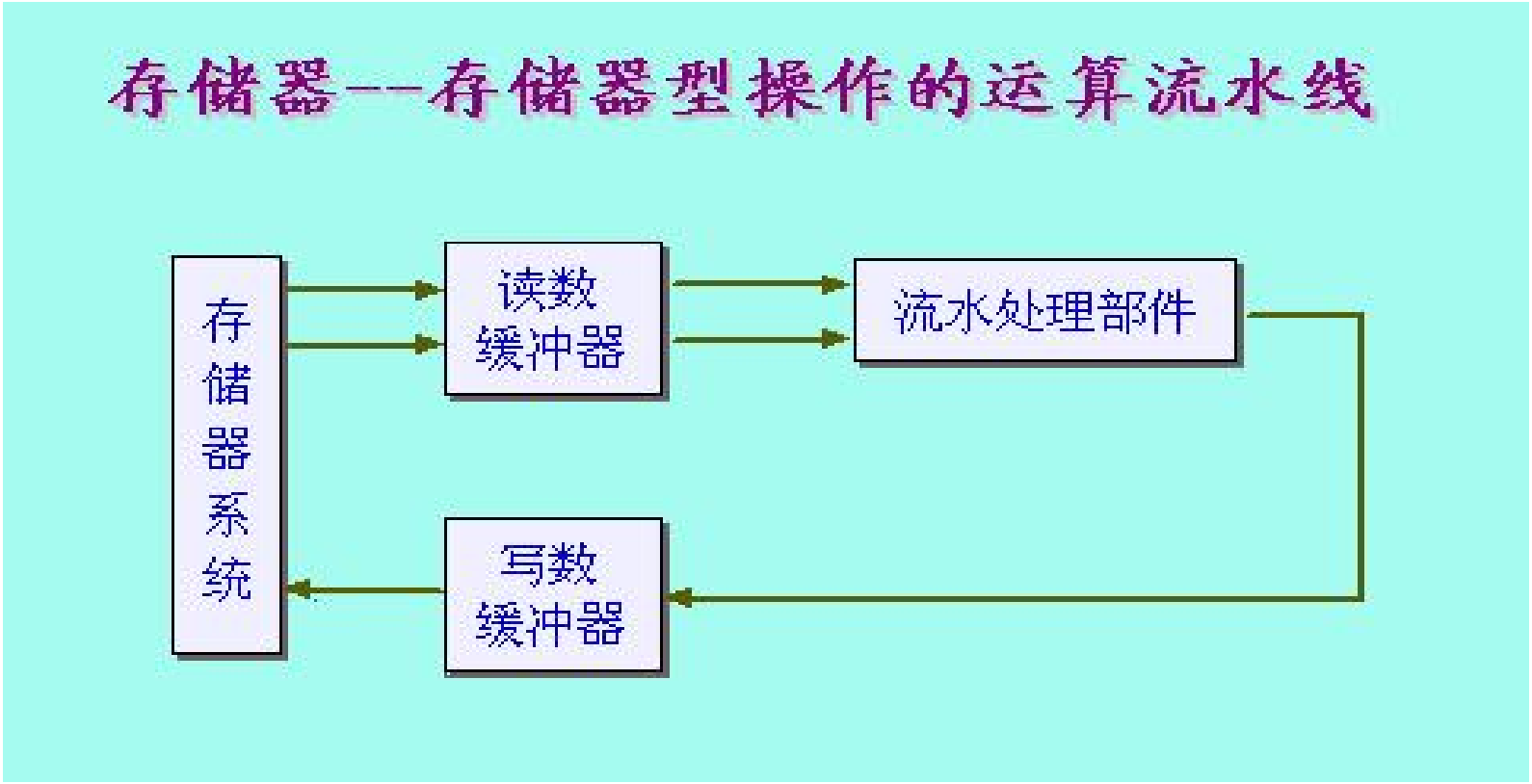
- 把向量分成长度为某个固定值的若干组，组内按纵向方式处理，依次处理各组
- 每组内：
  - 相关：1次
  - 功能切换：2次
- 对处理机结构的要求：有大量的寄存器，用来存放源向量、目的向量以及中间结果。纵横处理方法适合于寄存器-寄存器工作方式的向量处理机

9.4.2 向量处理机的结构

- 向量结构的一大优点就在于取一次指令可以完成一个很长的向量运算
- 要求向量计算机的存储器系统能提供给运算器连续不断的数据流以及接收来自运算器的连续不断的运算结果
- 参加运算的向量数据在存储器中，运算的结果也送到存储器中

9.4.2 向量处理机的结构

- 由于向量的长度不受限制，源向量和目的向量都存在存储器中，从而构成



- 这种结构对存储器的带宽要求很高

9.4.2 向量处理机的结构

1、向量处理机原理框图

- 向量处理机的基本思想是把两个向量的对应分量进行运算，产生一个结果向量。
- 如运算：C=A+B
  - 相当于$c_i=a_i+b_i$　$0 \leqslant i \leqslant N-1$
  - 读A、B、写C可以通过多端口存储器并行操作

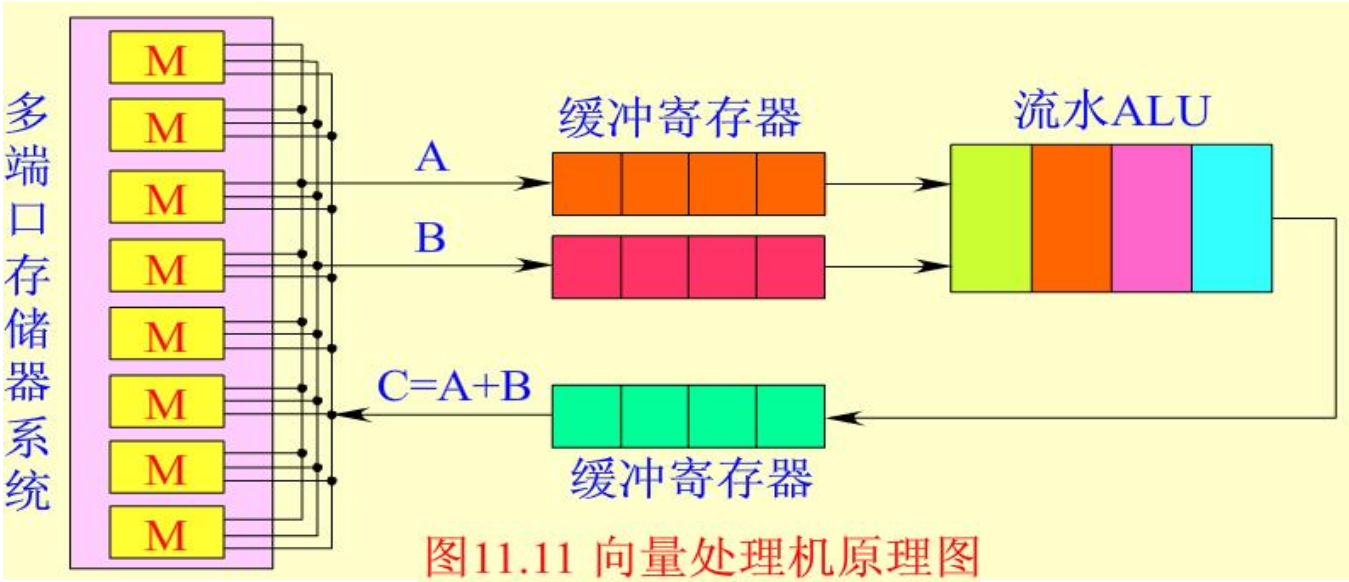


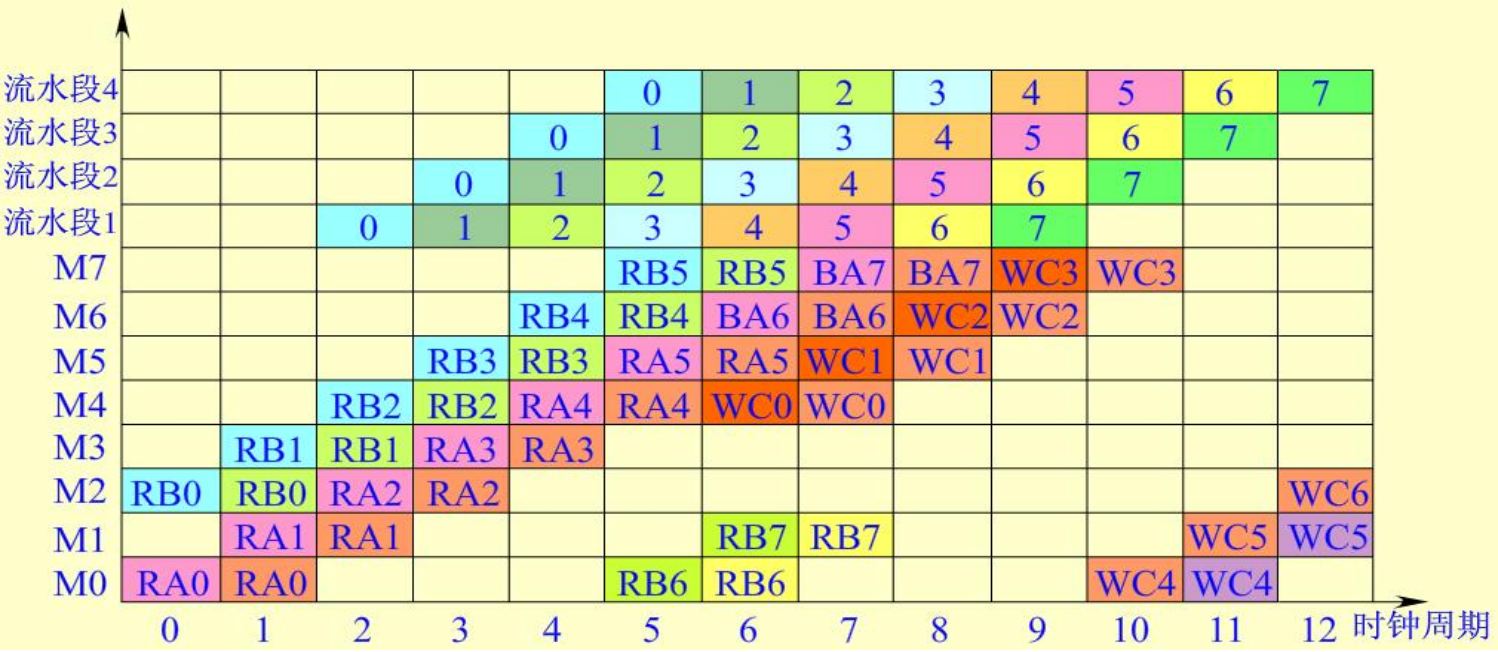
图11.11 向量处理机原理图

9.4.2 向量处理机的结构

- 流水线运算器与主存储器系统之间有三条相互独立的数据通路，各数据通路可以同时工作，但一个存储器模块在同一时刻只能为一个通路服务

9.4.2 向量处理机的结构

2、向量计算C=A+B的时空图



- 假定一个存取周期为两个CPU时钟周期，加法执行过程由4个流水段组成，向量ABC各有8个元素
- 在时钟周期6时有6个存储模块同时工作。此时运算器和存储器的工作衔接得非常好，在整个计算进行过程中没有任何冲突发生。
- 之所以如此，是特意将向量各元素按上述方式存放在各存储模块中

9.4.2 向量处理机的结构

3、寄存器-寄存器型向量处理机

- 然而实际情况并非与上图所示理想化的流水运行一样
- 读写冲突而断流
  - 有效方法是：由一级或多级中间存储器形成一个层次结构的存储器系统（中间存储器起着数据的中间存储作用，功能上相当于寄存器，因此称为寄存器-寄存器型向量处理机，通过寄存器寻址方式访问中间存储器，不需要像访问Cache时那样需要查Cache表）
- Cray-1系统
  - CRAY-1是一台典型的寄存器-寄存器结构的向量处理机，其运算速度达亿次/秒以上

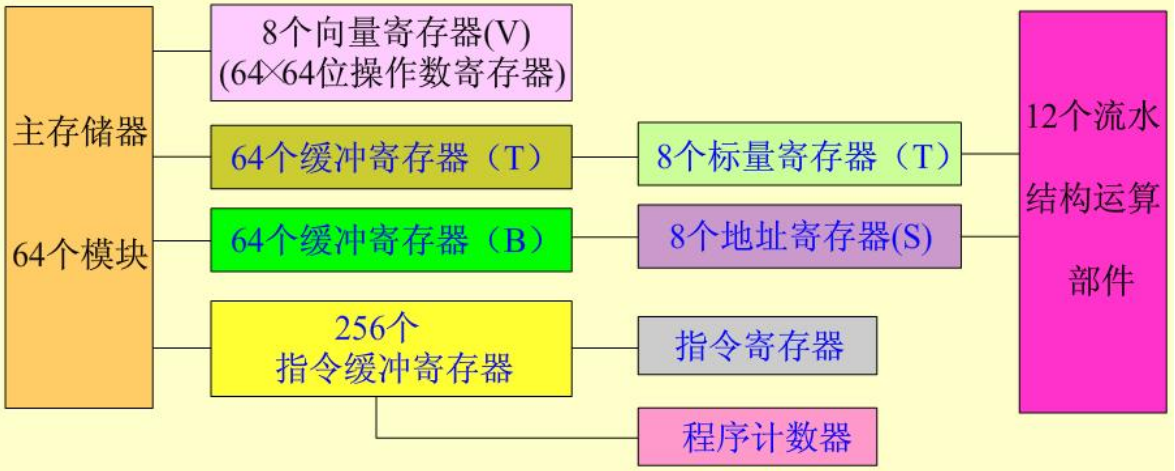


图11.13 Cray-1 寄存器-寄存器结构的向量处理机

9.4.2 向量处理机的结构

4、多功能部件的并行操作

- 在向量处理机中，为了加快向量操作，通常都采用独立的多个功能部件，并使它们并行工作

- Cray-1共4组12个单功能流水部件
- 上述12个功能部件都是独立的，每个功能部件的左边数字表示该部件的流水线延迟周期
- 只要满足一定的约束条件，它们可并行工作：
  - 不存在向量寄存器使用冲突
  - 不存在功能部件使用冲突

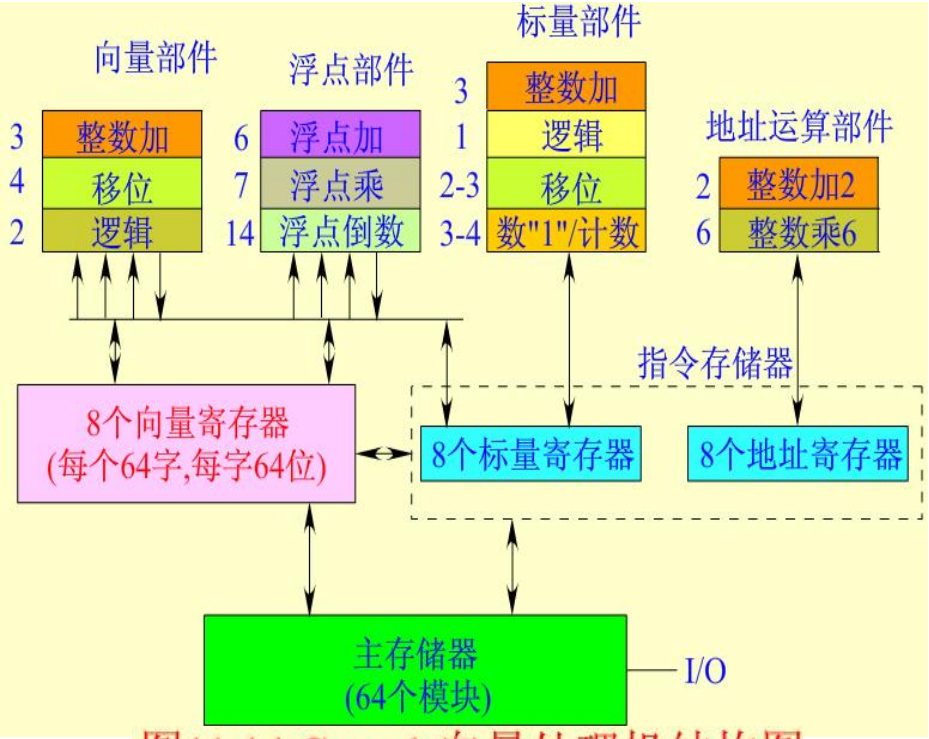


图11.14 Cray-1 向量处理机结构图

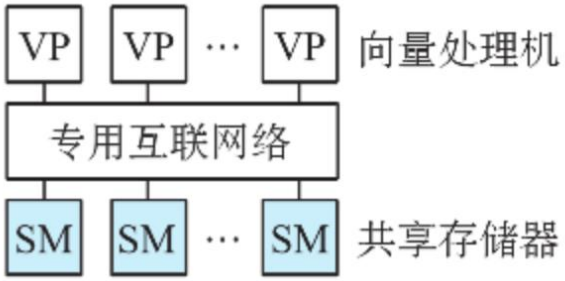
9.5 多处理机

提纲

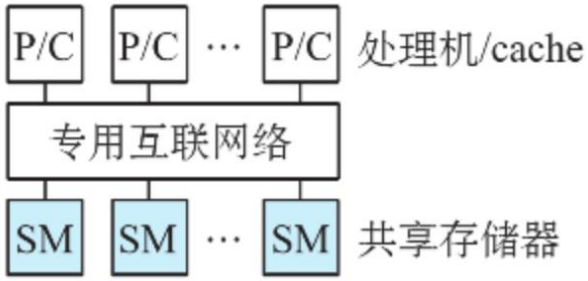
9.5.1 多处理机系统的分类

- 多处理机系统由多个独立的处理机组成，每个处理机能够独立执行自己的程序
- 分为四种类型：
  - 并行向量处理机
  - 对称多处理机
  - 大规模并行处理机
  - 分布共享存储器多处理机

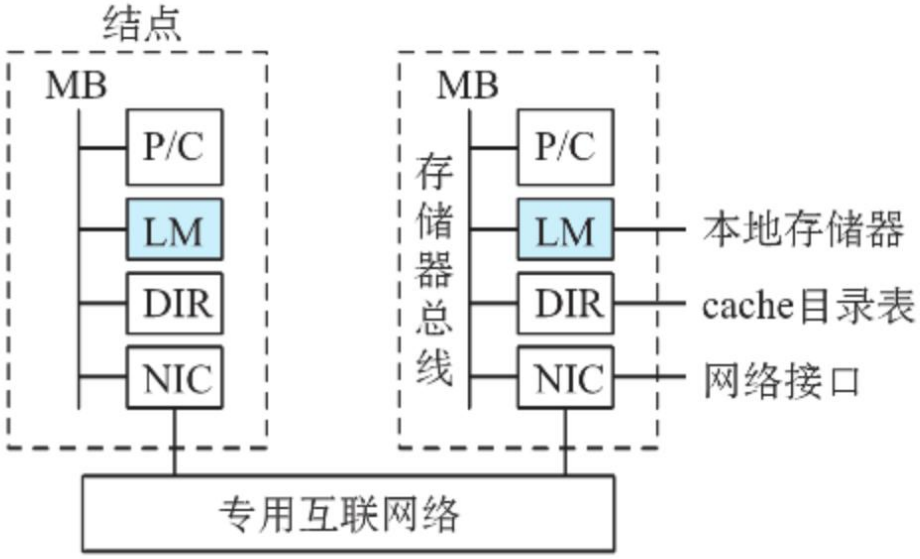
9.5.1 多处理机系统的分类



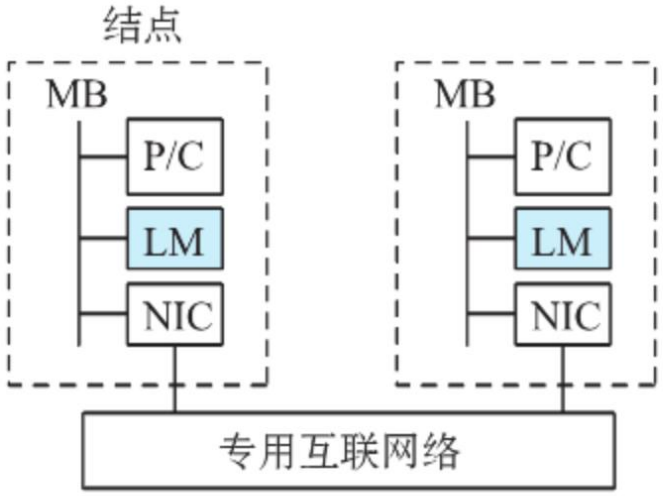
(a) 并行向量处理机(PVP)



(b) 对称多处理机(SMP)



(c) 分布共享存储器多处理机(DSM)



(d) 大规模并行处理机(MPP)

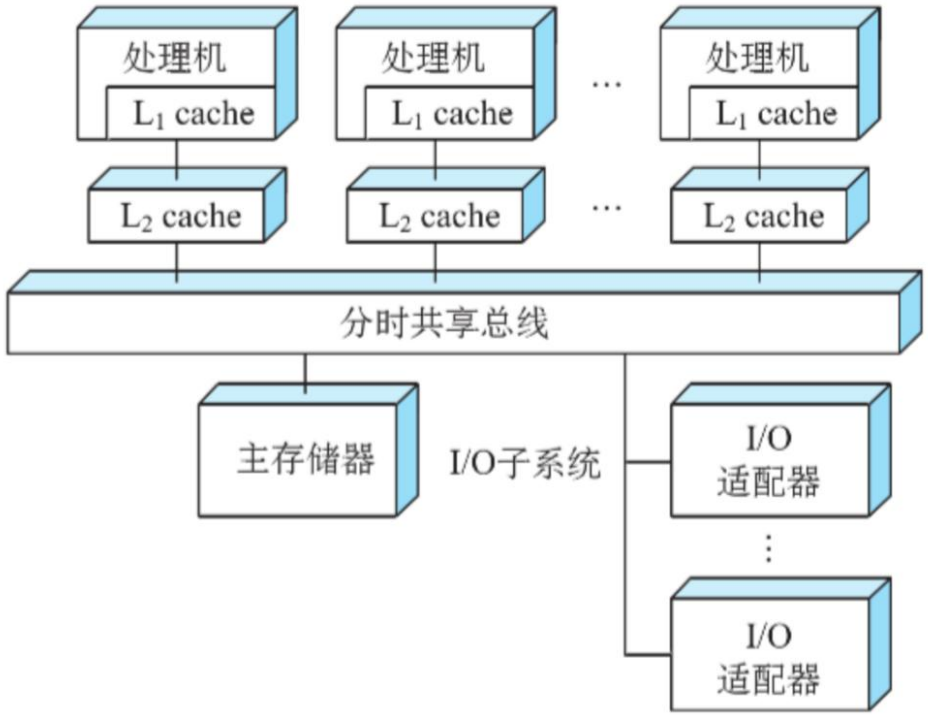
9.5.2 对称多处理机（SMP）

- SMP定义为具有如下特征的独立计算机系统：
  - 有两个以上功能相似的处理机
  - 这些处理机共享同一主存和I/O设施，以总线或其他内部连接机制互连在一起；这样，存储器存取时间对每个处理机都是大致相同的
  - 所有处理机共享对I/O设备的访问，或通过同一通道，或通过提供到同一设备路径的不同通道
  - 所有处理机能完成同样的功能
  - 系统被一个集中式操作系统（OS）控制。OS提供各处理机及其程序之间的作业级、任务级、文件级和数据元素级的交互

9.5.3 SMP的结构和实例

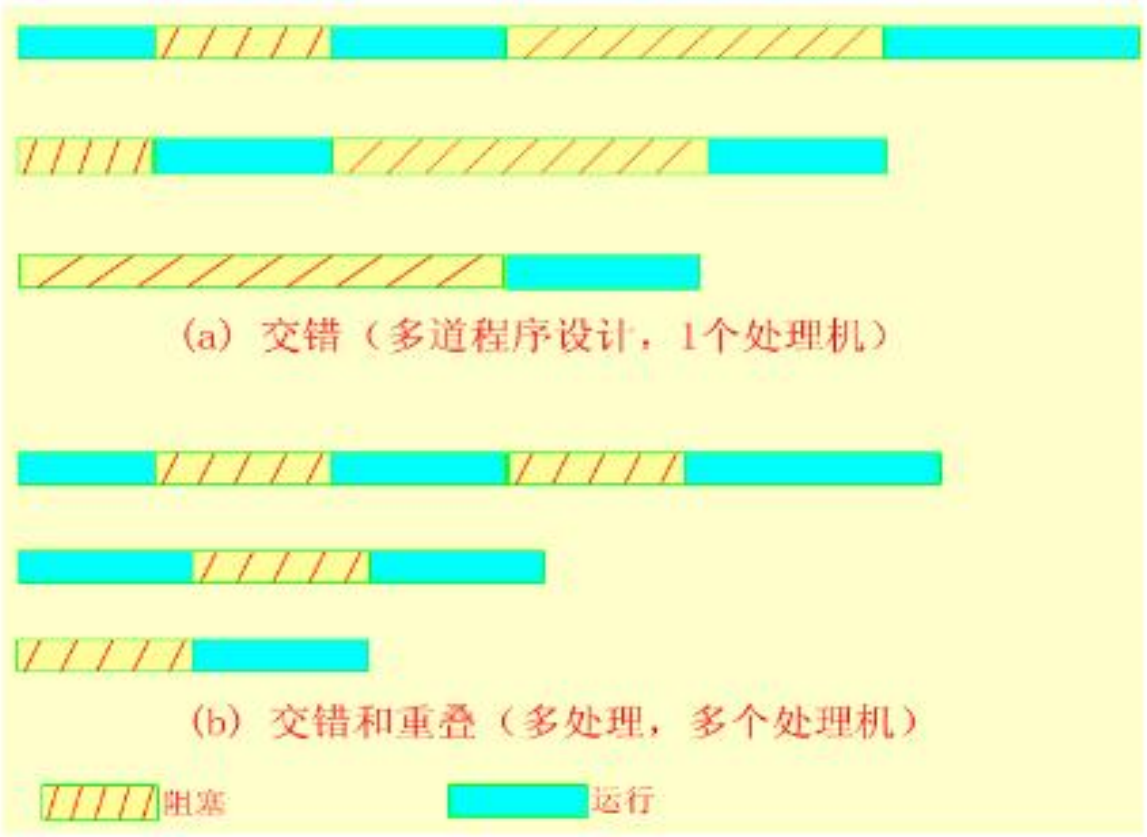
- 为便利来自I/O处理器的DMA传送，应提供如下特征：
  - 寻址：区别总线上各模块，确定数据的源和目标
  - 仲裁：任何I/O模块都能临时具备主控器的功能
  - 分时共享：保证分时共享总线

- 一般来说，工作站和个人机SMP都有两级Cache。现在，某些处理机还使用了L3 Cache



9.5.4 多处理机操作系统

- 一个SMP操作系统管理多个处理机和其它计算机资源，而使用户感觉到单一操作系统控制着系统资源。呈现为一个单处理机多道程序设计系统



9.5.4 多处理机操作系统

- 多道程序设计的关键问题
  - 同时并发进程管理：OS例程必须是可重入的，以允许几个处理机可同时执行同一指令流代码。要恰当地管理OS的表和其它结构，以避免死锁或无效操作
  - 调度：任何处理机都可完成调度，因此必须避免冲突
  - 同步：由于有多个活动进程可能访问共享地址空间或共享I/O资源，因此必须提供有效的同步
  - 存储管理：探查可用的硬件并行性，协调不同处理机上的分页机制
  - 可靠性和容错：必须识别每个处理机的失败，并相应地重构管理表

9.5.5 多处理机的Cache一致性

- 在单处理机中，Cache一致性问题只存在于Cache与主存之间
- 紧耦合多处理机中，需解决多处理机中Cache之间的一致性
- 产生内容不一致的原因
  - 可写数据的共享：修改某一数据块时，会引起其它处理机的Cache中同一副本的不一致
  - I/O活动：DMA数据传送也会导致Cache不一致
  - 进程迁移
- I/O活动和进程迁移的解决：禁止I/O通道或禁止进程迁移
- 可写数据的共享问题的解决办法：大体有两类
  - 软件方法：编译程序判断一致性问题并设置一致性指令
  - 硬件方法：通过硬件发现和解决Cache一致性问题
    - 目录协议法：维护有关数据块副本驻存在何处的信息
    - 监听协议法：数据修改时通过广播机构通知所有其它Cache

北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

9.6 机群系统（集群系统）

提纲

9.6.1 机群系统的定义和特点

1、机群系统的定义

- 机群系统（Cluster）是并行或分布计算机系统的一种类型，它是由一组完整的计算机通过高性能的网络或局域网互连而成的系统，它作为一个统一的计算资源一起工作，并能产生一台机器的印象
- 机群系统中的每台计算机一般称为结点

2、机群和大规模并行处理机（MPP）差别

- MPP的结点上采用的处理机往往比较简单，结点之间用频带较宽的专用网络互连
- 机群的结点则是一台完整的计算机，结点之间采用的一般是商品化的网络互连

9.6.1 机群系统的定义和特点

2、机群系统的特点

- 机群系统的优点：
  - 使用方便
  - 可靠性好，软硬件冗余，各结点有自己的操作系统
  - 可缩放性好
  - 性能价格好：结点和网络都是商品化的计算机产品
- 机群的不足之处
  - 维护工作量和费用较高
  - SMP管理员要维护的只是一个计算机系统
- 现在很多机群采用SMP作为结点

9.6.2 机群系统的体系结构

- 机群系统有两种不同的分类，依据是系统中的计算机是否共享对同一磁盘的存取：
  - 无共享（shared-nothing）配置，采用局域网连接
  - 共享磁盘（shared-disk）配置，共享磁盘一般用RAID

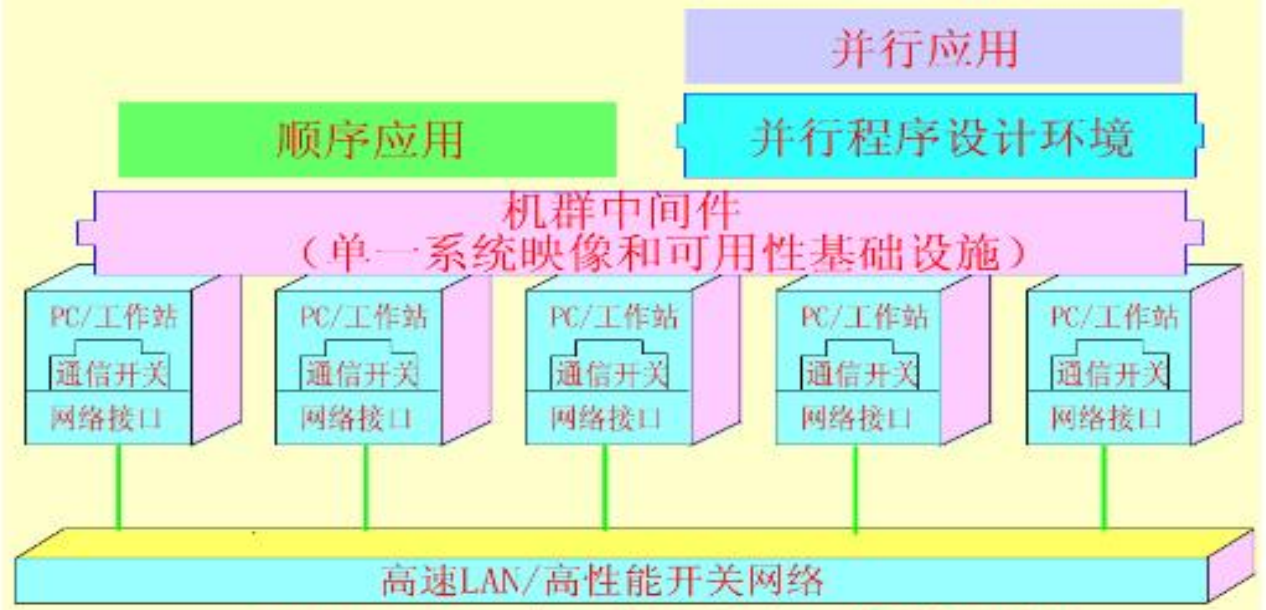


图11.21 机群计算机体系结构

9.6.2 机群系统的体系结构

- 各计算机都安装有中间件（middleware）和软件工具层
- 中间件的第一个作用是对用户提供了一个统一的系统映像
  - 单一文件层次：用户看到的是同一根目录下的单一文件目录层次体系
  - 单一虚拟网络：实际的机群配置可能由多个互连的网络组成，但任一结点都能存取机群中任何其他结点。因此有单一虚拟网络的操作
  - 单一存储空间：分布式共享存储器允许程序共享变量。
  - 单一作业管理系统：在机群作业调度程序管理下，用户提交作业无须指定执行此作业的宿主计算机
  - 单一用户接口：一个公共图形接口支持所有用户，不管用户由哪台工作站进入机群

9.6.2 机群系统的体系结构

- 中间件的第二个作用是保证机群系统的可用件
  - 单一I/O空间：任一结点都能远程访问任何IVO外设或磁盘设备，而无须知道它们的物理位置
  - 单一进程空间：使用一致的进程证实方案，任何结点上的进程都能在远程结点上生成一进程，并与之通信
  - 故障检查：该功能周期性地保存进程状态和中间计算结果，以在故障修复后重新运算
  - 进程迁移：该功能可使负载平衡。中间件机构需要识别机群系统不同成员上所出现的业务，并可将业务由一个成员移植到另一个成员上

9.6.3 IBM SP2系统

- 1997年战胜世界国际象棋冠军卡斯伯洛夫的“深蓝”，就是一个采用30个RS/6000工作站结点的IBM SP2机群
- SP2机群是异步的MIMD，具有分布式存储器体系结构
- SP2的结点通过网络接口开关（NIC）接到高性能开关（HPS），IBM把这种NIC叫做开关适配器

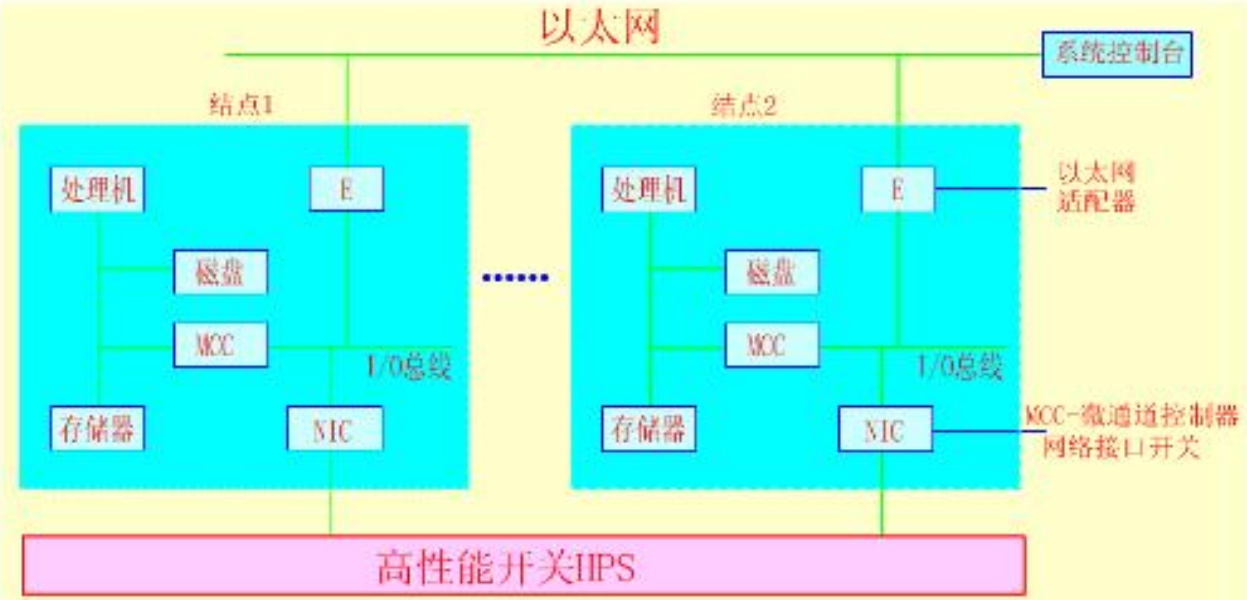


图11.22 SP2机群系统结构

9.6.4 超级刀片系统

- 深圳星盈科技公司研发的实时协作式超级刀片系统是最新一代的通用超级计算机系统设计

1、超级刀片机群在系统和应用方面的特点

- 革命性的设计理念：化繁为简
- 崭新的系统架构：使用普通的商业组件
- 高速运算和海量存储
- 安全性与可靠性
- 持久的生命力：计算刀片和存储单元可随时升级
- 扩展性和可塑性高
- 实时协作式应用和研发网格的信息平台
- 高效率的解决方案
- 广泛的使用领域

9.6.4 超级刀片系统

2、刀片的概念与技术规格

- 以通用的运算处理架构技术为基础，可以支持Intel Xeon EM64T处理机
- 每个刀片带有一个独立的管理控制模块
- 刀片的技术规格如下：
  - 处理机：支持2个IntelXeon EM64T处理机
  - 存储器：支持8GB内存
  - 硬盘数量：支持2个硬盘，最大支持800GB
  - Interconnect节点互连：InfiniBand 2X 10Gb接口
  - I/O扩展技术：支持任何PCI/O标准

3.可用的系统软件和开发工具

- 该系统的每个节点安装经过定制优化的RedHatLinuxAS3.1操作系统
- GNU编译器和调试器在Linux和Unix上广泛使用包括CC++、Fortran77，提供最好的程序兼容性