学习内容

- 5.1 重叠方式
- 5.2 流水方式
- 5.3 向量的流水处理与向量处理机
- 5.4 指令级高度并行的超级处理机
- 5.5 ARM流水线处理器举例

5.3 向量的流水处理与向量处理机

- 向量的特点
  - 元素之间无相关 ➔ 可连续流动
  - 对元素一般执行相同类型的操作 ➔ 单一功能，很少切换
- 所以向量很适合于用流水处理方式

5.3 向量的流水处理与向量处理机

- 一般将向量数据表示与流水处理方式结合在一起，构成向量流水处理机，也称其为向量处理机
- 向量处理机是解决数值计算问题的一种高性能计算机结构。
- 向量处理机一般都采用流水线结构，有多条流水线并行工作。
- 向量处理机通常属大型或巨型机，也可以用微机加一台向量协处理器组成。

5.3 向量的流水处理与向量处理机

- 一般向量计算机中包括有一台高性能标量处理机。
- 必须把要解决的问题转化为向量运算，向量处理机才能充分发挥作用。

5.3.1 向量的流水处理

- 向量的处理方式
  - 只有选择合适的向量的处理方式，才能充分发挥流水线的效能
  - 向量的流水处理所要研究的一个问题就是向量的处理方式

5.3.1 向量的流水处理

- 向量的处理方式
  - 但向量的处理方式与计算机的系统结构紧密相连，并互相影响
  - 不同的处理方式对流水处理机的系统结构、组成提出不同的要求，而系统结构、组成不同的向量处理机也会要求采用不同的向量的处理方式
  - 要根据向量运算的特点和向量处理机的类型选择向量的处理方式。

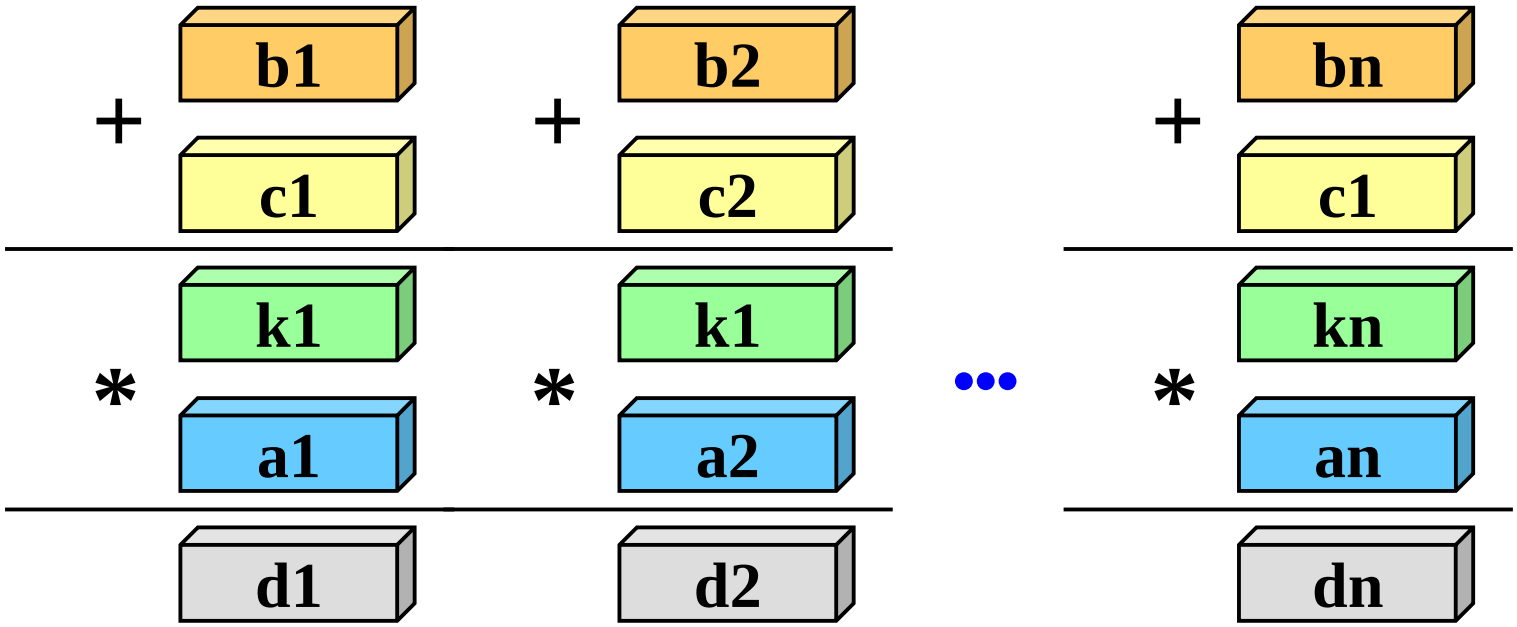
5.3.1 向量的流水处理

- 向量的处理方式主要有三种：
  - 水平（横向）处理方式
  - 垂直（纵向）处理方式
  - 分组（纵横）处理方式
- 下面以D=A*(B+C)为例说明向量的处理方式

5.3.1 向量的流水处理

- 水平（横向）处理方式
  - 即逐个求D向量元素的方法
    - 访存，取得三个操作数a_i，b_i，c_i
    - 先计算 $b_i + c_i \rightarrow k_i$
    - 再计算 $k_i * a_i \rightarrow d_i$
  - 标量机上通常采用此方式，使用循环程序实现

5.3.1 向量的流水处理



标量机上通常采用此方式，使用循环程序实现。

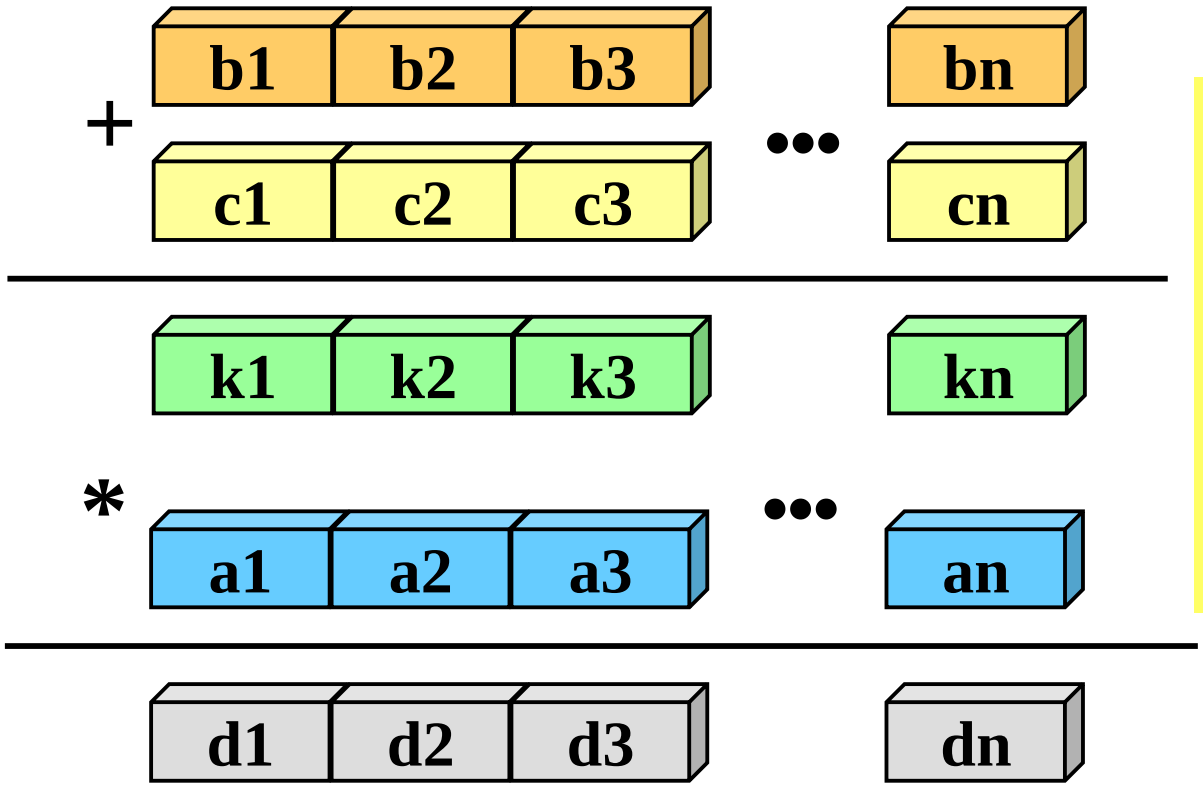
5.3.1 向量的流水处理

- 水平（横向）处理方式
  - 2个主要问题：
    - 每个D向量元素至少需两个操作（切换），使流水线难以连续流动
    - 都需要用到k，两条指令之间存在相关（先写后读）➔不适合流水处理

5.3.1 向量的流水处理

- 垂直（纵向）处理方式
  - 以向量为单位进行计算
    - 先计算所有的元素 $b_i + c_i \rightarrow k_i$（1 → N）
    - 再计算所有的元素 $k_i * a_i \rightarrow d_i$（1 → N）
  - 这样元素之间无相关，切换总共只需要一次

5.3.1 向量的流水处理



- 只要能为流水线提供连续输入，即可获得高吞吐率。
- 需要存储器具有较高的速度。

5.3.1 向量的流水处理

- 垂直（纵向）处理方式
  - 问题：
    - 是否能为流水线连续提供输入？
    - 如果主存速度不足，就无法连续提供输入

5.3.1 向量的流水处理

采用向量指令只需要2条：

VADD　B，C，T

VMUL　A，T，D

这种处理方式适用于向量处理机

数据相关不影响流水线连续工作。

不同的运算操作只需要切换1次。

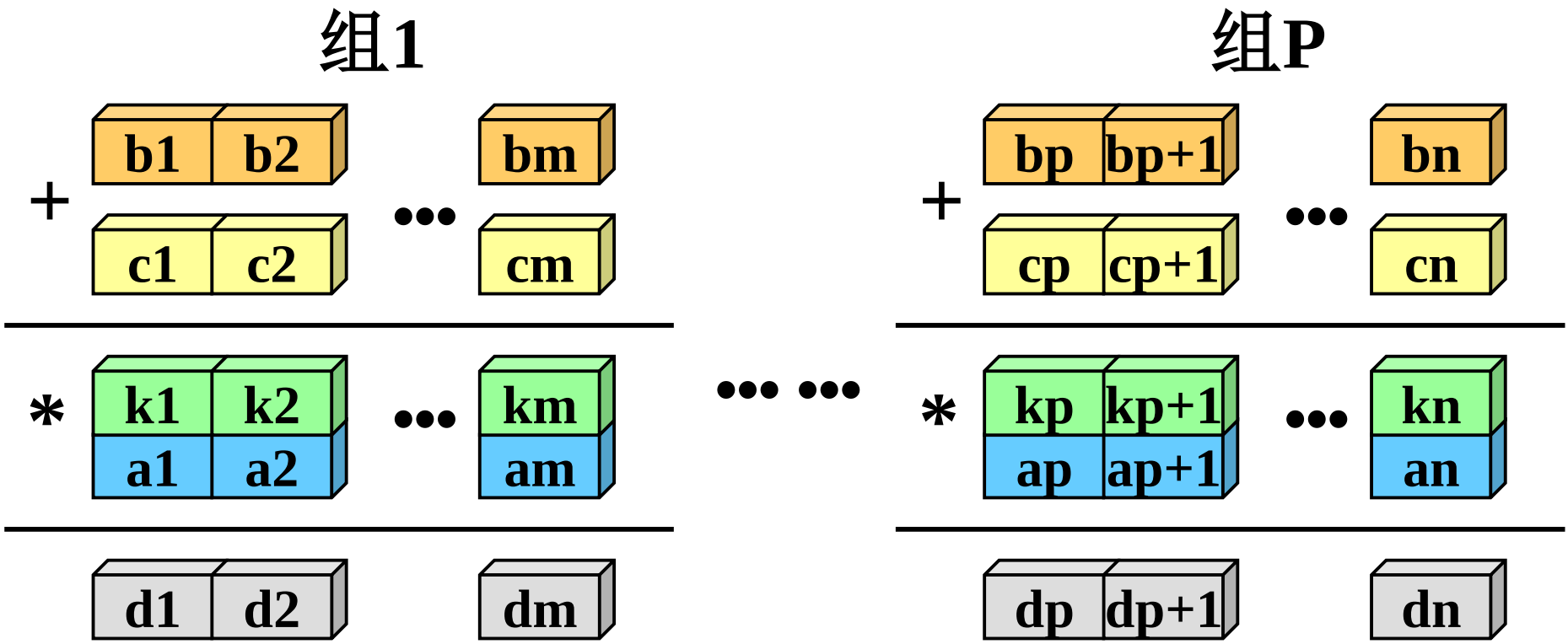
5.3.1 向量的流水处理

- 垂直（纵向）处理方式
  - 解决方法：
    - 采用多体交叉存储。这是一种面向存储器—存储器型结构的流水线处理机。由于很多通道也要使用主存，要保证连续提供数据很难，因此将主存直接连在流水线输入、输出端的做法并不是最好
    - 设置向量寄存器组。这是一种面向寄存器—寄存器型结构的流水线处理机。将流水线的输入、输出端直接连到大容量向量寄存器组，向量寄存器组与主存之间成组传送

5.3.1 向量的流水处理

- 分组（纵横）处理方式
  - 如果寄存器装得下整个向量，则采用垂直处理方式；
  - 如果向量太长，使得寄存器装不下整个向量，则将向量分组，使每组都能装入寄存器
    - 寄存器组内采用垂直处理方式
    - 寄存器组间采用水平处理方式
- 这种处理方式称为分组处理方式

5.3.1 向量的流水处理


组1
组P
+
b1 b2
c1 c2
bm
cm
bp bp+1
cp cp+1
bn
cn
*
k1 k2
a1 a2
km
am
kp kp+1
ap ap+1
kn
an
d1 d2
dm
dp dp+1
dn

5.3.2 向量流水处理机

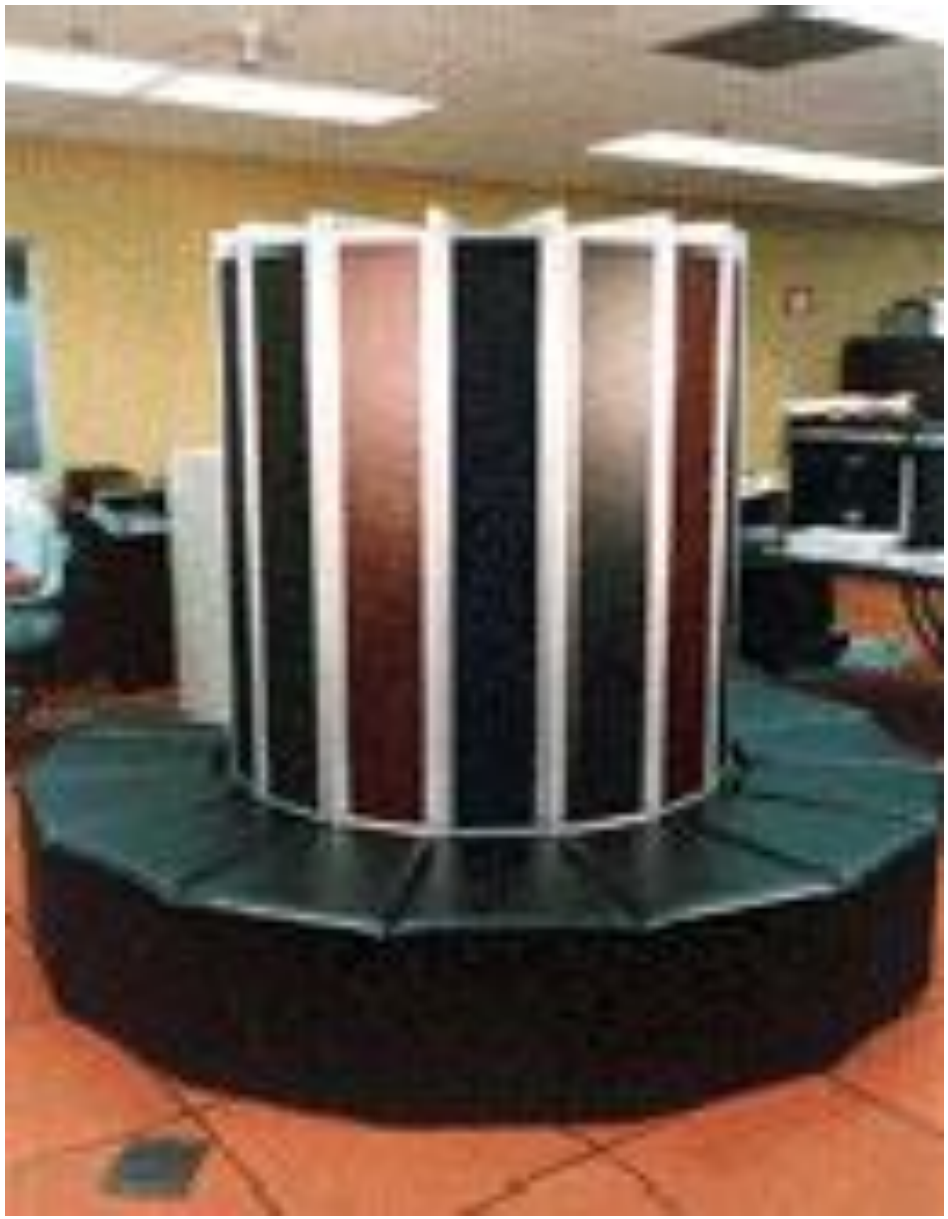
- 将向量数据表示与流水处理方式结合在一起，构成**向量流水处理机**，也称其为向量处理机。
- 向量处理机在70年代出现，经过80年代和90年代的发展，成为超级计算机的基础。

CDC STAR -100

世界上第一台使用向量处理器的计算机

Cray-1超级计算机

1976年，CRAY公司推出CRAY-1向量机，开始了向量机的蓬勃发展，其峰值速度为0.1 Gflops。



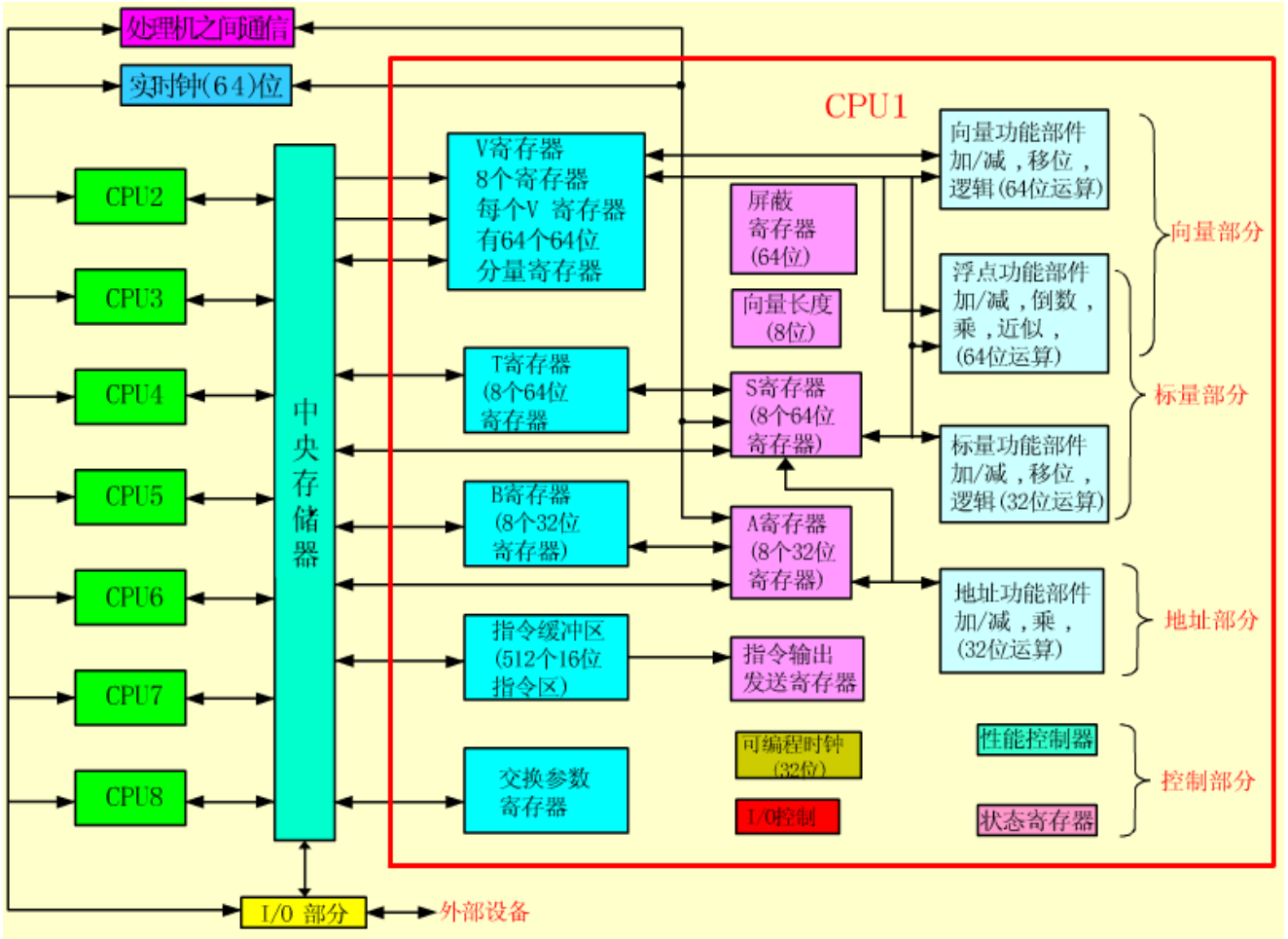
Cray 1

1985年，CRAY-2，1G flops
1990年，SX-3, 22G flops
1991年，Cray-YMP-C90，16Gflops

Cray 2

Cray XMP/4

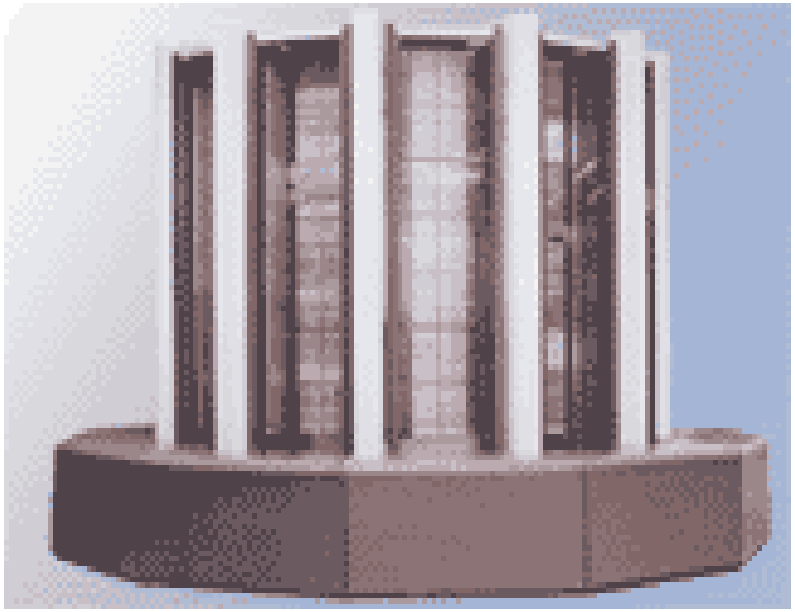
CRAY Y-MP816系统结构



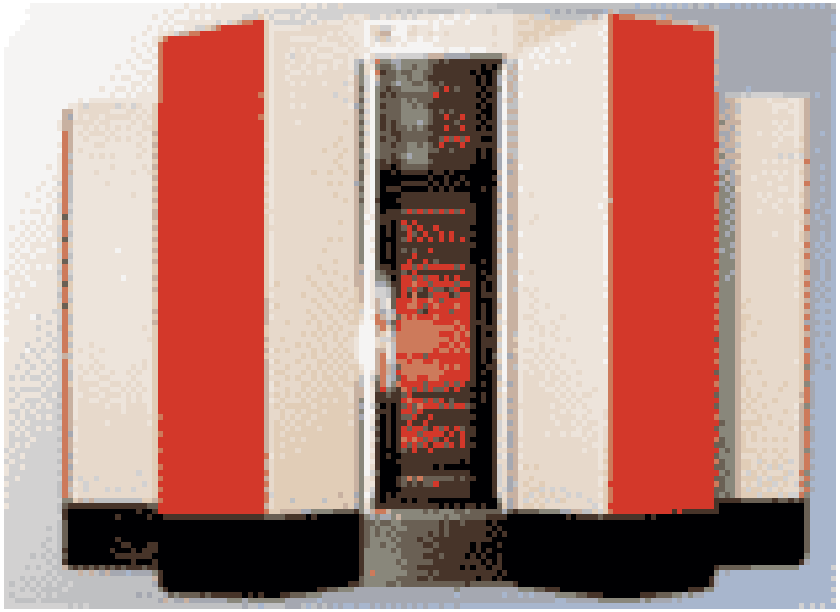
- 1991年
- 多向量处理器
- 时间并行+空间并行
- 256交叉存储
- 16MB－1GB
- 大量使用寄存器
- 64位浮点/定点

1983年12月，银河-I巨型计算机由国防科技大学计算机研究所研制成功。

1992年11月，银河-II并行巨型计算机由国防科技大学计算机研究所研制成功。



银河1



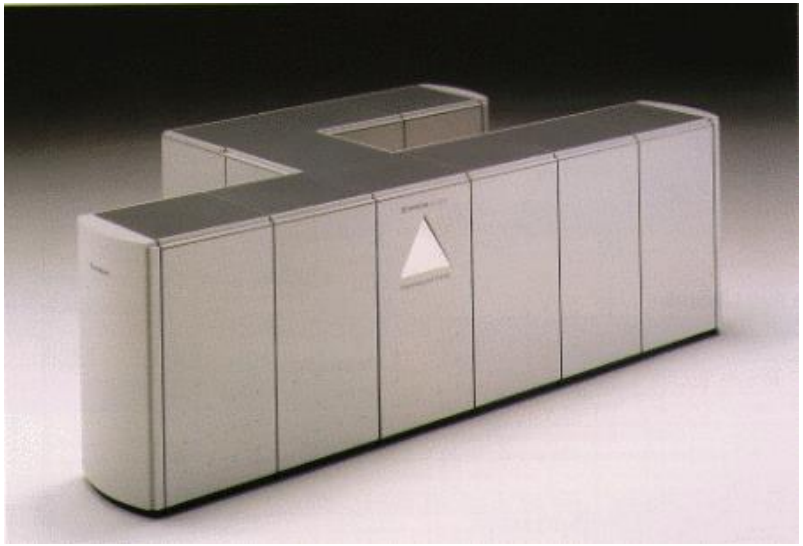
银河2

5.3.2 向量流水处理机

- **向量处理机特点：**
  - 具有向量处理指令，如向量运算、向量传送、向量压缩与恢复等，可以很有效的对向量进行处理；
  - 一般都采用流水线结构，有多条流水线并行工作；
  - 在执行向量操作时，一条指令可以同时对向量的多个元素进行运算。**向量处理机是单指令流多数据流（SIMD）处理机。**
- 向量处理机是解决数值计算问题的一种高性能计算机结构，**是向量并行计算、以流水线结构为主的并行处理计算机 。**

5.3.2 向量流水处理机

- 向量处理机通常属大型或巨型机，也可以用微机加一台向量协处理器组成。
- 一般向量计算机中包括有一台高性能标量处理机。



Turing Hitachi S3600
包括一台标量处理机和一台向量处理机

个人超级计算机

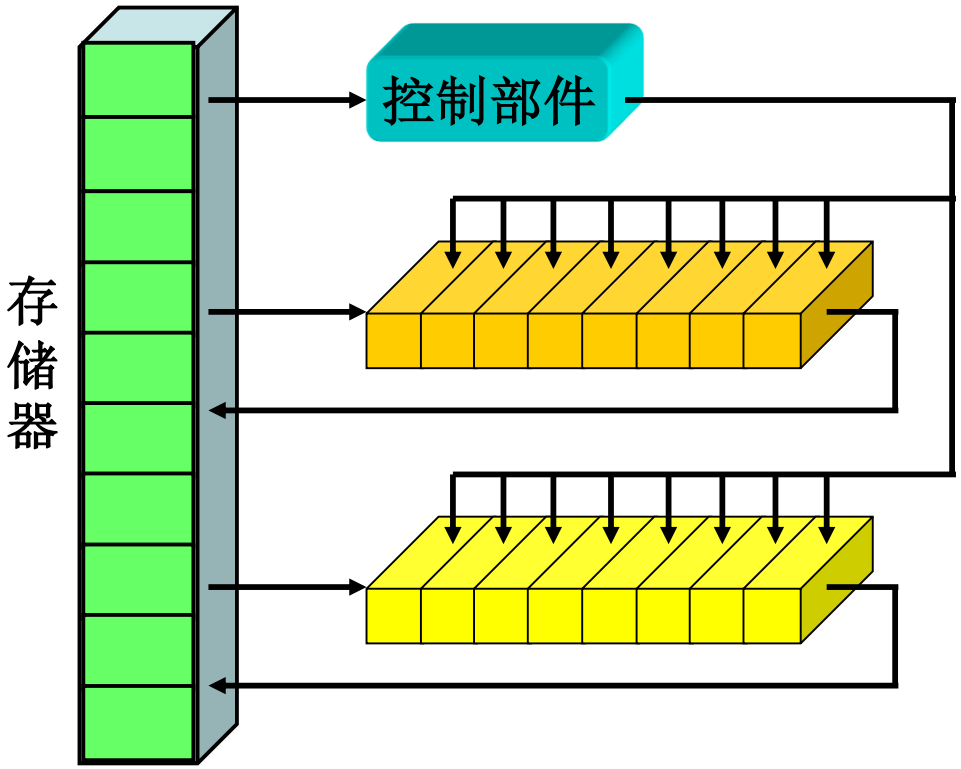
向量处理机的结构

- 向量处理机的**最关键问题**是存储器系统能够满足运算部件带宽的要求。
- 主要采用两种方法：
  - 1. 存储器－存储器结构
  - 2. 寄存器－寄存器结构

向量处理机的结构

- **1. 存储器－存储器结构**
  - 多个独立的存储器模块并行工作。
  - 处理机结构简单，对存储系统的访问速度要求很高。
- **2. 寄存器－寄存器结构**
  - 运算通过向量寄存器进行。
  - 需要大量高速寄存器，对存储系统访问速度的要求降低。

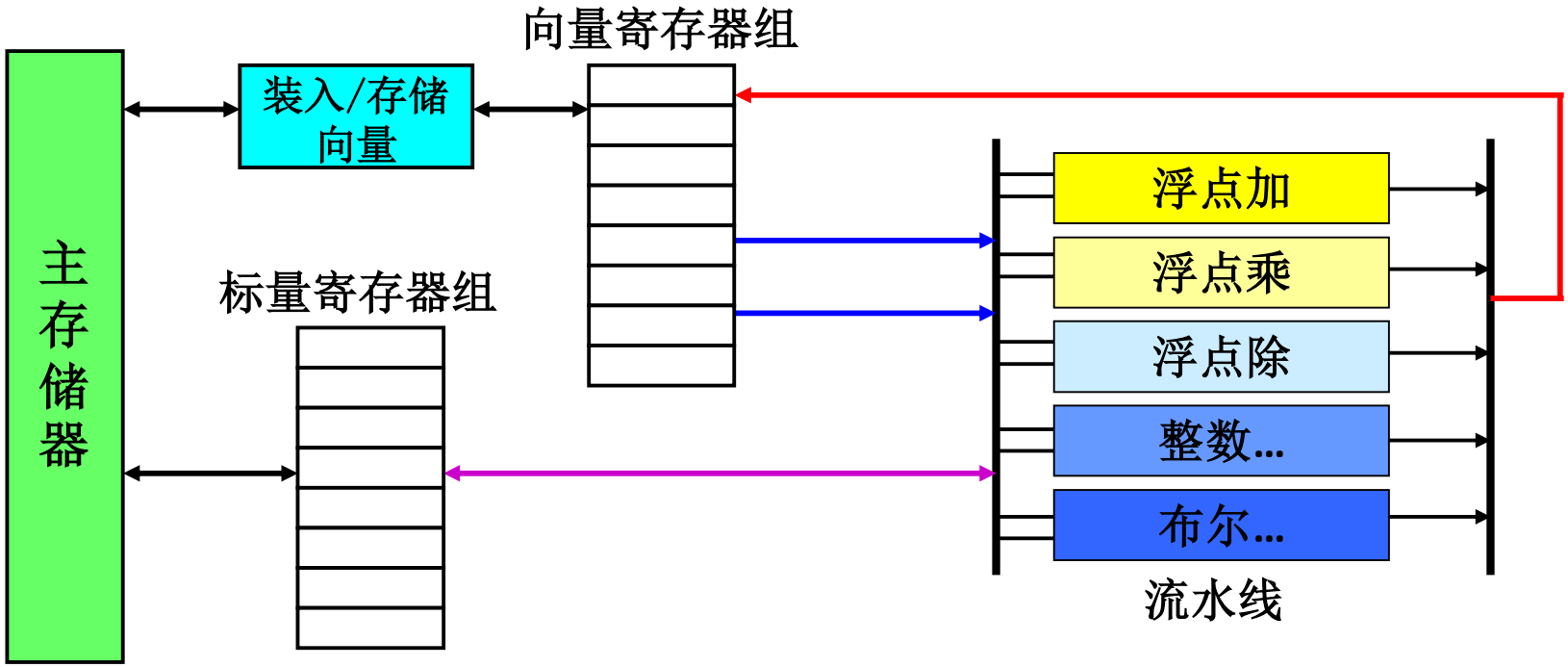
向量处理机的结构



- 为减少等待时间，存储器通常采用多体交叉器。多个独立的存储器模块并行工作。
- 向量元素存储在不同的存储器模块中，可以同时访问多个元素。

存储器 – 存储器结构的向量处理机

向量处理机的结构



寄存器 – 寄存器结构的向量处理机

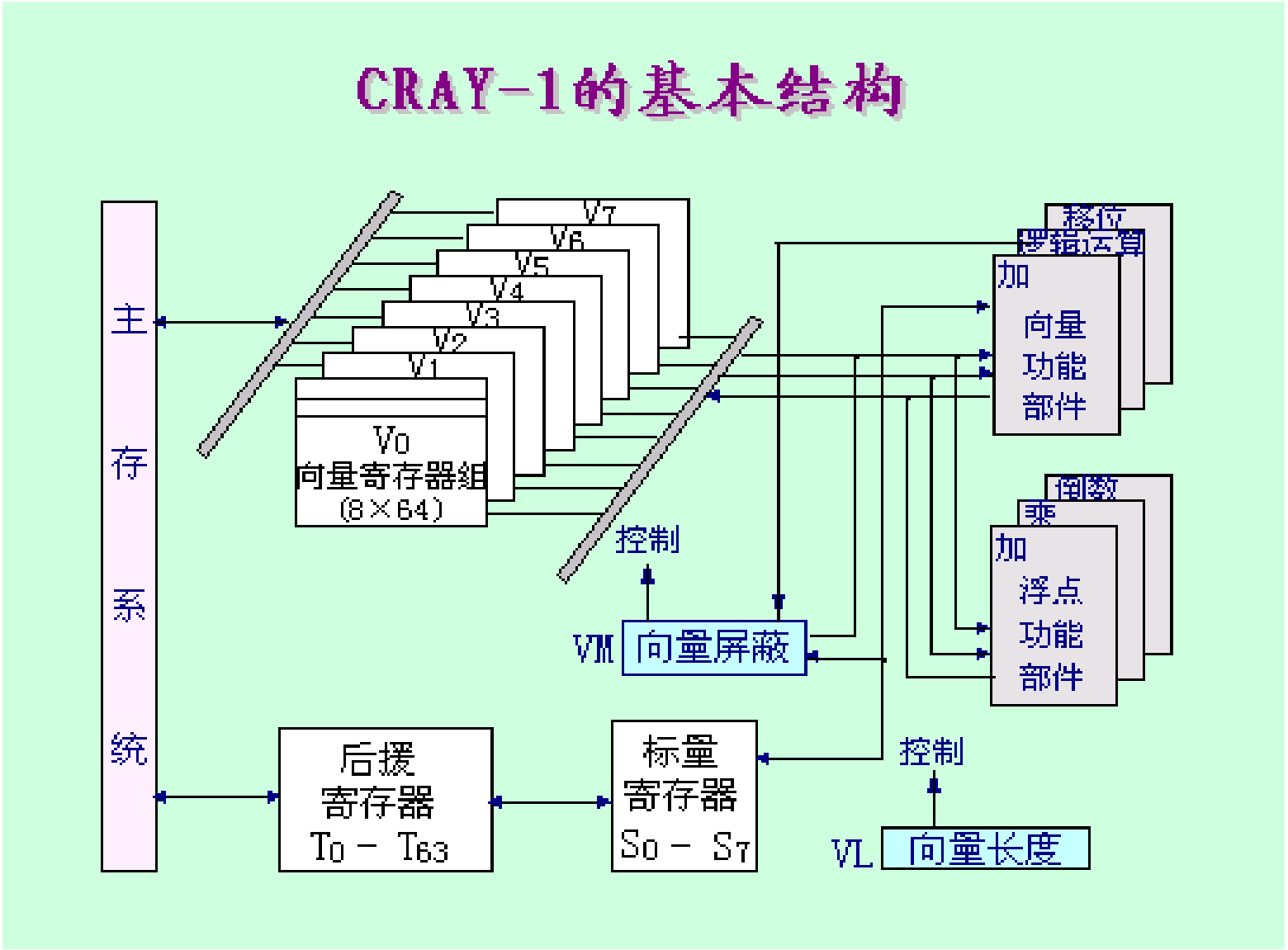
向量处理机的指令系统

- 一般包括向量型和标量型两类指令
- 向量型运算指令一般有以下几种：
  - 向量V1运算得向量V2
  - 向量V1运算得标量S
  - 向量V1与向量V2运算得向量V3
  - 向量V1与标量S运算得向量V2

向量处理机的指令系统

- 向量指令格式一般包括：
  - 操作码
  - 源或目的操作数地址
  - 地址偏移量
  - 地址增量
  - 向量长度等

CRAY-1


CRAY-1的基本结构
主
存
系
统
V7
V6
V5
V4
V3
V2
V1
V0
向量寄存器组
(8×64)
移位
逻辑运算
加
向量
功能
部件
倒数
乘
加
浮点
功能
部件
控制
VM
向量屏蔽
后援
寄存器
T0 - T63
标量
寄存器
S0 - S7
控制
VL
向量长度

CRAY-1结构

- 共有12条可并行工作的单功能流水线，可分别流水地进行地址、向量、标量的各种运算。
- 6个单功能流水部件：进行向量运算
  - 整数加（3拍）
  - 逻辑运算（2拍）
  - 移位（4拍）
  - 浮点加（6拍）
  - 浮点乘（7拍）
  - 浮点迭代求倒数（14拍）

CRAY-1结构

- 向量寄存组V
  - 由512个64位的寄存器组成，分成8块。
  - 编号：V0～V7
  - 每一个块称为一个向量寄存器，可存放一个长度（即元素个数）不超过64的向量。
  - 每个向量寄存器可以每拍向功能部件提供一个数据元素，或者每拍接收一个从功能部件来的结果元素。

CRAY-1结构

- 标量寄存器S和快速暂存器T
  - 标量寄存器有8个：S0～S7　64位
  - 快速暂存器T用于在标量寄存器和存储器之间提供缓冲。
- 向量屏蔽寄存器VM
  - 64位，每一位对应于向量寄存器的一个单元。
  - 作用：用于向量的归并、压缩、还原和测试操作、对向量某些元素的单独运算等。

CRAY-1特点

- 每个向量寄存器 Vi 都有连到 6 个向量功能部件的单独总线。
- 每个向量功能部件也都有把运算结果送回向量寄存器组的总线。
- 只要不出现 Vi 冲突和功能部件冲突，各 Vi 之间和各功能部件之间都能并行工作，大大加快了向量指令的处理。

CRAY-1特点

- **V_j冲突：**并行工作的各向量指令的源向量或结果向量使用了相同的V_j。**例如：源向量相同**

$$V_3 \leftarrow V_1 + V_2$$

$$V_5 \leftarrow V_4 \wedge V_1$$

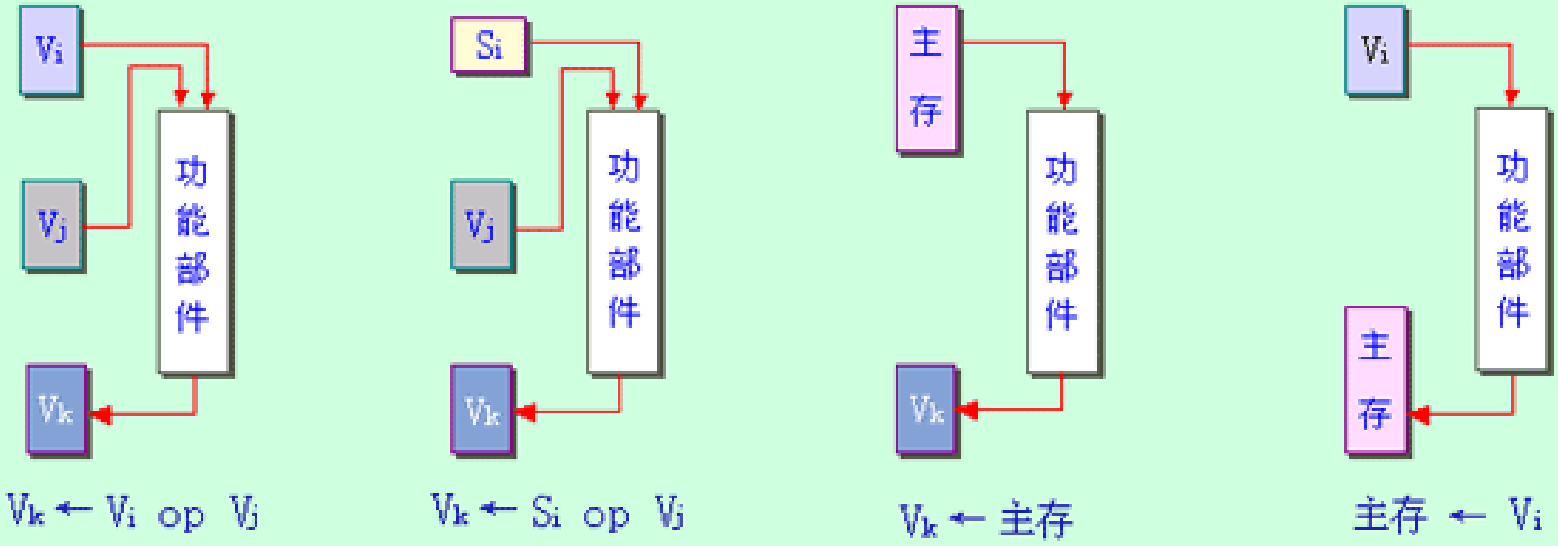
- **功能部件冲突：**并行工作的各向量指令使用同一个功能部件。**例如：都需使用乘法功能部件**

$$V_3 \leftarrow V_1 \times V_2$$

$$V_5 \leftarrow V_4 \times V_6$$

CRAY-1向量指令类型

$V_k \leftarrow V_i \text{ op } V_j$
$V_k \leftarrow S_i \text{ op } V_j$
$V_k \leftarrow$ 主存
主存 $\leftarrow V_i$



提高向量处理机性能的方法

- 设置多个功能部件，使它们并行工作。
- 向量与标量性能的平衡
  - 向量平衡点(vector balance point)：
    - 为了使向量硬件设备和标量硬件设备的利用率相等，一个程序中向量代码所占的百分比。
- 采用链接技术，加快一串向量指令的执行。
- 采用循环开采技术，加快循环的处理。
- 采用多处理机系统，进一步提高性能。