**第六章 并行处理机和互联网络**

# 6.1 并行处理机的原理

并行处理机是并行处理计算机中的一种重要结构， 它主要通过资源重复实现同时性并行。

并行处理 机也 称为 阵列处理机。

## 6.1.1 并行处理机的定义及特点

1. **并行处理机定义：**

重复设置多个同样的处理单元（PE），并将它们按照 一定方式互相连接，在统

一的控制部件（CU） 作用下，各自对分配的数据并行地完成同一条指令所规定

的操作，实现操作级并行的 SIMD 计算机。

1. **并行处理机由五个部分组成：**

一个控制器CU 、多个处理单元 PE 、多个存储器模块 M 、一个互连网络 ICN、 一台输入输出处理机 IOP

* 从**CU**看，指令是串行执行的。
* 从**PE**看，数据是并行处理的。

1. **并行处理机的主要特点：**

速度快、模块性好、可靠性高、效率低（与流水线、向量处理机相比）、依赖互连网络和并行算法、需要有高性能的标量处理机。



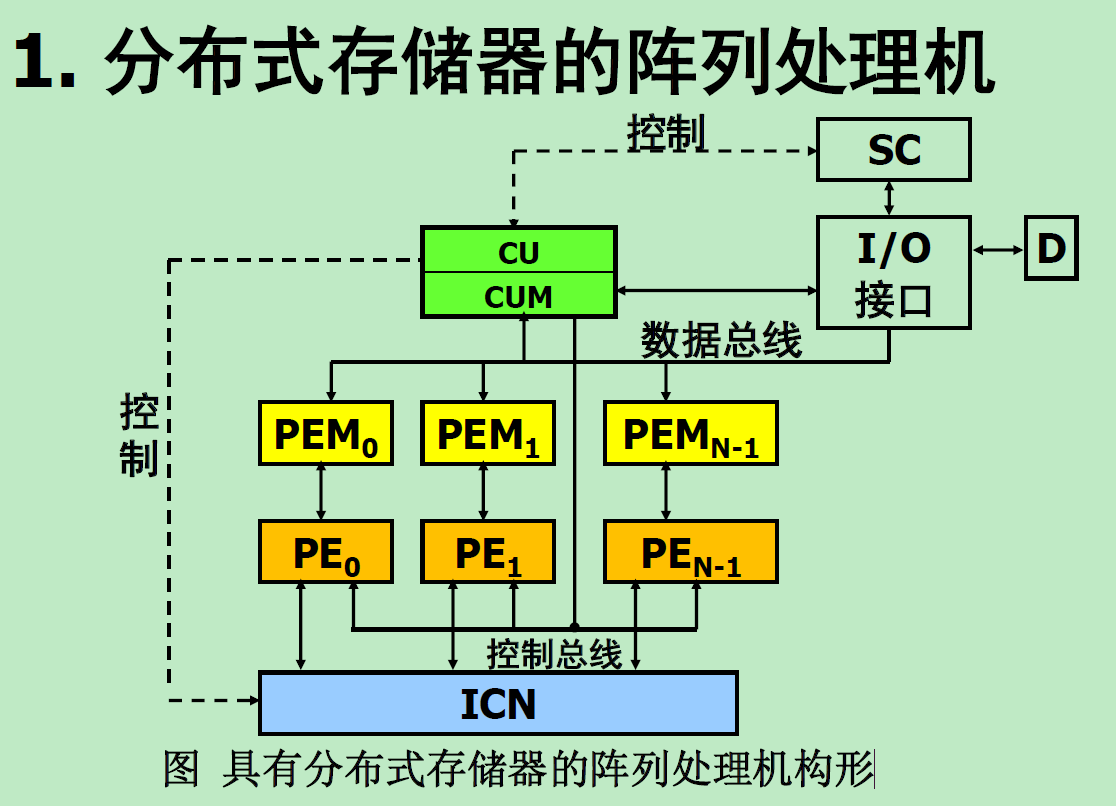
* 1. 速度快 ，而且潜力大
  2. 模块性好 ，生产和维护方便。
  3. 可靠性高 ，容易实现容错和重构。
  4. 效率低 （与流水线处理机、向量处理机等比较）。通常作为 专用计算机 ，因此，在很大程度上依赖于并行算法。
  5. 依赖于互连网络和并行算法。互连网络决定了PE之间的连接模式，也决定了并行处理机能够适应的算法。
  6. 需要有一台高性能的标量处理机。使向量计算与标量计算平衡。

## 6.1.2阵列处理机基本构形与特点

1. 根据存储器组成方式的不同 阵列处理有 两种基本构形：

分布式存储器 的阵列处理机 集中式共享存储器 的阵列处理机

1. 分布式存储器的阵列处理机：
   1. 每个**PE**有自己的局部存储器**PEM**；**PEM**只能被本**PE**访问；
   2. 整个系统在**CU**控制下运行；
   3. **所有指令都在CU中译码。**译码后把适合于并行处理的向量类指令“播送”给各个**PE**，让处于“活跃”的那些**PE**并行地执行。因此**CU中的指令基本上是单指令流**；
   4. CU 也可以采用流水线工作方式 进一步让多条向量指令在时间上重叠执行；
   5. 为了有效地对向量数据进行高速处理要求能把数据合理地预分配到各个 PEM中；**PE 之间通过互连网络 ICN 来交换数据**



* 1. 这种构形是**SIMD**的主流

比较容易构成**MPP**（**Massively Parallel Processor**），几十万个**PE**。

必须依靠并行算法来提高**PE**的利用率。因此，应用领域很有限。

1. 集中式共享存储器的阵列处理机
   1. 与分布式存储器的阵列处理机相比 有两处不同 ：
      1. 系统存储器由 K 个存储体构成 经互连网络 ICN 为全部 N 个 PE 所共享 。 为使各个PE 对长度为 N 向量的各个元素同时处理存储体数 K 应等于或大于 PE 数 N 即 K>=N 。 为避免访存冲突 需要合适的算法将数据合理地分配到各个存储体中.
      2. ICN 的作用不同 。 ICN 用于在 PE 和 MM之间构建数据通路 使各个 PE 可以高速 、灵活 、 动态地 与不同的存储体相连 。 因此有时被称为 **对准网络**
2. 阵列处理机的特点：
   1. 阵列机以**单指令流多数据流**方式工作。
   2. 阵列机采用**资源重复**方法引入空间因素.即在系统中设置多个相同的处理单元来实现并行性 这与利用时间重叠的向量流水处理机是不一样的 。 此外 阵列机利用并行性中的同时性 所有处理单元必须同时进行相同的操作 。
   3. 它使用简单而规整的ICN来确定PE之间的连接模式。ICN限定了并行处理机适用的解题算法的类型,也对整个系统的性能产生明显影响。因此 **ICN是设计重点**。
   4. 阵列机是以某一类算法（如图像处理）为背景的专用计算机。这是由于阵列机中通常都采用简单、规整的互连网络来实现处理单元间的连接操作，从而限定了其所适用的求解算法类别。
   5. 阵列机的研究必须与并行算法的研究密切结合，以使其求解算法的适应性更强一些，应用面更广一些。
   6. 从处理单元来看 由于结构都相同因而可将阵列机看成是一个同构型并行机。但其控制器实质上是一个标量处理机而**为了完成I/O操作及操作系统管理**尚需一个**前端机**。因此实际的阵列机系统是由上述三部分构成的一个异构型多处理机系统。

# 6.2 阵列处理机的并行性算法

**ILLIAC IV 是最先采用SIMD 结构的并行机。**

* **ILLIAC IV组成： ILLIAC IV 阵列 及 ILLIAC IV 输入输出系统**
* **ILLIAC IV是由3种类型处理及组成的多机系统：**
  + **专门对付数组运算的 处理单元阵列 (processing element array)**
  + **阵列控制器 (array control unit)，它既是处理单元阵列的控制部分，又可以视为一台相对独立的小型标量处理机；**
  + **一台 标准的 Burroughs B6700 计算机 ，担负 ILLIACIV 输入 输出系统和操作系统管理功能。**

**累加和： log2N次加法，加速比N/log2N.(** **这种方法也称为 级联求和 或 递归求和 。**

* **与流水线中采用的方法相同， 它利用结合律来提高并行度。)  
  SIMD与向量计算机的区别**



# 6.3 互连网络的基本概念

**ICN**限定了并行处理机适用的解题算法的类型，也对整个系统的性能产生明显影响

**ICN**结构复杂性：反映成本

**ICN**结构灵活性：反映性能

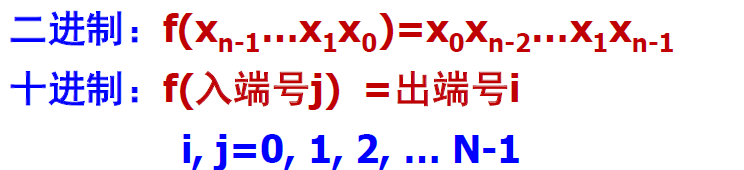
## 6.3.1 互连网络的设计目标与互连函数

1. 互连网络的定义（ICN）： 互连网络是一种由**开关元件**按照一定的拓扑结构和控制方式构成的网络，用来实现计算机系统内部多个处理机或多个功能部件之间的相互连接。
2. 互连网络的作用：

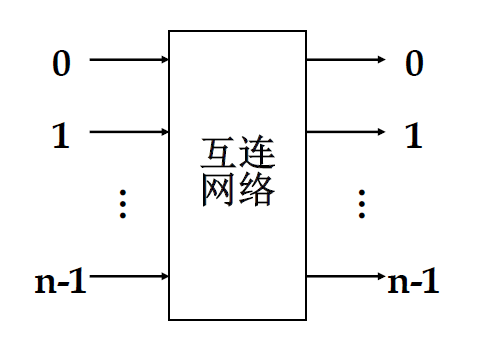
用来实现计算机系统内部多个处理机或多个功能部件之间的相互连接。互连网络已经成为并行处理系统的核心组成部分。对整个计算机系统的性价比有决定性影响。

1. 互连网络的表示方法：

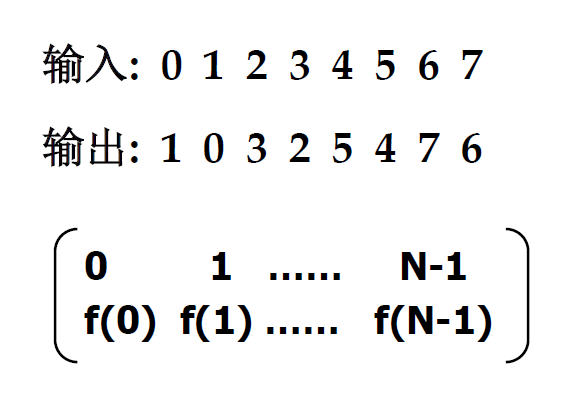
**（1）互连函数表示法**



（2）图形表示法



（3）输入输出对应表示法



1. 互连网络的主要特性（互连网络是设计重点！）

（1）网络规模：网络中结点个数

（2）结点度：与结点相连接的边数（=入度+出度）

（3）距离：两个结点之间相连的最少边数

（4）网络直径：网络中任意两个结点间距离的最大值

（5）结点间的线长：两个结点间连线的长度

（6）等分宽度： 某一网络被切成相等的两半时，沿切口的最小边数称为该网络的等分宽度

（7）对称性：从任何结点看到拓扑结构都是一样的网络称为对称网络

1. 互连网络消息传输：
   1. 发送方：
      1. (1) 用户程序把要发送的数据拷贝到操作系统的缓冲区。
      2. (2) 操作系统把缓冲区中的数据打包，并发送到网络接口部件。
      3. (3) 网络接口硬件开始发送消息。
   2. 接收方:
      1. (1) 把数据包从网络接口部件拷贝到操作系统缓冲区。
      2. (2) 检查收到的数据包，如果正确，给发送方发回答信号。
      3. (3) 把接收到的数据拷贝到用户地址空间。
2. **互连网络在传输方面的主要参数：**

（1）频带宽度（每秒传输的比特数）

（2）传输时间（消息长度/带宽）

（3）飞行时间（第一位信息到达接收方所需的时间）

（4）传输时延（飞行时间+传输时间）

（5）发送方开销（处理器将消息放到互连网络消耗的时间）

（6）接收方开销（处理器将消息从互连网络取出消耗的时间）

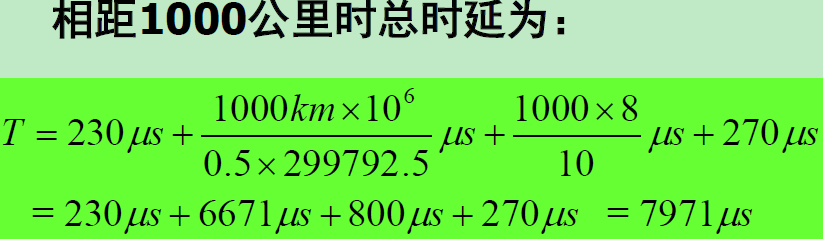
**总时延=发送方开销+飞行时间+消息长度/带宽+接收方开销**

假设一个网络的频宽为 10Mb/s ，发送方开销为 230µs ，接收方开销为 270 µs。如果两台机器相距 100 米，现在要发送一个 1000 字节的消息给另一台机器，试计算总时延。如果两台机器相距 1000公里，那么总时延为多大？

总时延 = 发送方开销+飞行时间+消息长度/带宽+接收方开销

= 230µs + 100/（0.5\*3\*10^8\*10^6）+1000\*8/10Mbps+270µs

= 1301µs



## 6.3.2 设计互连网络时应考虑的问题

1. 确定 PE 之间通信的互连网络的因素

包括：

**操作方式 控制策略 交换方法 网络拓扑结构等**

* 1. **操作方式**
     1. **同步**
     2. **异步**
     3. **同步/异步**
     4. **现在的阵列处理机均采用同步操作方式 让所有PE按时钟同步 。**
     5. **多处理机一般采用异步或同步 异步组合方式**
  2. **控制策略**
     1. **典型的 ICN 由开关单元和互连线路组成**
     2. **互连通路的路径选择通过设置开关单元的状态来控制**
     3. **开关单元的状态的设置有 两种控制策略：**
        1. **集中控制策略**
        2. **分布控制策略**
     4. **多数 SIMD 的 ICN 采用集中控制策略**
  3. **交换方法**
     1. **线路交换**
     2. **包交换**
     3. **线路/包交换**
     4. **SIMD 的 ICN 一般采用线路交换 。**
     5. **多处理机和计算机网络一般采用包交换**
  4. **网络拓扑结构**
     1. **网络拓扑结构指互连网络输入 、 输入端可以实现连接的模式**
     2. **有两种：**
        1. **静态**

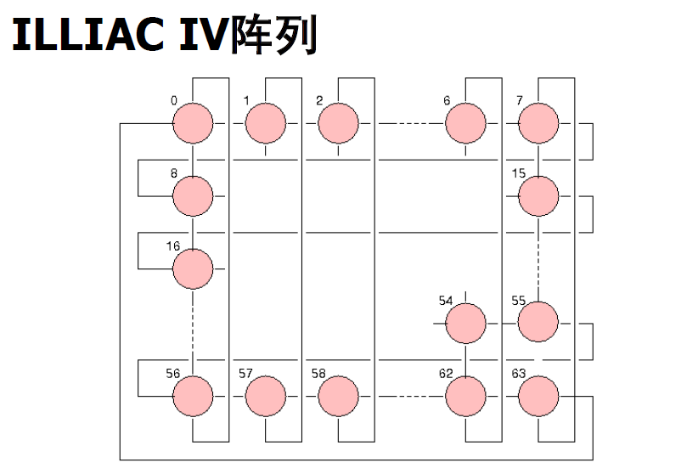
**在静态连接模式中两个PE之间的连接是固定的不能重新配置成与其他 PE连接**

* + - 1. **动态：单级 和 多级**

**在动态连接模式中PE 之间的连接可以通过设置开关单元的状态重新配置**

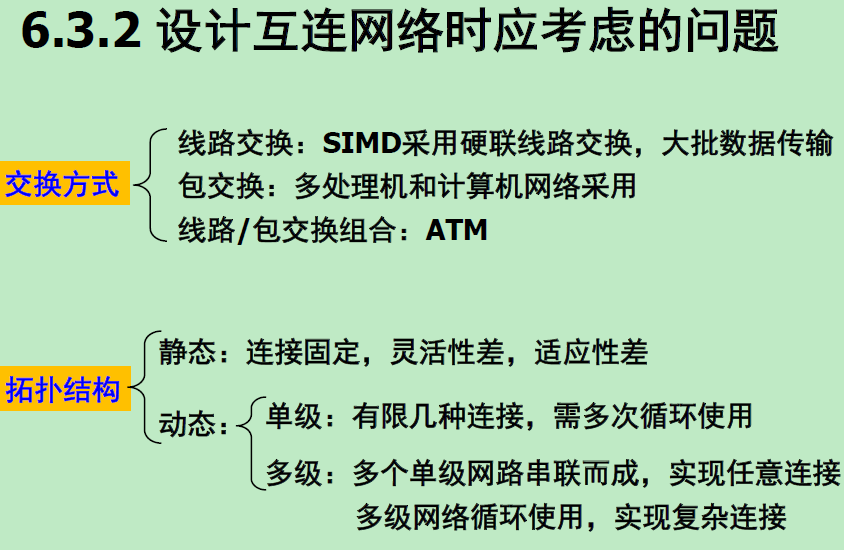
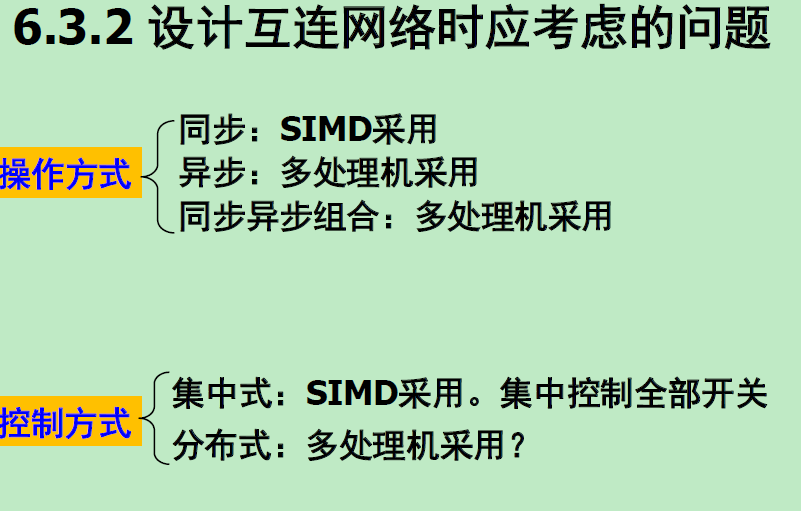
7.SIMD的ICN一般采用线路交换（大批量数据），多处理机和计算机网络通常用包交换（短数据）。

8.ILLIAC IV阵列



N\*N的二维阵列中，任意两个结点间最短距离不会超过（N-1）

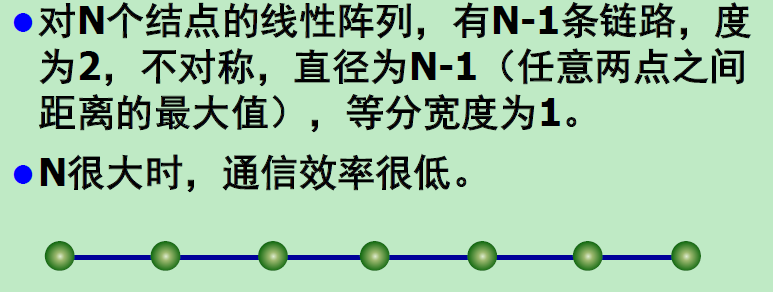
小结：



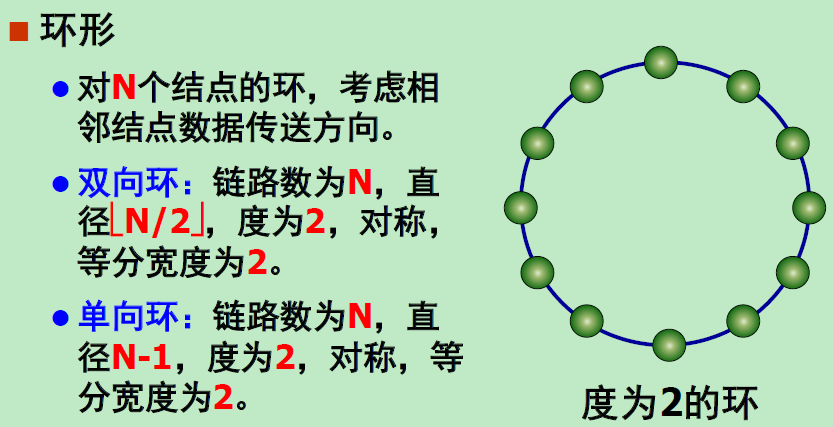
# 6.4 互连网络的种类

## 6.4.1 静态互连网络

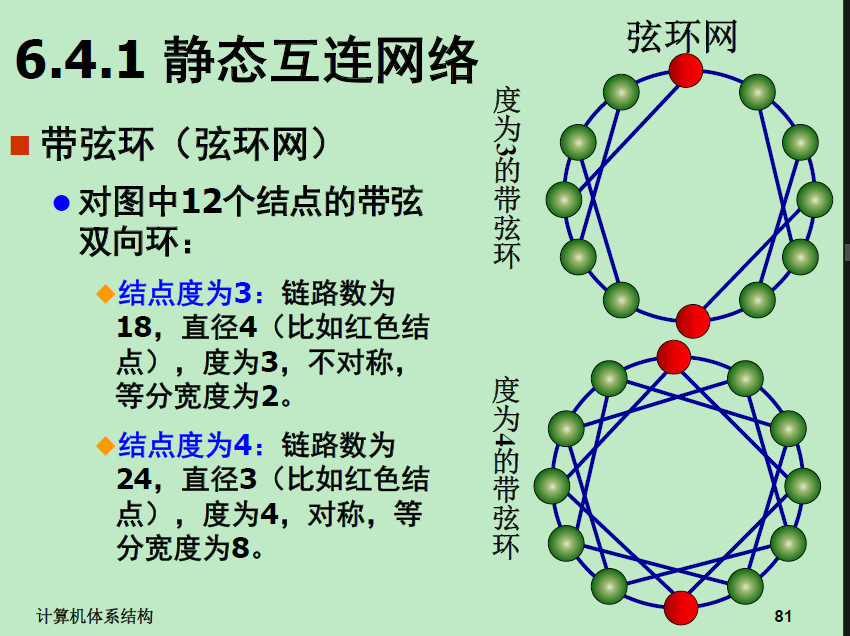
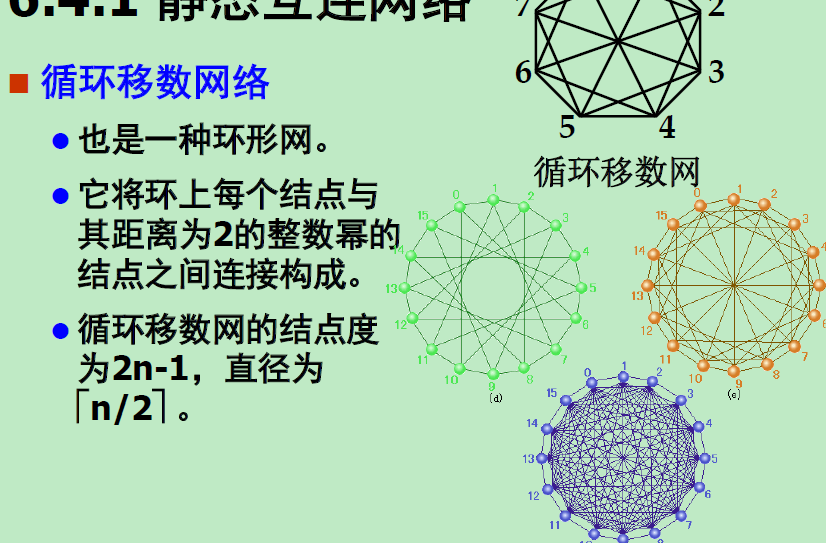
1. 一维线性阵列



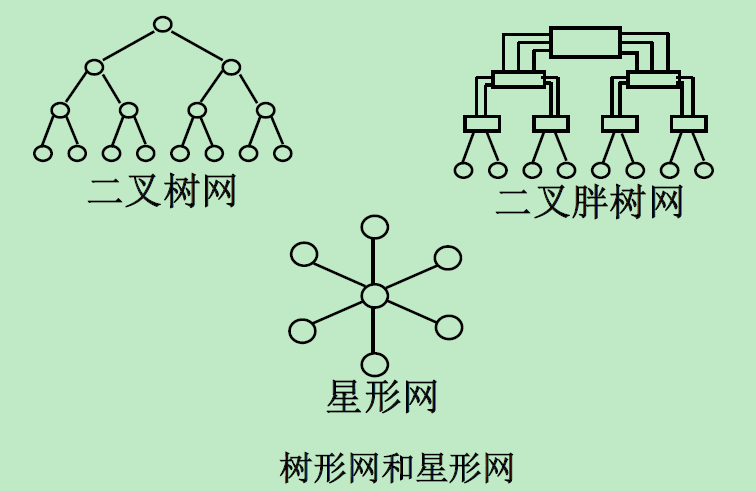
1. 环形



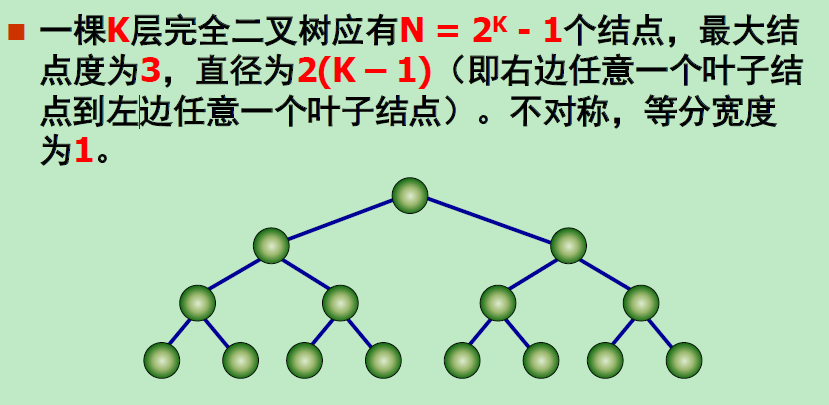
1. 弦环网和循环移数网络

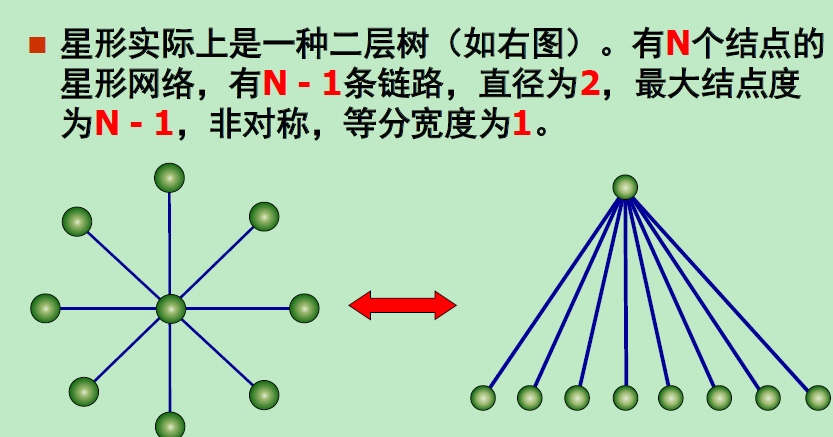
1. 其他



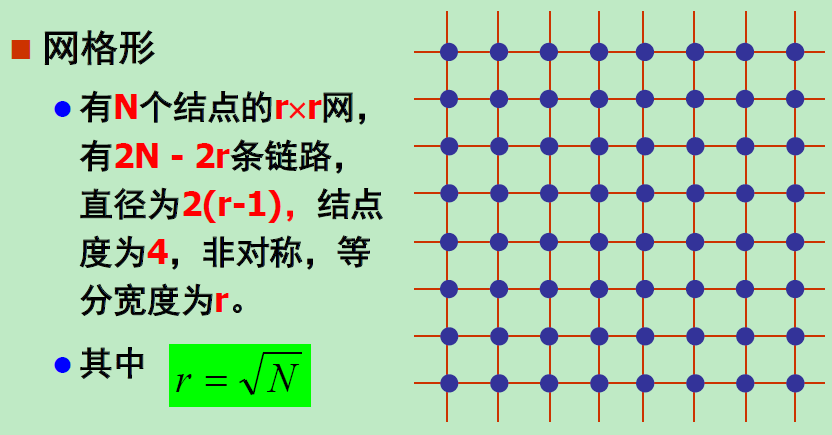
1. 二叉树



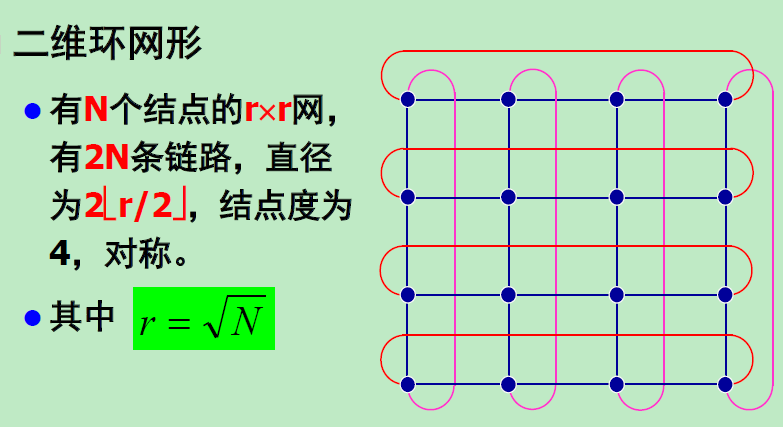
1. 星型扩展



1. 网格型



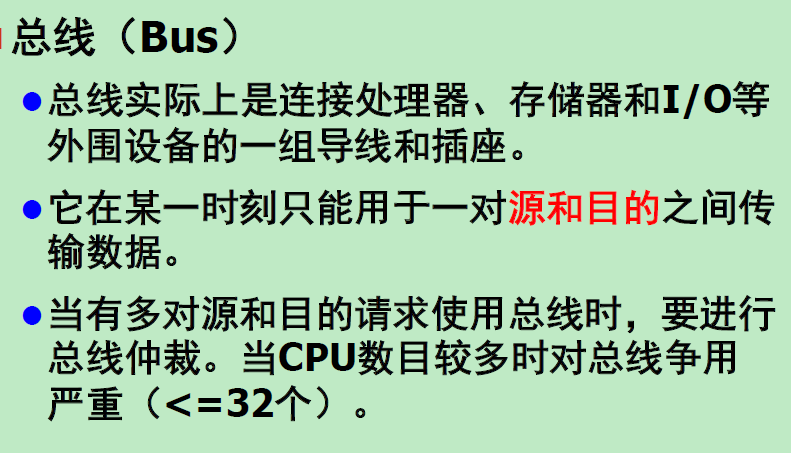
1. 二维环型网



## 6.4.2 动态互连网络

* 动态网络中的连接不固定，在程序执行过程中可根据需要改变。
* 网络的开关元件有源，链路可通过设置这些开关的状态来重构。
* 只有在网络边界上的开关元件才能与处理机相连。
* 动态网络主要有：总线、交叉开关、循环网络，多级交换网络

1. 总线

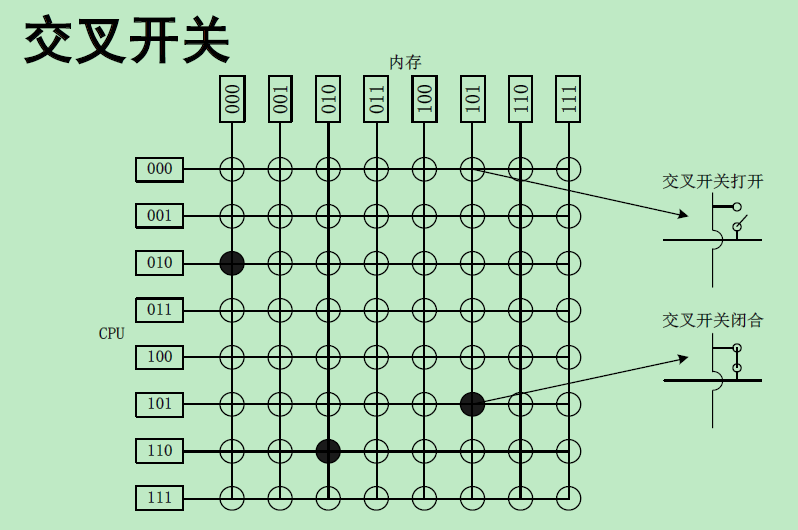


* 1. 线性阵列与总线的区别
  2. 线性阵列： 允许不同的源结点和目的结点对并发 使用系统的不同部分。
  3. 总线： 通过切换与其相连的许多结点来实现时分特性，同一时刻只有一对结点在传送数

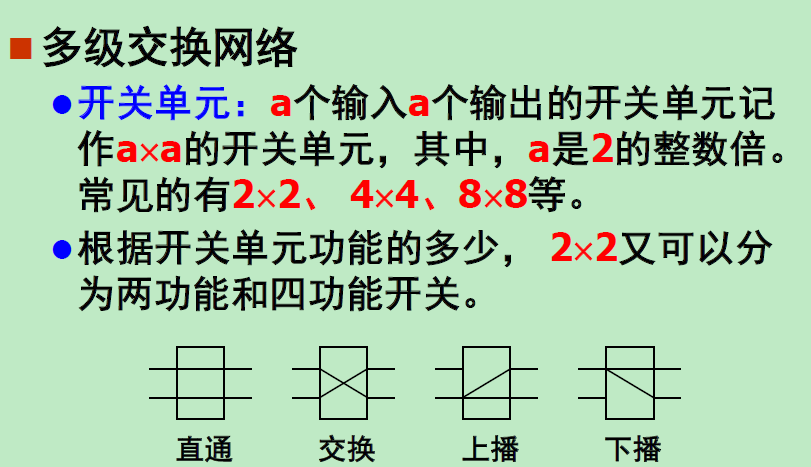
1. 交叉开关（ Crossbar Switcher



* 1. 交叉开关是一种高带宽网络，它可以在输入端和输出端之间建立动态连接.
  2. 在每个输入端和输出端的交叉点上都有交叉点开关。该开关可以根据需要置为 开 或 关 状态，从而使不同的输入端和输出端导通
  3. 交叉开关的硬件复杂性为 n^2 数量级，造价昂贵。**但是其带宽和寻径性能在这三种动态网络中最好**
  4. 如果网络规模小，它是一种理想的选择（ （<=64个）



1. 循环网络
   1. 将同一套单级互连网络循环使用，组成循环互连网络。
   2. 经过多次循环，实现 PE 之间的连接。
2. 多级交换网络



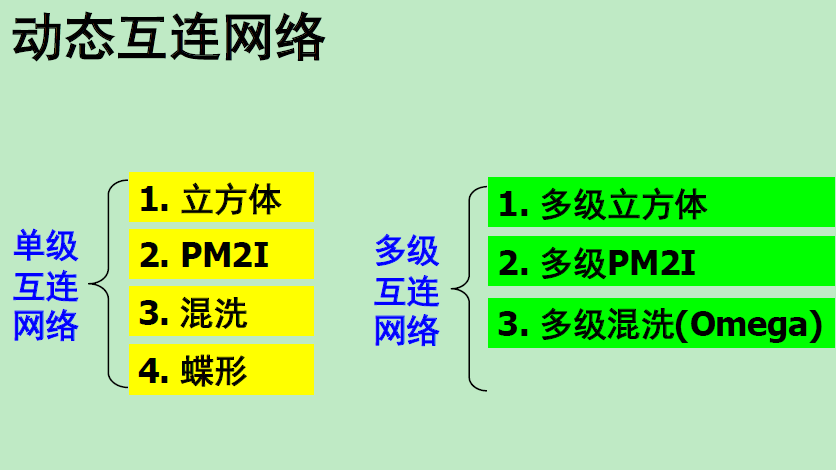


* 级间互连模式
  + 混洗、蝶式、纵横开关及立方体连结等
* 控制方式
  + 级控制： 每级只有一个控制信号
  + 单元控制： 每个开关一个控制信号
  + 部分级控制： 几个开关合用一个控制信号



多级交换网络是总线和交叉开关的折衷。它的主要优点在于采用模块结构，可扩展性好（ >64）

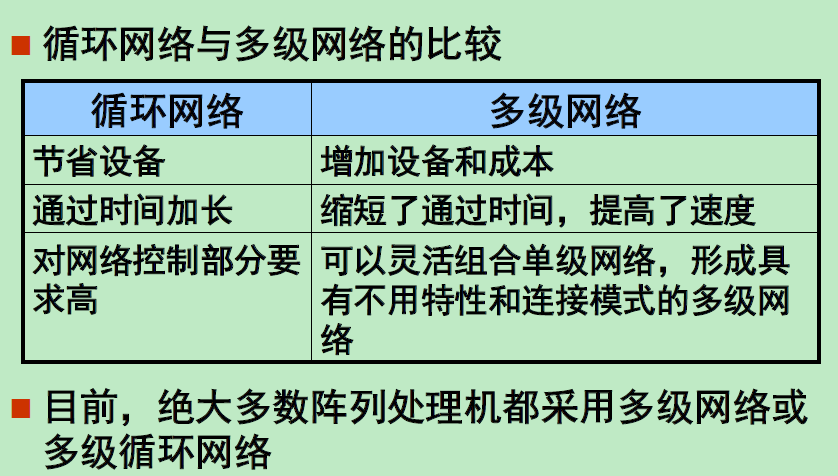
**MIMD和 SIMD 计算机都使用多级网络**



## 6.4.3循环互连网络

* 单级互连网络只能实现有限几种基本连接，并不能实现任意处理器之间的互连
* 有两种解决办法：
  + 循环互连网：多次重复使用同一个单级互连网络。**牺牲时间换取设备**
  + 多级互连网：将多套相同单级互连网络连接起来。**以设备换取时间**

循环互连网络与多级互连网络相比，节省了重复的设备，但加长了通过时间。

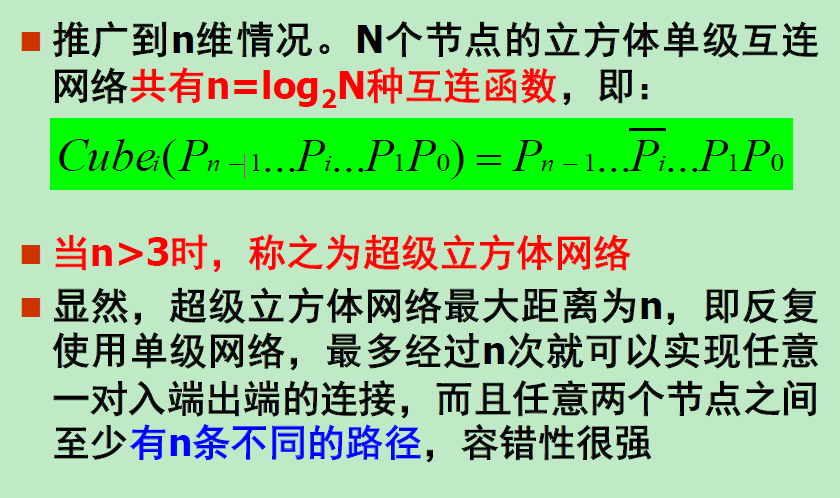
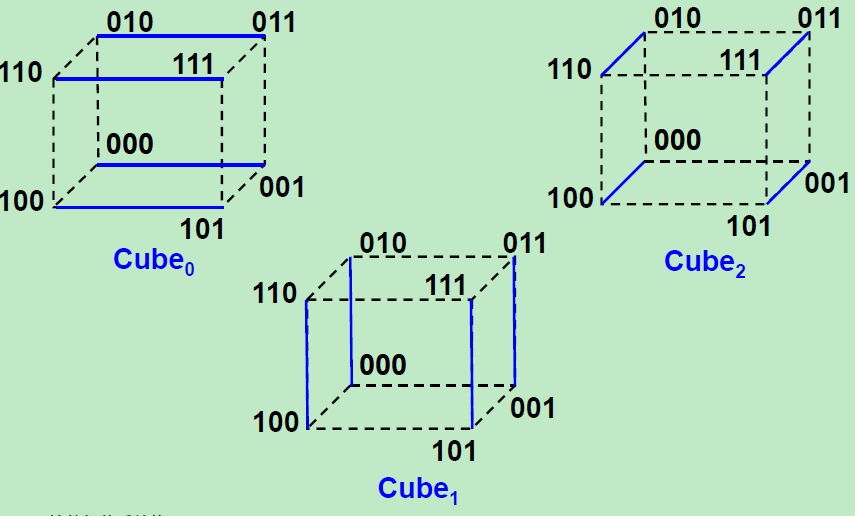


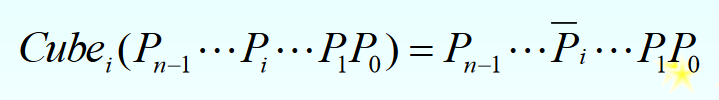
目前 绝大多数阵列处理机都采用多级网络或多级循环网络

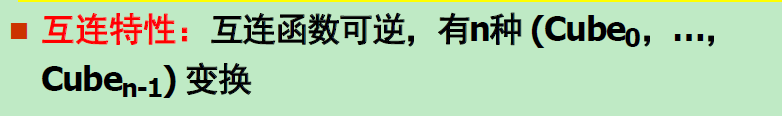
## 6.4.4 基本的单级互连网络

* 立方体 单级互连网络
* PM2I单级互连网络
* 混洗交换 单级互连网络
* 蝶形 单级互连网络

1. 立方体单级网络

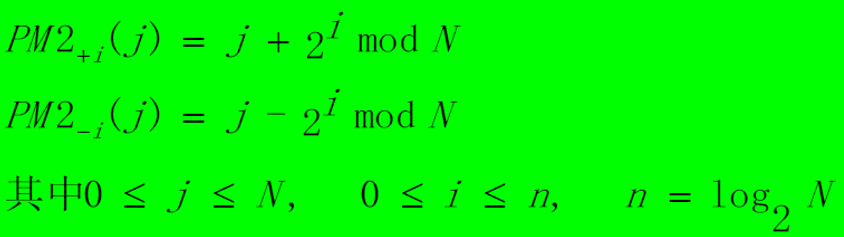
 

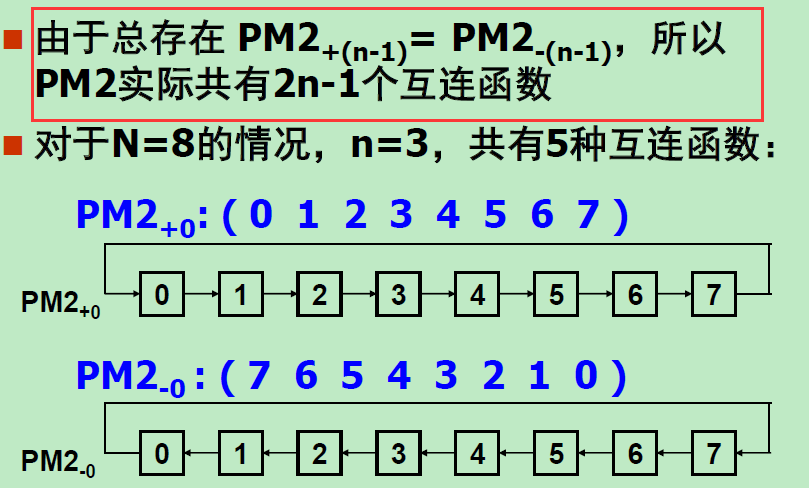


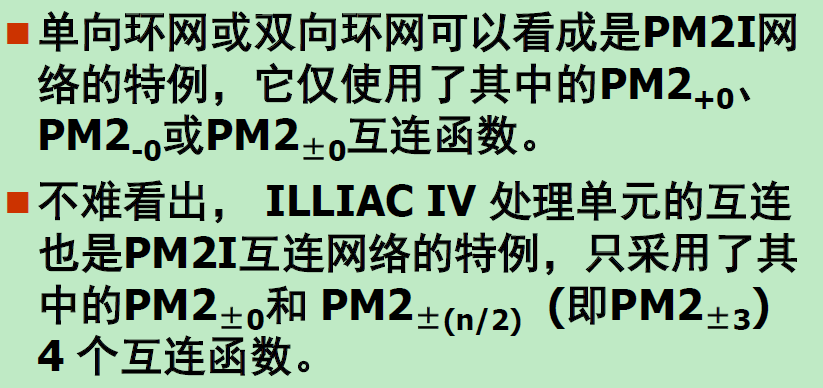


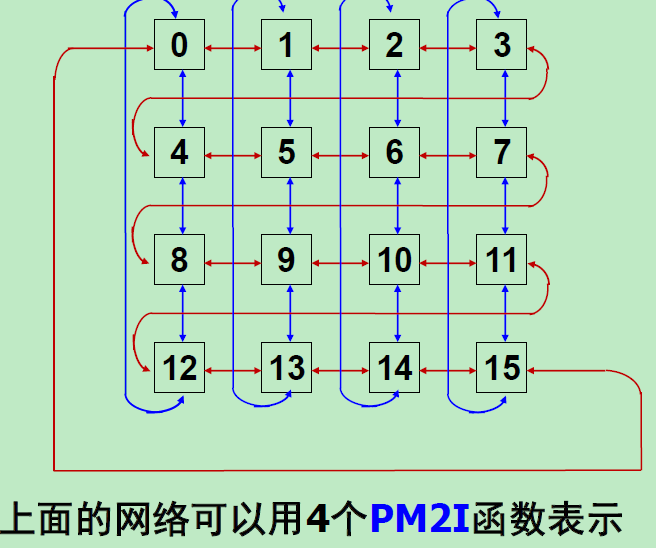
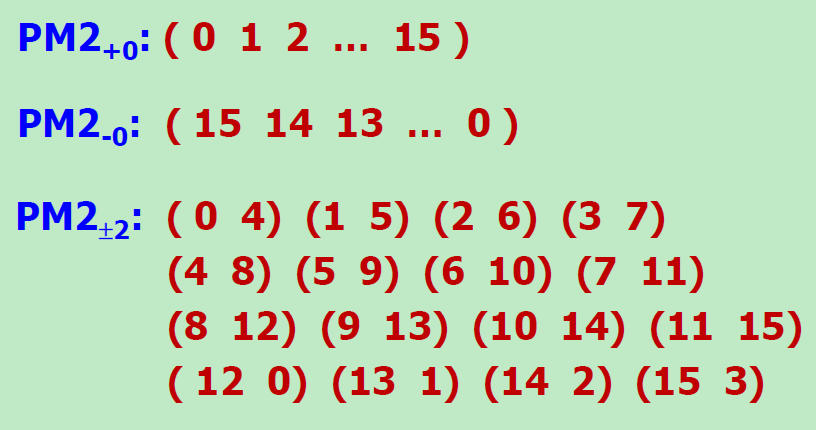
1. PM2I单级互连网络

**PM2I**是**Plus-Minus2(**加减**2)**单级网络的简称，能实现**j**号**PE**直接与**j**±**2**号**PE**连接，即







* **PM2I**单级网络的最大距离为**[n/2]**。
* 对于**n=3**，最多使用两次，既可以实现任意一对入端出端的连接

1. 混洗交换单级互连网络

混洗交换 shuffle Exchange单级互连

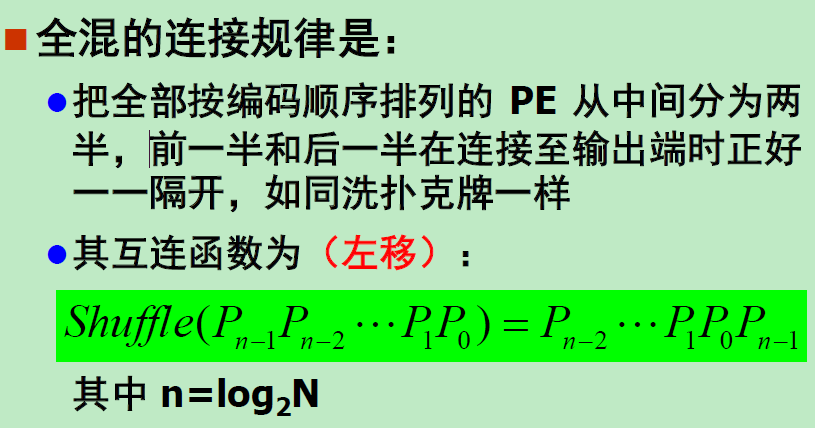
网络包含两个互连函数：

全混 Perfect Shuffle

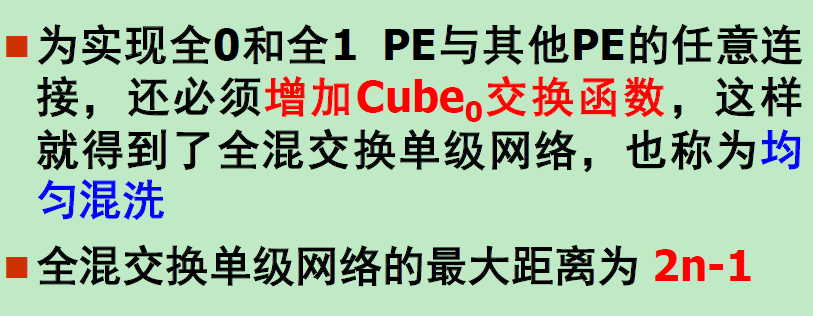
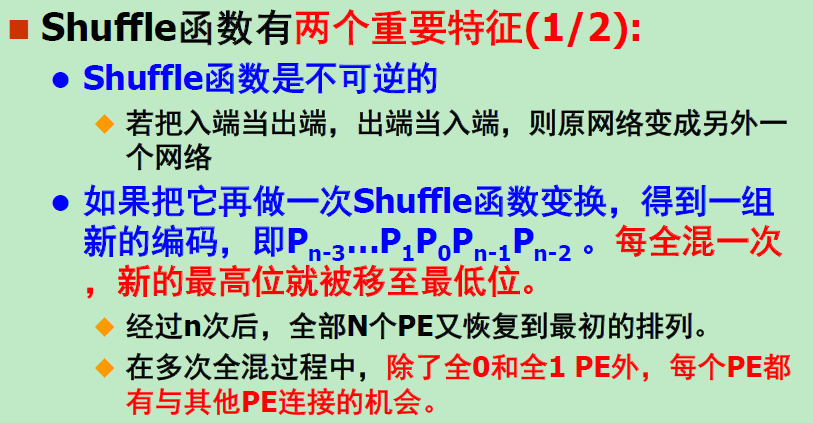
交换 Exchange

**先混洗 再交换**

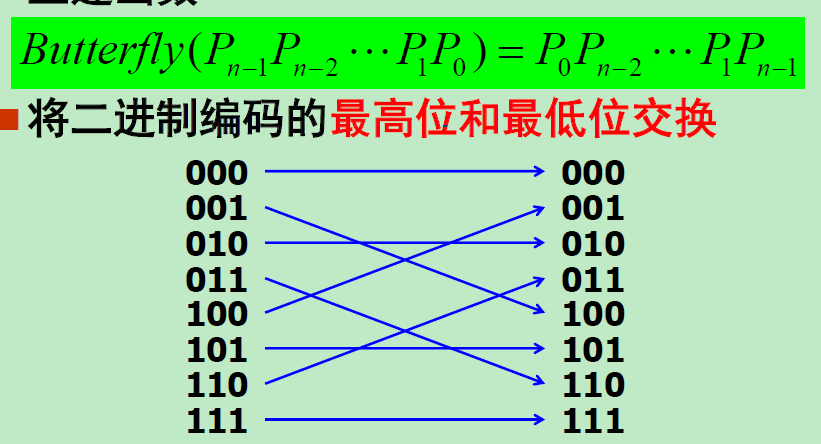
* **全混**



* Shuffle函数



1. 蝶形单级互连网络（非重点）



## 6.4.5 多级互连网络

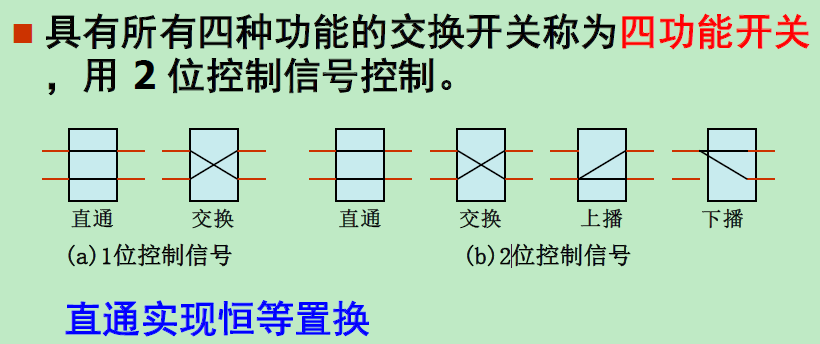
* 多级互连网络采用的关键技术：

 **(1)** 交换开关

 **(2)** 交换开关之间的拓扑连接

 **(3)** 对交换开关的控制方式

* 交换开关
  + 一个**a**×**b**交换开关有**a**个输入和**b**个输出
  + 最常用的二元开关：**a=b=2**
  + 每个输入可与一个或多个输出相连，但是在输出端必须避免发生冲突（多个输入同时连接到一个输出）。
  + 一对一和一对多映射是容许的；但不容许有多对一映射（冲突）。
  + **只容许一对一映射时称为置换连接，称这种开关为n×n交叉开关**。
  + 具有直通和交换两种功能的交换开关称为二功能开关 或 交换开关 。 用 1 位控制信号控制。
  + 具有所有四种功能的交换开关称为 四功能开关用 2 位控制信号控制



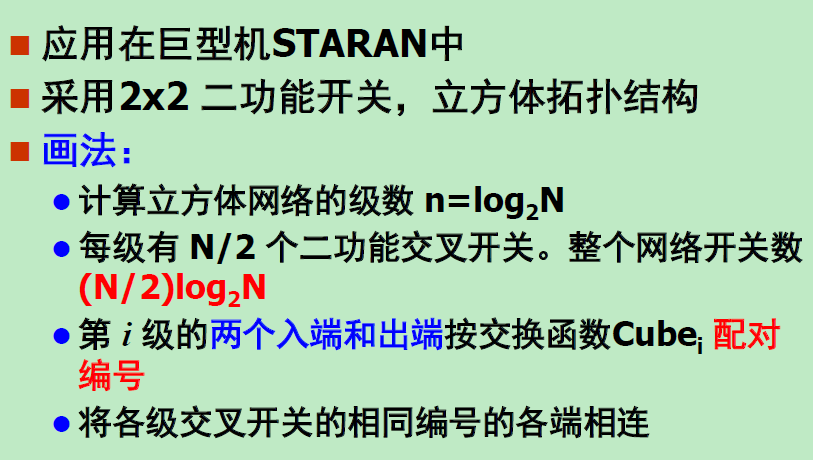
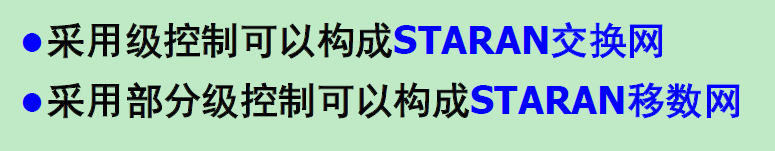
* 在多级互连网络中有多级交换开关，每一级又有多个交换开关。通常有三种控制方式：

**(1)** 级控制：同一级交换开关使用同一个控制信号控制。

**(2)** 单元级控制：每个交换开关分别控制。

**(3)** 部分级控制：例如，第**i**级使用**i+1**个控制信号控制**(0** **i** **n-1)**。

**同一个多级互连网络分别常用三种不同的控制方式，可以构成三种不同的互连网络**

1. 多级立方体网络
   1. 单级立方体网络的最大距离为 n 。即重复n 次使用单级互连网络，可以实现任意两个结点的连接。
   2. 因此，用n个单级立方体网络相串联，让每一级实现一种互连函数Cubei，其中i =0，1，… n-1，则同样能实现任意两个结点的连接。
   3. 
   4. 
   5. 

## 6.4.6 全排列网络

* 如果互连网络是从**N**个入端到**N**个出端的一到一的映射，就可以把它看成是对此**N**个端的重新排列。因此，互连网络的功能实际上就是用新排列来置换**N**个入端原有的排列。
* 循环互连网络和多级互连网络不能实现同时多对入端与出端之间的互连。因为会发生数据通路冲突。
* 多级立方体：**0**->**5**与**1**->**7**在开关**A**发生冲突

 **N=8**的多级混洗交换网络：**0**->**6**和**4**->**7**，**3**->**0**和**5**->**1**，**3**->**0**和**7**->**3**，**5**->**0**和**7**->**1**等有冲突

* 要同时实现两对或多对入端与出端之间的连接时，有可能因争用数据传送路径而发生冲突。我们称具有这类性质的互连网络为**阻塞式网络(BlockingNetwork)**。
* 反之，不具有这类性质的互连网络为**非阻塞式网络，或称为全排列网络**。非阻塞式网络连接的灵活性好，但连线多，控制复杂，成本高。

# 6.5并行存储器的无冲突访问

