第9章 互连网络

9.1 互连函数

9.2 互连网络的结构参数与性能指标

9.3 静态互连网络

9.4 动态互连网络

Switched networks are replacing buses as the normal means of communication between computers, between I/O devices, between boards, between chips, and even between modules inside chips.

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中结点之间的相互连接。

- 结点：处理器、存储模块或其它设备。
- 在拓扑上，互连网络为输入结点到输出结点之间的一组互连或映象。
- SIMD计算机和MIMD计算机的关键组成部分。
- 3大要素：互连结构，开关元件，控制方式。

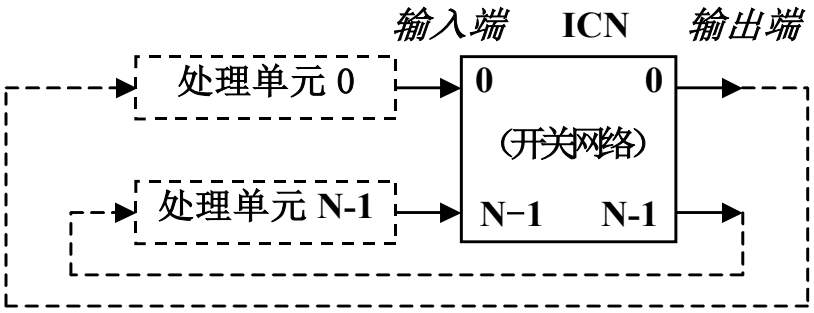
设计目标

The ultimate goal of computer architects is to design interconnection networks of the *lowest possible cost* that are capable of transferring the *maximum amount of available information* in the *shortest possible time.*

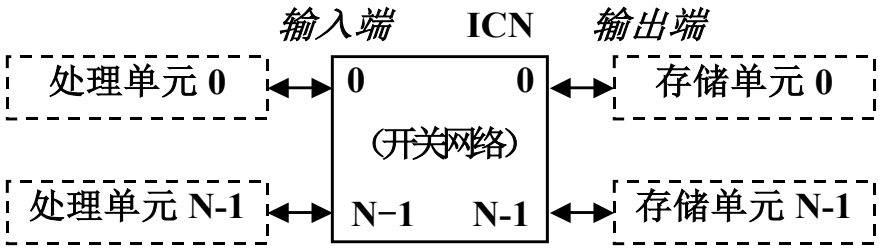
ICN目的与作用

(1) 当前提高计算速度的主要措施，一是改进器件，二是多处理单元并行计算。ICN是供多处理单元传输数据的高速通路，对并行计算时间影响很大。

(2) ICN与处理单元的连接模型



(a) 处理单元/处理单元的连接



(b) 处理单元/存储单元的连接

(3) ICN的主要操作：置换(N－N)，广播(1 － N)，选播(1 － N')。

互连网络的分类

(1)通用网/专用网

通用网（原用于计算机之间交换信息的普通网络）

专用网（专用于并行计算系统各处理单元之间并行交换数据的特殊网络）；

(2)串行网/并行网

串行网（多个结点的发送操作在时间上不能重叠）

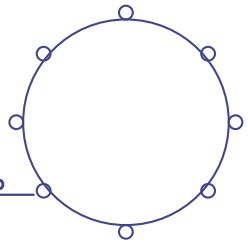
并行网（多个结点的发送操作在时间上可以重叠）；

(3)同步网/异步网（并行网再细分）

同步网（多个结点必须朝同一方向、以同一距离、同时开始发送）

异步网（多个结点可以朝不同方向、以不同距离、不同时开始发送，可能冲突）；

以右图为例，环形网可以设计成串行网、并行同步网、并行异步网。



(4)静态网/动态网（P273）

静态网（结点之间有固定连接）

动态网（结点之间的连接关系不固定，须通过开关导向或地址识别来确定当前的目的结点）；

互连网络的类型

On-chip networks (OCNs) — Also referred to as network-on-chip (NoC), Interconnect microarchitecture functional units, register files, caches, compute tiles, and processor and IP cores within chips or multichip modules.

System/storage area networks (SANs)— used for interprocessor and processor-memory interconnections within multiprocessor and multicomputer systems, and also for the connection of storage and I/O components within server and data center environments.

Local area networks (LANs)—used for interconnecting computer systems distributed across a machine room or throughout a building or campus environment. ***Ethernet*** has a 10 Gbps standard version that supports maximum performance over a distance of 40 km.

Wide area networks (WANs)—connect computer systems distributed across the globe, which requires internetworking support. WANs connect many millions of computers over distance scales of many thousands of kilometers.

9.2 互连网络的结构参数与性能指标

9.2.1 互连网络的结构参数

1. 网络通常是用有向边或无向边连接有限个结点的图来表示。
2. 互连网络的主要特性参数有：
   - 网络规模N：网络中结点的个数。

     表示该网络所能连接的部件的数量。
   - 结点度d：与结点相连接的边数（通道数），包括入度和出度。
     - 进入结点的边数叫入度。
     - 从结点出来的边数叫出度。

- 结点距离：对于网络中的任意两个结点，从一个结点出发到另一个结点终止所需要跨越的边数的最小值。
- 网络直径D：网络中任意两个结点之间距离的最大值。

  网络直径应当尽可能地小。
- 等分宽度b：把由N个结点构成的网络切成结点数相同（N/2）的两半，在各种切法中，沿切口边数的最小值。
  - **线等分宽度：**B＝b×w
    - **其中：w为通道宽度（用位表示）**
    - **该参数主要反映了网络最大流量。**
- 对称性：从任何结点看到的拓扑结构都相同的网络称为对称网络。

对称网络比较容易实现，编程也比较容易。

9.3 静态互连网络

互连网络通常可以分为两大类：

- 静态互连网络

  各结点之间有固定的连接通路、且在运行中不能改变的网络。

- 动态互连网络

  由交换开关构成、可按运行程序的要求动态地改变连接状态的网络。

下面介绍几种静态互连网络。

（其中：N表示结点个数）

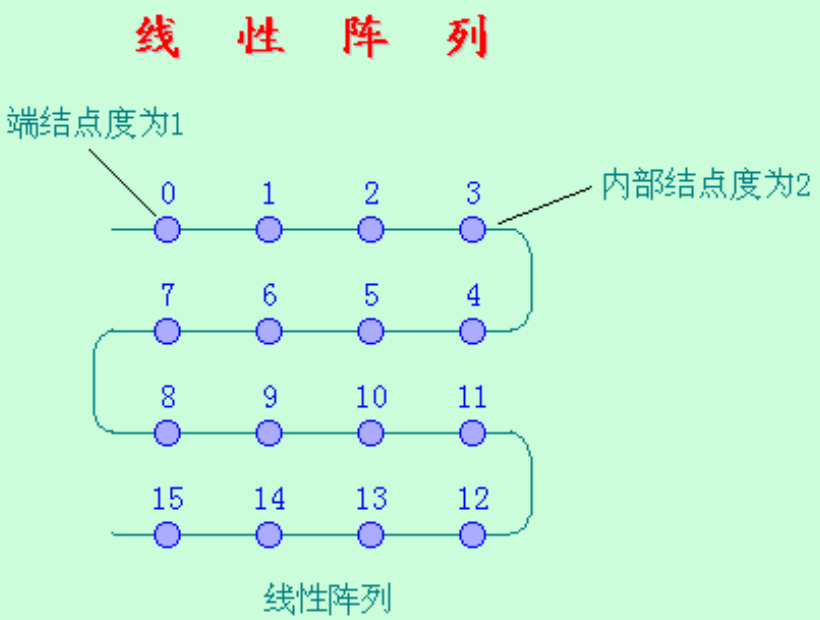
1. 线性阵列 一种一维的线性网络，其中N个结点用N-1个链路连成一行。

端结点的度：1

其余结点的度：2

直径：N－1

等分宽度b=1



线性阵列与总线的区别：

总线是通过切换与其连接的许多结点来实现时分特性的

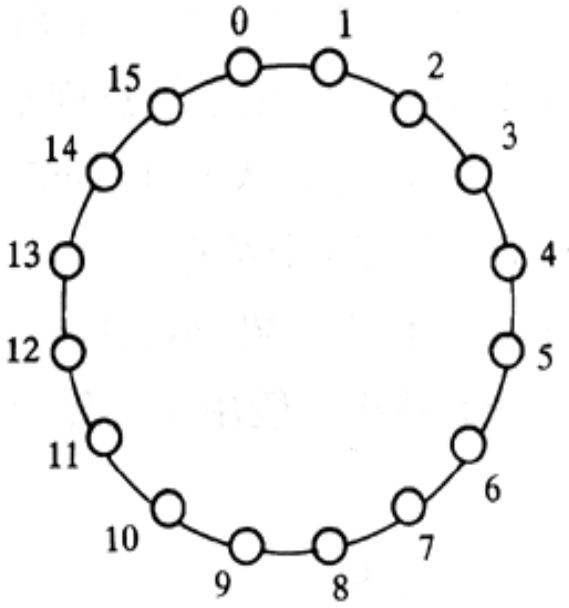
线性阵列允许不同的源结点和目的结点对并行地使用其不同的部分

2. 环和带弦环

➢ 环

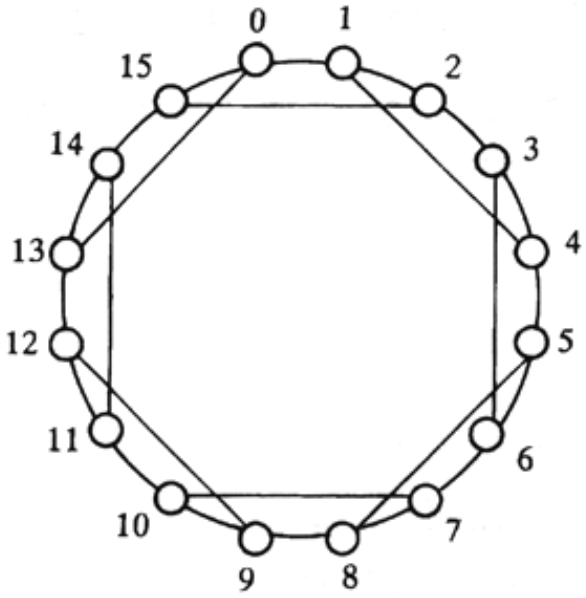
用一条附加链路将线性阵列的两个端点连接起来而构成。可以单向工作，也可以双向工作。

- 对称
- 结点的度：2
- 双向环的直径：N/2
- 单向环的直径：N
- 环的等分宽度b=2

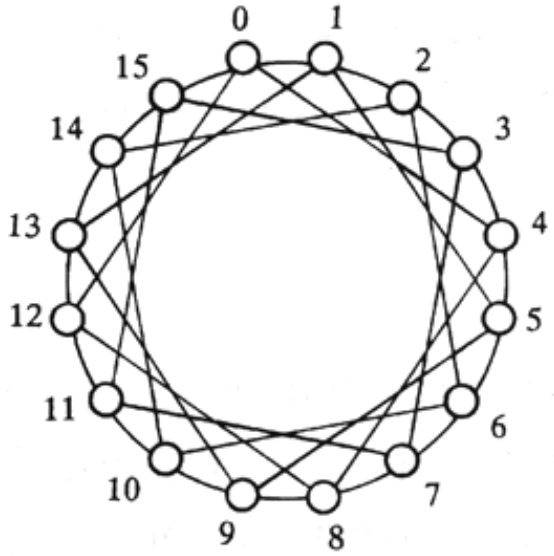


(b)环

带弦环　增加的链路愈多，结点度愈高，网络直径就愈小。

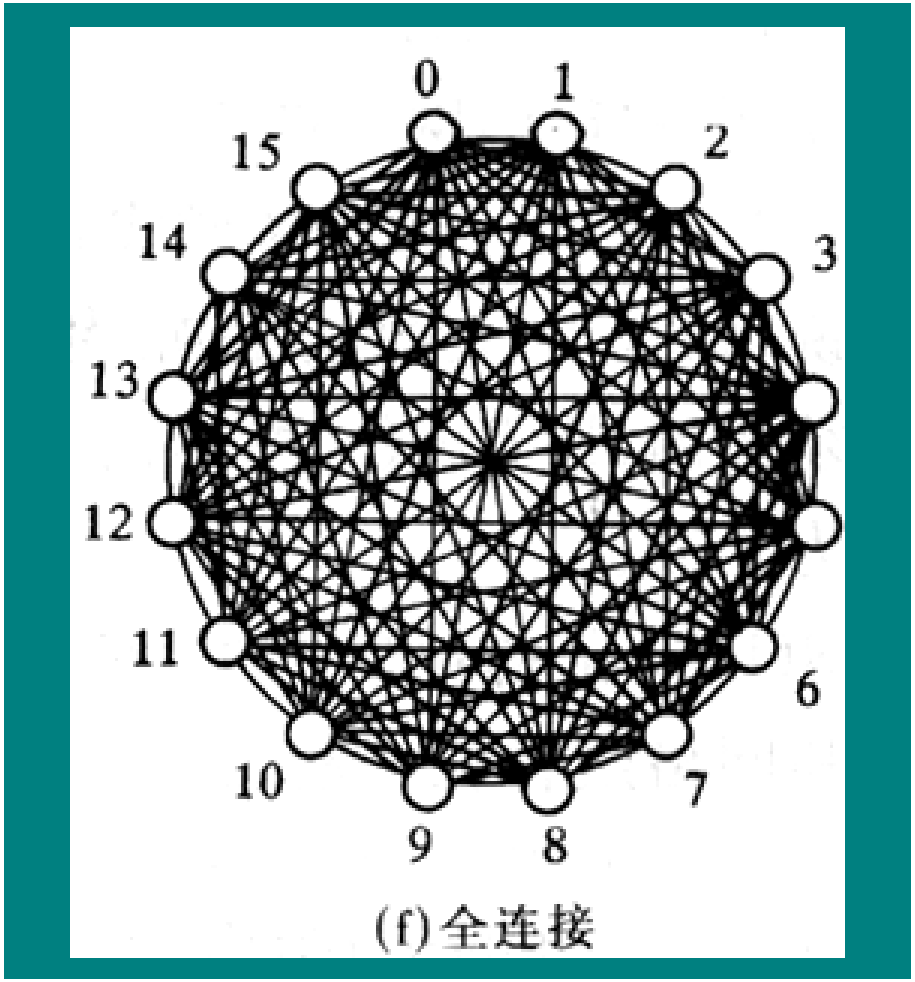


(c)度为 3 的带弦环



(d)度为 4 的带弦环(与 Illiac 网相同)

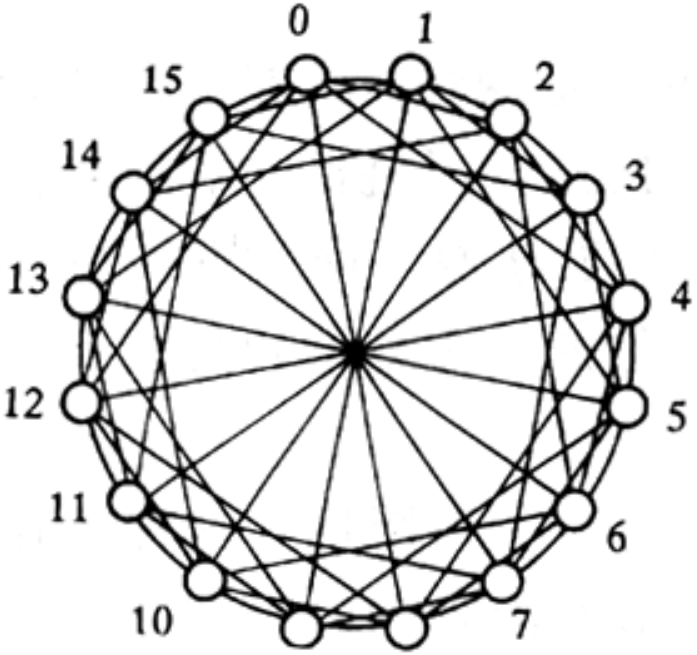
- 全连接网络
  - 结点度：15
  - 直径为1。



(f)全连接

3. 循环移数网络

- 通过在环上每个结点到所有与其距离为2的整数幂的结点之间都增加一条附加链而构成。



N=16　结点度：7;直径：2

(e)循环移数网络

1. 一般地，如果 | j−i | $=2^r$

（r=0, 1, 2, …, n−1，n=$\log_2 N$），

则结点i与结点j连接。

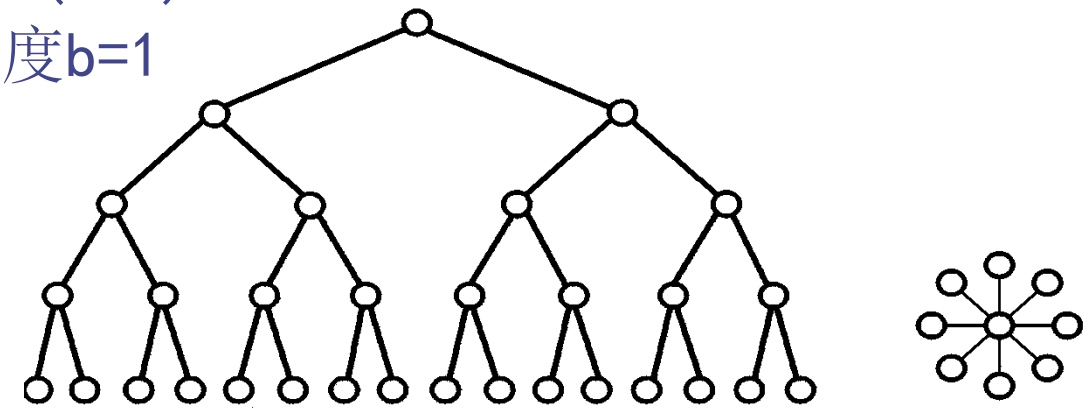
- 结点度：2n－1
- 直径：n/2
- 网络规模$N=2^n$

4. 树形和星形

树形:一棵5层31个结点的二叉树

一般，一棵k层完全平衡的二叉树有N=2k−1个结点。

- 最大结点度：3
- 直径：2(k−1)
- 等分宽度b=1

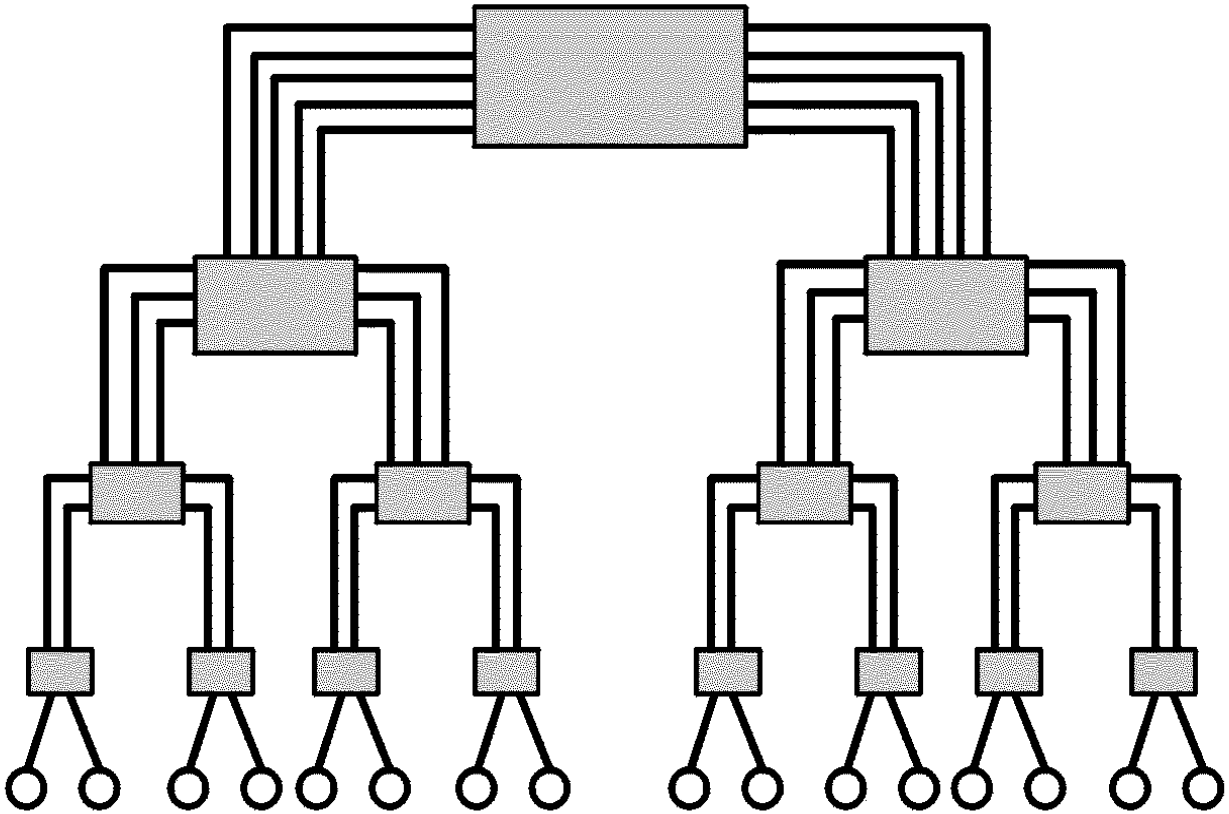


(a) 二叉树　　(b) 星形

星形:

- 结点度较高，为N－1。
  - 直径较小，是一常数2。等分宽度b=⌊N/2⌋
  - 可靠性较差，中心结点出故障，整个系统就会瘫痪。

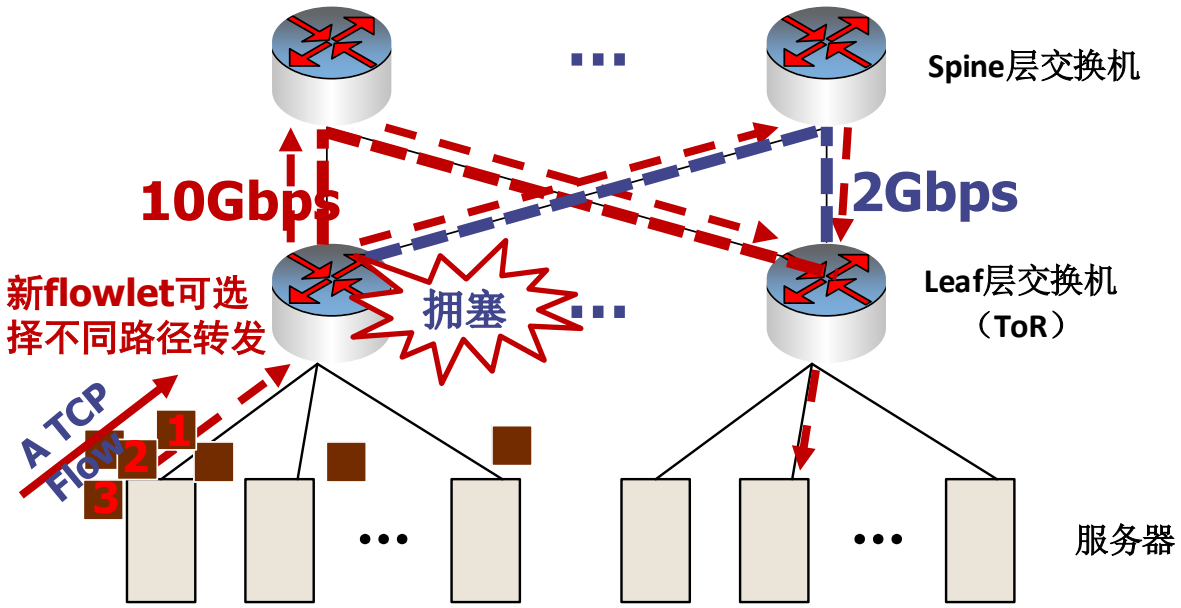
5. 胖树形



(c) 二叉胖树

数据中心内部复杂网络环境

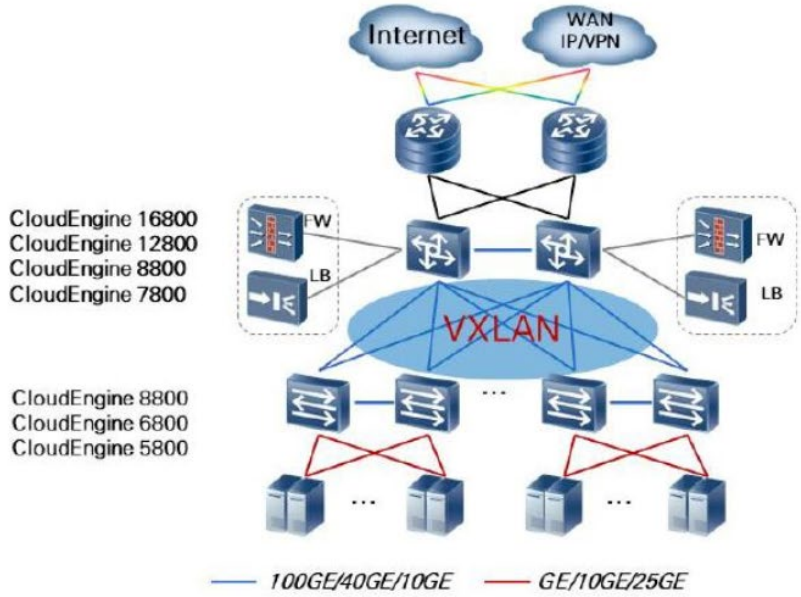
- 低时延链路（微秒级别）
- 多路径（交换机支持大量转发端口）
- 流量具备突发性
- 高带宽（10 – 几百Gbps）
- 网络不对称（如平行链路可用带宽不同）
- 不同网络需求的应用流量混合



Leaf-Spine
(数据中心典型网络拓扑之一)

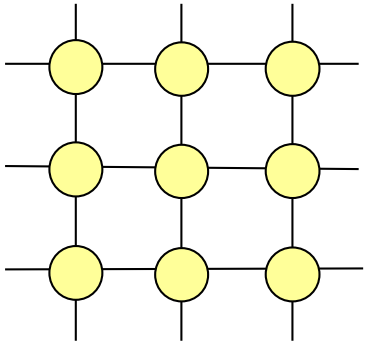
在数据中心典型应用

在数据中心的典型组网中， 采用 CloudEngine 16800 作为网络的核心交换机， CloudEngine 8800/CloudEngine 6800/CloudEngine 5800 作为 TOR 交换机，与 CloudEngine 16800 通过 100GE/40GE/10GE 端口互联。采用 VXLAN 等协议组建无阻塞大二层网络，保证虚拟机的大范围迁移以及用户业务的灵活部署。



华为公司数据中心交换机CloudEngine16800. https://e.huawei.com/cn/products/enterprise-networking/switches/data-center-switches/ce16800

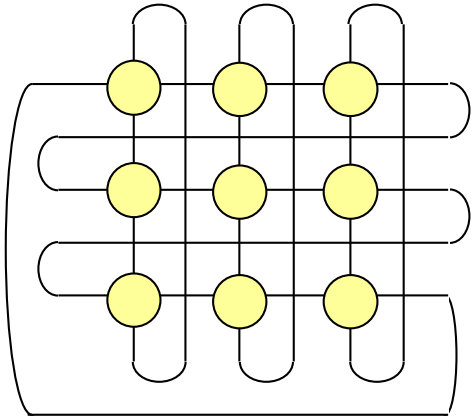
6. 网格形和环网形



（a）网格形

➢ 网格形

- 一个3×3的网格形网络
- 一个规模为$N=n\times n$的2维网格形网络
  - 内部结点的度d=4
  - 边结点的度d=3
  - 角结点的度d=2
  - 网络直径D=2（n-1）
  - 等分宽度b=n
- 一个由$N=n^k$个结点构成的k维网格形网络（每维n个结点）的内部结点度d=2k，网络直径D=k(n-1)。



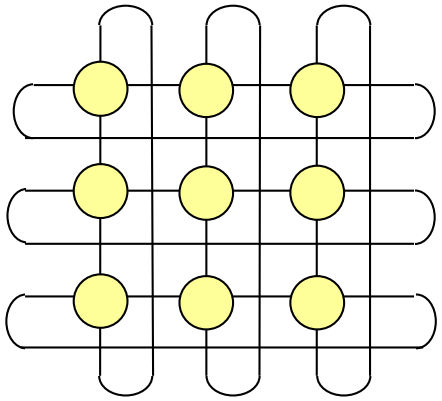
Illiac网络

- 名称来源于采用了这种网络的Illiac IV计算机
- 把2维网格形网络的每一列的两个端结点连接起来，再把每一行的尾结点与下一行的头结点连接起来，并把最后一行的尾结点与第一行的头结点连接起来。
- 一个规模为n×n的Illiac网络
  - 所有结点的度d=4
  - 网络直径D=n-1

    Illiac网络的直径只有纯网格形网络直径的一半。
  - 等分宽度：2n

➢ 环网形

- 可看作是直径更短的另一种网格。
- 把2维网格形网络的每一行的两个端结点连接起来，把每一列的两个端结点也连接起来。
- 将环形和网格形组合在一起，并能向高维扩展。
- 一个n×n的环网形网
  - 结点度：4
  - 网络直径：$2\times\lfloor n/2\rfloor$
  - 等分宽度**b=2n**



（c）环网形

9.1 互连函数

9.1.1 互连函数

变量x：输入（设x=0，1，...，N－1）

函数f(x)：输出

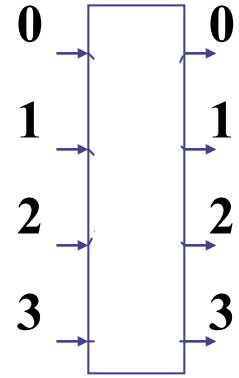
通过数学表达式建立输入端号与输出端号的连接关系。即在互连函数f的作用下，输入端x连接到输出端f(x)。

- 互连函数反映了网络输入数组和输出数组之间对应的置换关系或排列关系。

（有时也称为**置换函数**或**排列函数**）

互连函数有多种表示方式，如下例所示：

f(0)=1
f(1)=2
f(2)=0
f(3)=3



f= 0 1 2 3

f=(0,1,2)(3)

a.枚举法　　b.开关状态图　　c.列表法　　d.循环函数

一个网络通过开关切换可以形成多个映射关系，所以要用“互连函数族”来定义一个网络。

9.1.2 几种基本的互连函数

介绍几种常用的基本互连函数及其主要特征。

1. 恒等函数

- 恒等函数：实现同号输入端和输出端之间的连接。

$$I(x_{n-1}x_{n-2}\cdots x_1x_0)=x_{n-1}x_{n-2}\cdots x_1x_0$$

2. 交换函数

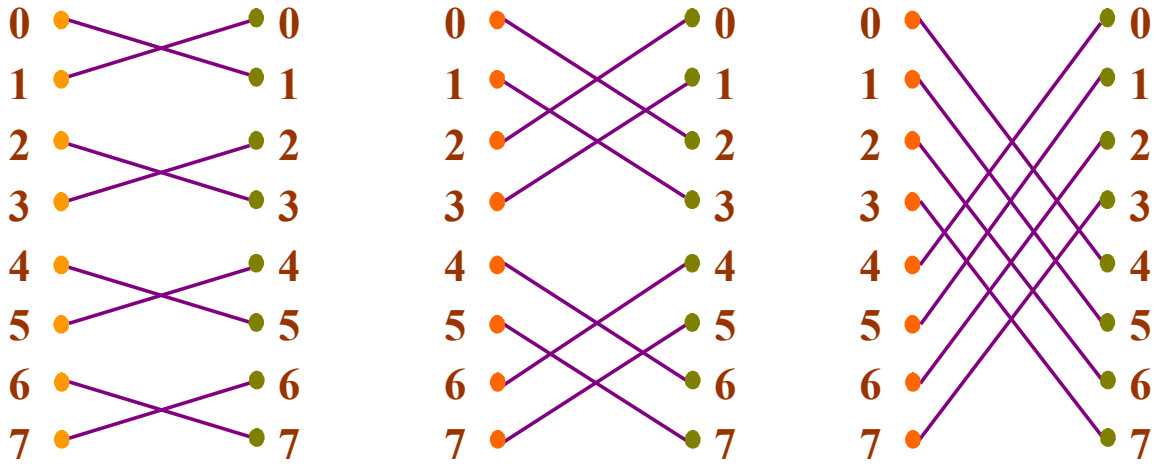
- 交换函数：实现二进制地址编码中第k位互反的输入端与输出端之间的连接。

$$E(x_{n-1}x_{n-2}\cdots x_{k+1}x_kx_{k-1}\cdots x_1x_0)=x_{n-1}x_{n-2}\cdots x_{k+1}\bar{x}_kx_{k-1}\cdots x_1x_0$$

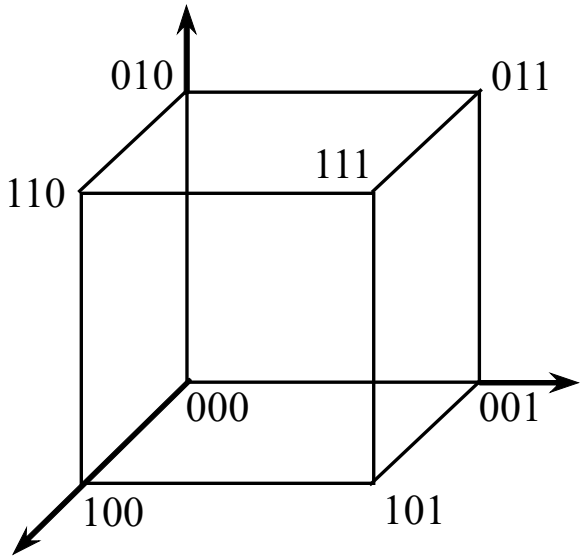
- 主要用于构造立方体互连网络和各种超立方体互连网络。
- 它共有$n=\log_2 N$种互连函数。（N为结点个数）
- 当N=8时，n=3，可得到常用的立方体互连函数：

$$Cube_0(x_2x_1x_0)=x_2x_1\overline{x}_0$$
$$Cube_1(x_2x_1x_0)=x_2\overline{x}_1x_0$$
$$Cube_2(x_2x_1x_0)=\overline{x}_2x_1x_0$$

N=8 的立方体交换函数



(a) $Cube_0$ 交换函数　(b) $Cube_1$ 交换函数　(c) $Cube_2$ 交换函数



立方体网络

3. 均匀洗牌函数

- 均匀洗牌函数：将输入端分成数目相等的两半，前一半和后一半按类似均匀混洗扑克牌的方式交叉地连接到输出端（输出端相当于混洗的结果）。
  - 也称为混洗函数（置换）
  - 函数关系

$$\sigma\left(x_{n-1}x_{n-2}\cdots x_1x_0\right)=x_{n-2}x_{n-3}\cdots x_1x_0x_{n-1}$$

即把输入端的二进制编号循环左移一位。

- 互连函数（设为s）的第k个子函数：把s作用于输入端的二进制编号的低k位。
- 互连函数（设为s）的第k个超函数：把s作用于输入端的二进制编号的高k位。

例如：对于均匀洗牌函数

第k个子函数：

$$\sigma_{(k)}(x_{n-1}\cdots x_k \mid x_{k-1}x_{k-2}\cdots x_0) = x_{n-1}\cdots x_k \mid x_{k-2}\cdots x_0x_{k-1}$$

即把输入端的二进制编号中的低k位循环左移一位。

第k个超函数：

$$\sigma^{(k)}(x_{n-1}x_{n-2}\cdots x_{n-k} \mid x_{n-k-1}\cdots x_1x_0) = x_{n-2}\cdots x_{n-k}x_{n-1} \mid x_{n-k-1}\cdots x_1x_0$$

即把输入端的二进制编号中的高k位循环左移一位。

下列等式成立：

$$\sigma^{(n)}(X)=\sigma_{(n)}(X)=\sigma(X)$$

$$\sigma^{(1)}(X)=\sigma_{(1)}(X)=X$$

- 对于任意一种函数f(x)，如果存在g(x)，使得

$$f(x)\times g(x)=I(x)$$

则称g(x)是f(x)的逆函数，记为$f^{-1}(x)$。

$$f^{-1}(x)=g(x)$$

- 逆均匀洗牌函数：将输入端的二进制编号循环右移一位而得到所连接的输出端编号。

- 互连函数

$$\sigma^{-1}\left(x_{n-1}x_{n-2}\cdots x_1x_0\right)=x_0x_{n-1}x_{n-2}\cdots x_1$$

  - 逆均匀洗牌是均匀洗牌的逆函数

➢ 当N=8时，有：

$$\sigma(x_2x_1x_0)=x_1x_0x_2$$

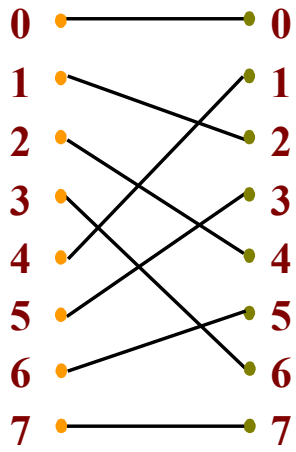
$$\sigma_{(2)}(x_2x_1x_0)=x_2x_0x_1$$

$$\sigma^{(2)}(x_2x_1x_0)=x_1x_2x_0$$

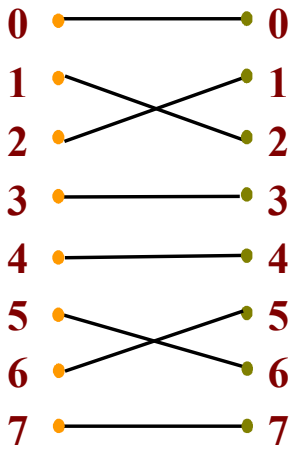
$$\sigma^{-1}(x_2x_1x_0)=x_0x_2x_1$$

性质1：shuffle（Xn-1Xn-2……X0）= Xn-2……X0Xn-1（循环左移）
性质2：shuffle^n（j）= j

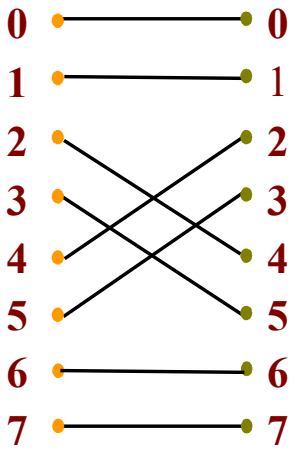
- N=8 的均匀洗牌和逆均匀洗牌函数



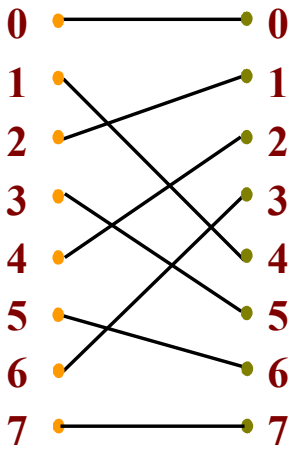
(a) 均匀洗牌函数 σ



(b) 子洗牌函数 $\sigma_{(2)}$



(c) 超洗牌函数 $\sigma^{(2)}$



(d) 逆均匀洗牌函数 σ^{-1}

N=8 的均匀洗牌函数

4. 碟式函数

- 蝶式互连函数：把输入端的二进制编号的最高位与最低位互换位置，便得到了输出端的编号。

$$\beta(x_{n-1}x_{n-2}\cdots x_1x_0)=x_0x_{n-2}\cdots x_1x_{n-1}$$

- 第k个子函数

$$\beta_{(k)}(x_{n-1}\cdots x_kx_{k-1}x_{k-2}\cdots x_1x_0)=x_{n-1}\cdots x_kx_0x_{k-2}\cdots x_1x_{k-1}$$

把输入端的二进制编号的低k位中的最高位与最低位互换。

- 第k个超函数

$$\beta^{(k)}(x_{n-1}x_{n-2}\cdots x_{n-k+1}x_{n-k}x_{n-k-1}\cdots x_1x_0)=x_{n-k}x_{n-2}\cdots x_{n-k+1}x_{n-1}x_{n-k-1}\cdots x_1x_0$$

把输入端的二进制编号的高k位中的最高位与最低位互换。

- 下列等式成立

$$\beta^{(n)}(X)=\beta_{(n)}(X)=\beta(X)$$

$$\beta^{(1)}(X)=\beta_{(1)}(X)=X$$

- 当N=8时，有：

$$\beta(x_2x_1x_0)=x_0x_1x_2$$

$$\beta_{(2)}(x_2x_1x_0)=x_2x_0x_1$$

$$\beta^{(2)}(x_2x_1x_0)=x_1x_2x_0$$

- 蝶式变换与交换变换的多级组合可作为构成方体多级网络的基础。

5. 反位序函数

- 反位序函数：将输入端二进制编号的位序颠倒过来求得相应输出端的编号。
  - **互连函数**

$$\rho\left(x_{n-1}x_{n-2}\cdots x_1x_0\right)=x_0x_1\cdots x_{n-2}x_{n-1}$$

- 第k个子函数

$$\rho_{(k)}(x_{n-1}\cdots x_kx_{k-1}x_{k-2}\cdots x_1x_0)=x_{n-1}\cdots x_kx_0x_1\cdots x_{k-2}x_{k-1}$$

即把输入端的二进制编号的低k位中各位的次序颠倒过来。

- 第k个超函数

$$\rho^{(k)}(x_{n-1}x_{n-2}\cdots x_{n-k+1}x_{n-k}x_{n-k-1}\cdots x_1x_0)=x_{n-k}x_{n-k+1}\cdots x_{n-2}x_{n-1}x_{n-k-1}\cdots x_1x_0$$

即把输入端的二进制编号的高k位中各位的次序颠倒过来。

- 下列等式成立

$$\rho^{(n)}(X)=\rho_{(n)}(X)=\rho(X)$$

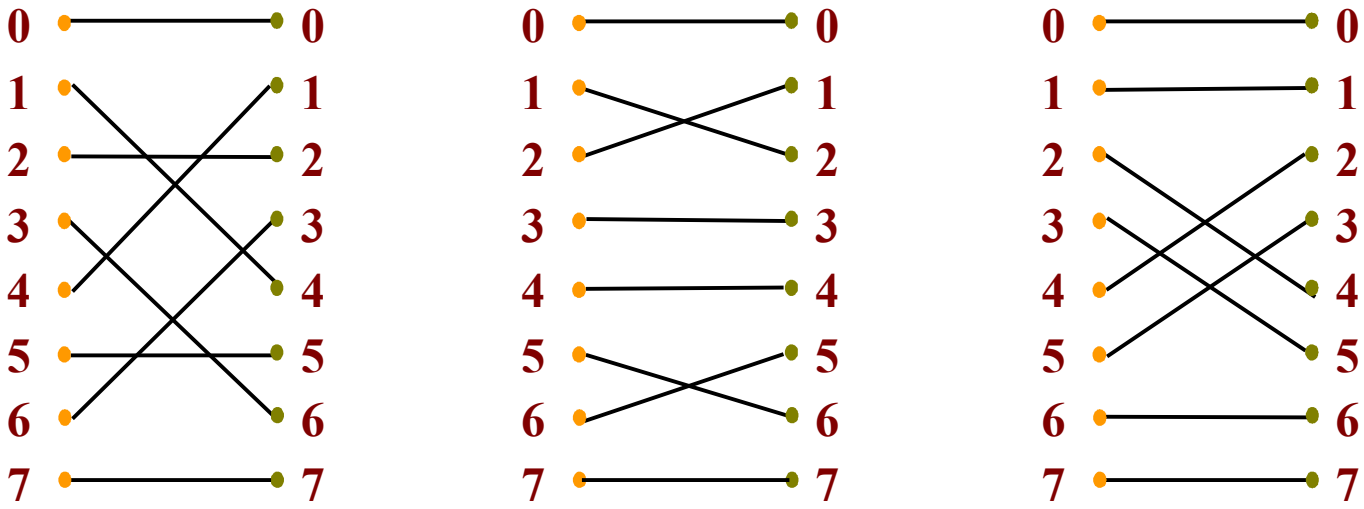
$$\rho^{(1)}(X)=\rho_{(1)}(X)=X$$

- 当N=8时，有：

$$\rho(x_2x_1x_0)=x_0x_1x_2$$

$$\rho_{(2)}(x_2x_1x_0)=x_2x_0x_1$$

$$\rho^{(2)}(x_2x_1x_0)=x_1x_2x_0$$



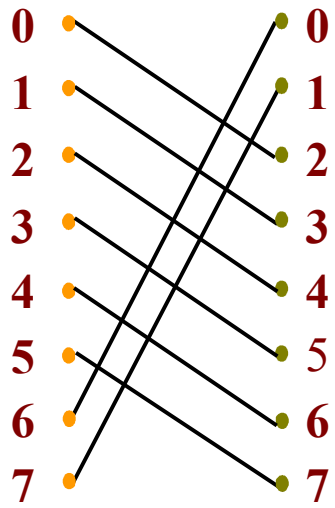
(a) $\beta = \rho$　　(b) $\beta_{(2)} = \rho_{(2)}$　　(c) $\beta^{(2)} = \rho^{(2)}$

N=8 的蝶式函数和反位序函数

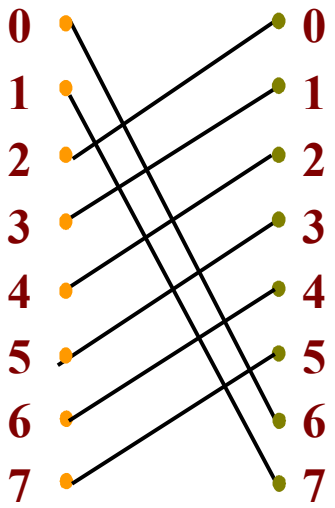
6. 移数函数

- 移数函数：将各输入端都错开一定的位置（模N）后连到输出端。
  - 函数式

$$\alpha(x)=（x\pm k）\bmod N \quad 1\le x\le N-1，1\le k\le N-1$$



(a) 左移移数函数 k＝2



(b) 右移移数函数 k＝2

7. PM2I函数

- P和M分别表示加和减，2I表示2^i。
  - 该函数又称为“加减2^i”函数。
- PM2I函数：一种移数函数，将各输入端都错开一定的位置（模N）后连到输出端。
- 互连函数

$$PM2_{+i}(x) = x + 2^i \bmod N$$

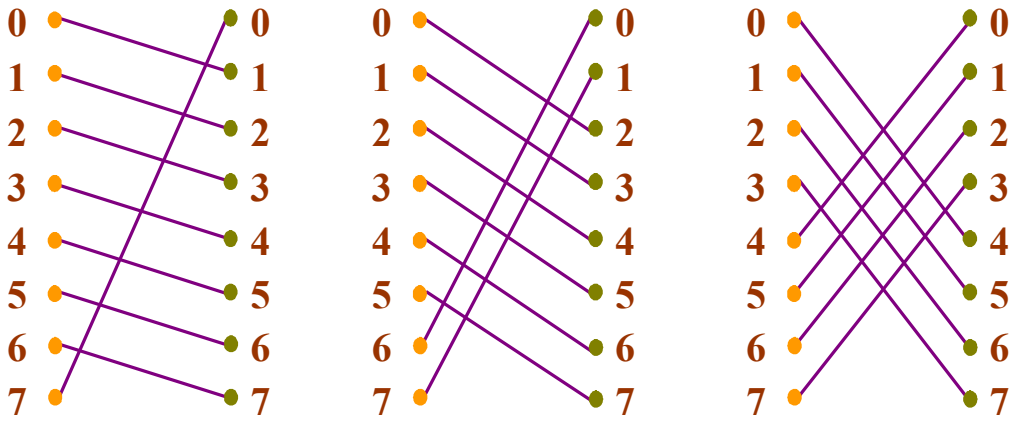
$$PM2_{-i}(x) = x - 2^i \bmod N$$

其中：

$0 \leqslant x \leqslant N-1$，$0 \leqslant i \leqslant n-1$，$n=\log_2 N$，N为结点数。

- PM2I互连网络共有2n个互连函数。

- 当N＝8时，有6个PM2I函数：
  - $PM2_{+0}$ ： (0 1 2 3 4 5 6 7)
  - $PM2_{-0}$ ： (7 6 5 4 3 2 1 0)
  - $PM2_{+1}$ ： (0 2 4 6) (1 3 5 7)
  - $PM2_{-1}$ ： (6 4 2 0) (7 5 3 1)
  - $PM2_{+2}$ ： (0 4) (1 5) (2 6) (3 7)
  - $PM2_{-2}$ ： (4 0) (5 1) (6 2) (7 3)



(a) $PM2_{+0}$ (b) $PM2_{+1}$ (c) $PM2_{+2}$

N=8 的PM2I函数

单级ICN

定义：单级ICN只使用一级开关，如下图所示。

开关的每种接通组合方式可用一个互连函数表示。

$f(j_{入}) = j_{出}$，$0 \leqslant j \leqslant N-1$

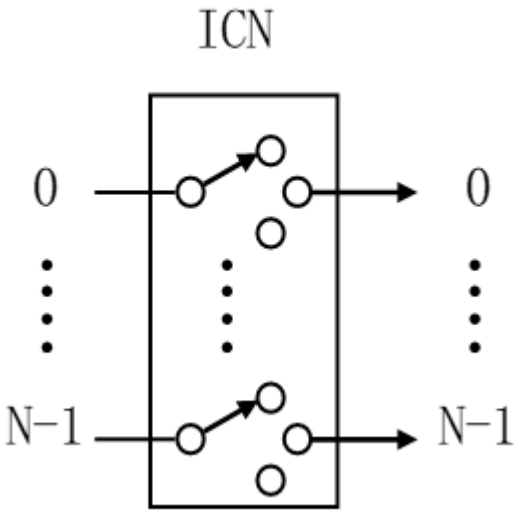
在互连函数中，记：

N —— 结点数

$n = \log_2 N$ —— 维数

$j = X_{n-1} \ldots\ldots X_0$ —— 结点

编号的二进制形式，位数为n。



互连函数族的组成必须使网络成为***连通图***。

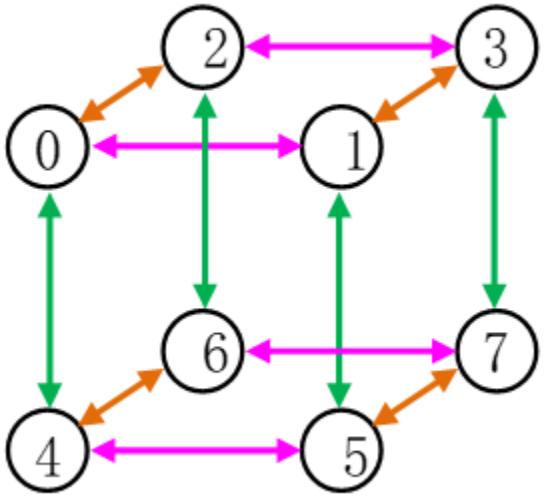
1、单级立方体网（Cube网）

该网络由立方体函数定义，立方体函数族有n个成员：$Cube_0$，$Cube_1$，…，$Cube_{n-1}$。

$Cube_i(X_{n-1}\ldots X_{i+1}X_iX_{i-1}\ldots X_0)=X_{n-1}\ldots X_{i+1}\overline{X_i}X_{i-1}\ldots X_0$，其中$0\leqslant i\leqslant n-1$

例如：$Cube_0(0)=1$，$Cube_3(7)=15$。

n=3的单级立方体网络拓扑形状如右图所示。



最坏情况下的传输需对输入结点编号的全部n位取反，所以单级立方体网络直径是n。成本$N\cdot\log_2N$ 。

立方体函数性质：结合律、交换律以及自反律（Cubei重复使用2次的结果与原始自变量相同）。

2、单级混洗-交换网

该网络由*混洗函数*（shuffle）与*交换函数*（exchange即$Cube_0$）定义，或者说它的互连函数族只有这两个成员。

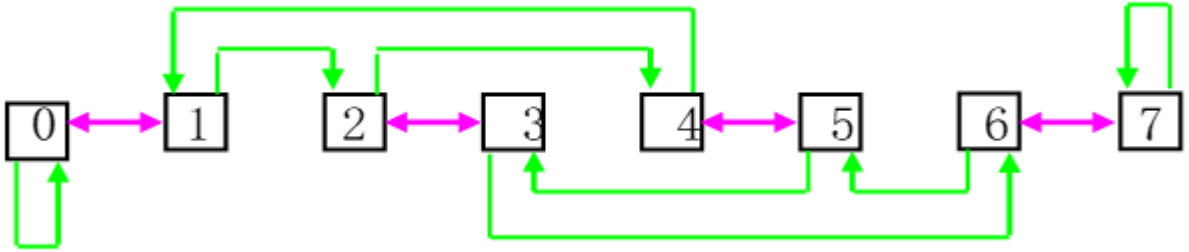
- ***混洗函数***定义：

$$shuffle(j) = \begin{cases} 2j \mod (N-1), & 当 j < N-1 \\ N-1, & 当 j = N-1 \end{cases}$$

例如：当N=8时，shuffle（0）= 0，shuffle（1）= 2，shuffle（7）= 7。

n=3的混洗网络拓扑形状如下图绿线所示，可以看出它不是一个*连通图*，所以还需要增加一个交换函数***交换函数***（图中红线所示），才能构成完整的单级混洗—交换网络。

单级混洗—交换网络的直径是2n-1。成本=N·2 。



3、单级加减2^i网（PM2I网，移数网）

该网络由PM2I函数定义，PM2I函数共有n对成员，分别是$PM2_{\pm 0}$，$PM2_{\pm 1}$，……，$PM2_{\pm (n-1)}$。

PM2I函数定义：$PM2_{\pm i}$的功能是对入端结点编号加或减2^i，然后再作模N运算

$$\begin{cases} PM2_{+i}(j) = j + 2^i \mod N \\ PM2_{-i}(j) = j - 2^i \mod N \end{cases}$$

其中j = 0 ～ N - 1，i = 0 ～ n - 1。

例如：当N = 8时，

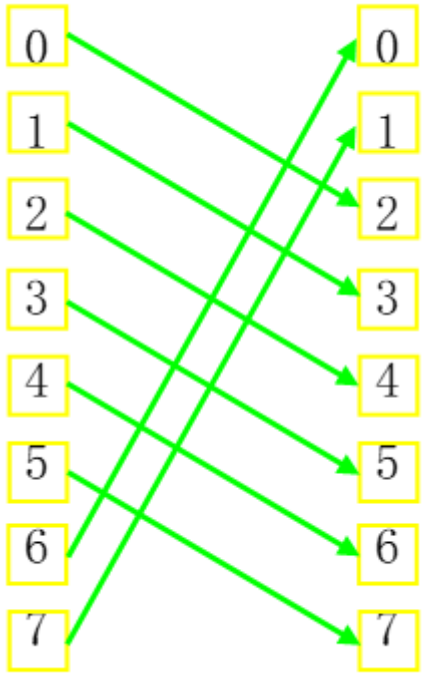
$PM2_{+0}(0) = 0 + 2^0 = 1$，

$PM2_{+0}(1) = 1 + 2^0 = 2$，

$PM2_{+0}(7) = 7 + 2^0 = 0$，

$PM2_{+1}(0) = 0 + 2^1 = 2$。

N = 8的$PM2_{+1}(j)$函数开关状态如右图所示，其连接规律是把各入端结点编号加上相同的增量2^1（mod N），获得出端结点编号。



N = 8的$PM2_{+i}$网络拓扑形状如图所示，

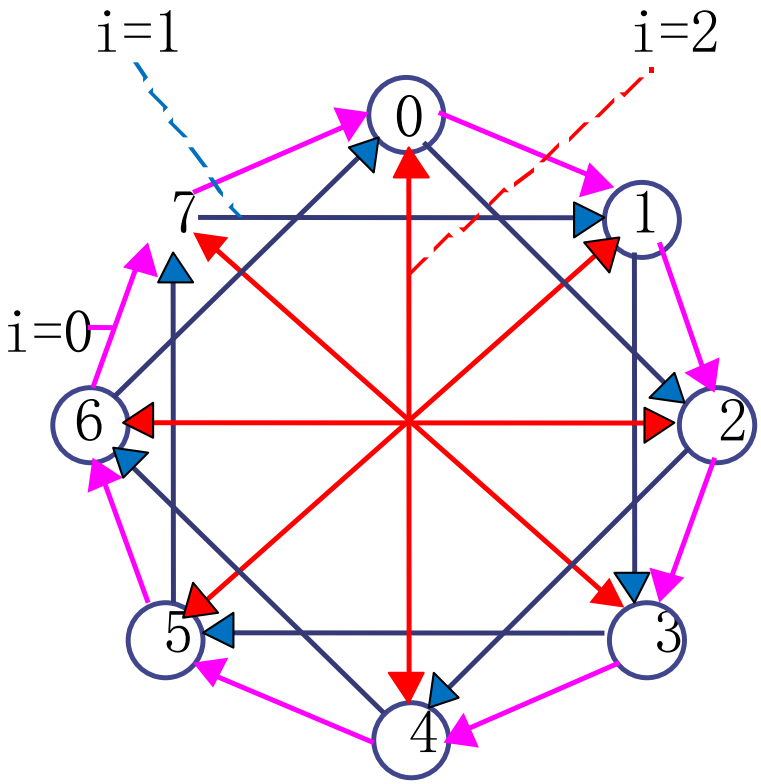
- 性质1：对相同的i值，$PM2_{+i}$与$PM2_{-i}$函数的传送路径相同，方向相反（右图中所有箭头反向即为$PM2_{-1}$的拓扑形状）；

- 性质2：$PM2_{+(n-1)} = PM2_{-(n-1)}$。

根据性质2，我们知道单级PM2I网络实际上只能实现2n-1种不同的置换。

单级PM2I网络的直径是$\lceil n/2 \rceil$。

成本=$N(2\cdot\log_2 N - 1)$ 。



可以看出它包含多个强连通子图（即除去若干边以后仍能保证任何一对结点互相可达），所以这2n个函数并不是实现互连网的最小集合。实际应用中为了降低造价，人们往往取它们的一个子集来构造互连网。

各种互连函数总结

交换置换函数定义　$E(X_{n-1}X_{n-2}\ldots X_1X_0)=X_{n-1}X_{n-2}\ldots X_1\overline{X_0}$，
其中$0\leqslant i\leqslant n-1$

立方体函数定义：Cubei的功能是对入端结点编号二进制形式的第i位取反

$Cube_i(X_{n-1}\ldots X_{i+1}X_iX_{i-1}\ldots X_0)=X_{n-1}\ldots X_{i+1}\overline{X_i}X_{i-1}\ldots X_0$，其中$0\leqslant i\leqslant n-1$

均匀洗牌

$$shuffle(X_{n-1}X_{n-2}\ldots\ldots X_0)=X_{n-2}\ldots\ldots X_0X_{n-1}$$（循环左移）

PM2I函数定义：$PM2_{\pm i}$的功能是对入端结点编号加或减2^i，然后再作模N运算

$$PM2_{+i}(X)=X+2^i \mod N$$
$$PM2_{-i}(X)=X-2^i \mod N$$

其中 X = 0 ～ N - 1，i = 0 ～ n - 1。

例9.1 现有16个处理器，编号分别为0，1，…，15，用一个N=16的互连网络互连。处理器i的输出通道连接互连网络的输入端i，处理器i的输入通道连接互连网络的输出端i。当该互连网络实现的互连函数分别为：

（1）$Cube_3$

（2）$PM2_{+3}$

（3）$PM2_{-0}$

（4）σ

（5）σ（σ）

时，分别给出与第13号处理器所连接的处理器号。

解：（1）由 $Cube_3(x_3x_2x_1x_0) = \bar{x}_3x_2x_1x_0$

得$Cube_3$（1101）= 0101，即处理器13连接到处理器5。

令$Cube_3$（ $x_3x_2x_1x_0$ ）= 1101，得$x_3x_2x_1x_0$= 0101，故与处理器13相连的是处理器5。

所以处理器13与处理器5双向互连。

（2）由$PM2_{+3} = j+2^3 \bmod 16$，得$PM2_{+3}$（13）$= 13+2^3 = 5$， 即处理器13连接到处理器5。

令$PM2_{+3}$（j）$= j+2^3 \bmod 16 = 13$，得j= 5，故与处理器13相连的是处理器5。

所以处理器13与处理器5双向互连。

（3）由$PM2_{-0}$（j）$= j-2^0 \bmod 16$，得$PM2_{-0}$（13）$= 13-2^0 = 12$，即处理器13连接到处理器12。

令$PM2_{-0}$（j）$= j-2^0 \bmod 16 = 13$，得j= 14，故与处理器13相连的是处理器14。

所以处理器13连至处理器12，而处理器14连至处理器13。

（4）由$\sigma(x_3x_2x_1x_0)= x_2x_1x_0x_3$，得$\sigma(1101)= 1011$，即处理器13连接到处理器11。

令$\sigma(x_3x_2x_1x_0)= 1101$，得$x_3x_2x_1x_0= 1110$，故与处理器13相连的是处理器14。

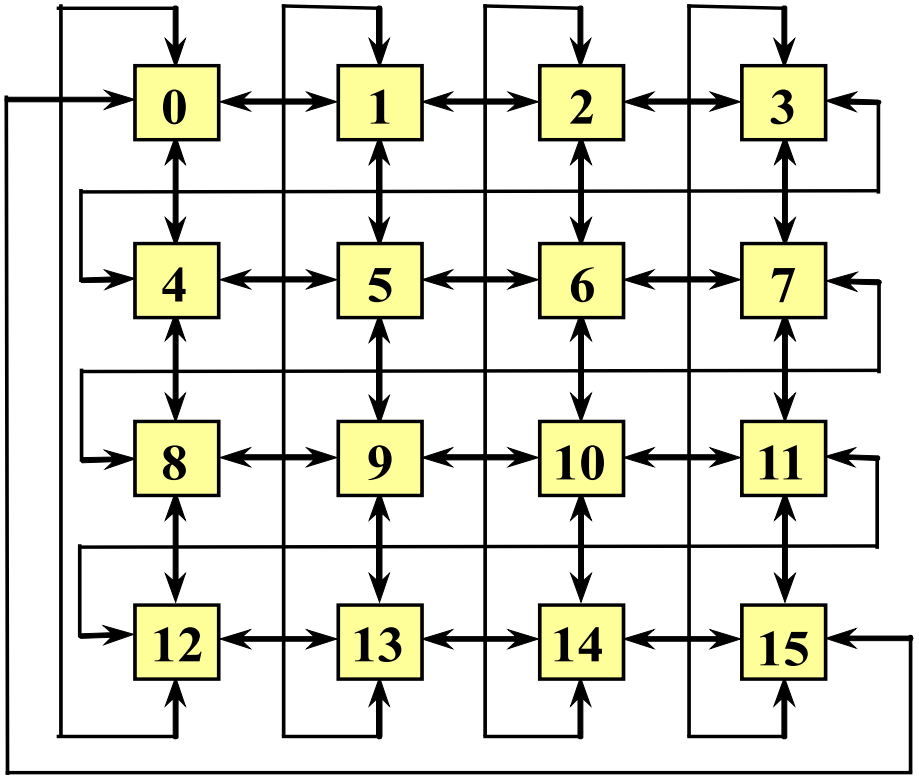
所以处理器13连至处理器11，而处理器14连至处理器13。

（5）由$\sigma(\sigma(x_3x_2x_1x_0))= x_1x_0x_3x_2$，得$\sigma(\sigma(1101))= 0111$，即处理器13连接到处理器7。

令$\sigma(\sigma(x_3x_2x_1x_0))= 1101$，得$x_3x_2x_1x_0= 0111$，故与处理器13相连的是处理器7。

所以处理器13与处理器7双向互连。

➢ 阵列计算机ILLIAC Ⅳ



用移数函数构成ILLIAC Ⅳ 阵列机的互连网络

例9.2 已知有16台个处理器用Illiac网络互连，写出Illiac网络的互连函数，给出表示任何一个处理器PU_i（0≤i≤15）与其他处理器直接互连的一般表达式。

解：Illiac网络连接的结点数N=16，组成4×4的阵列。每一列的4个处理器互连为一个双向环，第1列～第4列的双向环可分别用循环互连函数表示为：

（0 4 8 12）　　（12 8 4 0）

（1 5 9 13）　　（13 9 5 1）

（2 6 10 14）　　（14 10 6 2）

（3 7 11 15）　　（15 11 7 3）

其中，传送方向为顺时针的4个单向环的循环互连函数可表示为：

$$PM2_{+2}（X）=（X+2^2）\bmod N=（X+4）\bmod 16$$

传送方向为逆时针的4个单向环的循环互连函数可表示为：

$$PM2_{-2}(X) = (X - 2^2) \bmod N = (X - 4) \bmod 16$$

16个处理器由Illiac网络的水平螺线互连为一个双向环，用循环互连函数表示为：

（0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15）

（15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0）

其中，传送方向为顺时针的单向环的循环互连函数可表示为：

$$PM2_{+0}(X) = (X + 2^0) \bmod N = (X + 1) \bmod 16$$

传送方向为逆时针的单向环的循环互连函数可表示为：

$$PM2_{-0}(X) = (X - 2^0) \bmod N = (X - 1) \bmod 16$$

所以，N=16的Illiac网络的互连函数有4个：

$PM2_{\pm 0}(X)$ 和 $PM2_{\pm 2}(X)$

由互连函数可得任何一个处理器i直接与下述4个处理器双向互连：

$i \pm 1 \bmod 16$

$i \pm 4 \bmod 16$

符号约定：起点$x = x_{n-1}\cdots\cdots x_0$，终点$y = y_{n-1}\cdots\cdots y_0$。

1. 单级立方体网

二者间路径由地址逻辑差$y_i \oplus x_i$决定。$y_i \oplus x_i$="1"代表$Cube_i$维需要走一步，各维先后顺序可任意安排，"1"的个数即是总步数。

2. 单级混洗-交换网

二者间路径也由地址逻辑差$y_i \oplus x_i$决定。如果$y_i \oplus x_i$="1"，表明x_i须先经n-i步shuffle到最低位，1步$Cube_0$求反，再经i步shuffle回到原位变成y_i。如有多位"1"，可以合并shuffle，具体算法须灵活设计。

1010 → 0110 ？

课堂练习

1、
写出用单级混洗交换网模拟单级立方体网的算法，并求模拟步数的最大、最小值。

2、
求N个结点网络的最少广播步数。
(1)单级立方体网，每步只能执行一个Cube函数；
(2)单级混洗交换网，每步只能执行Shuffle函数、Exchange函数中的一个。

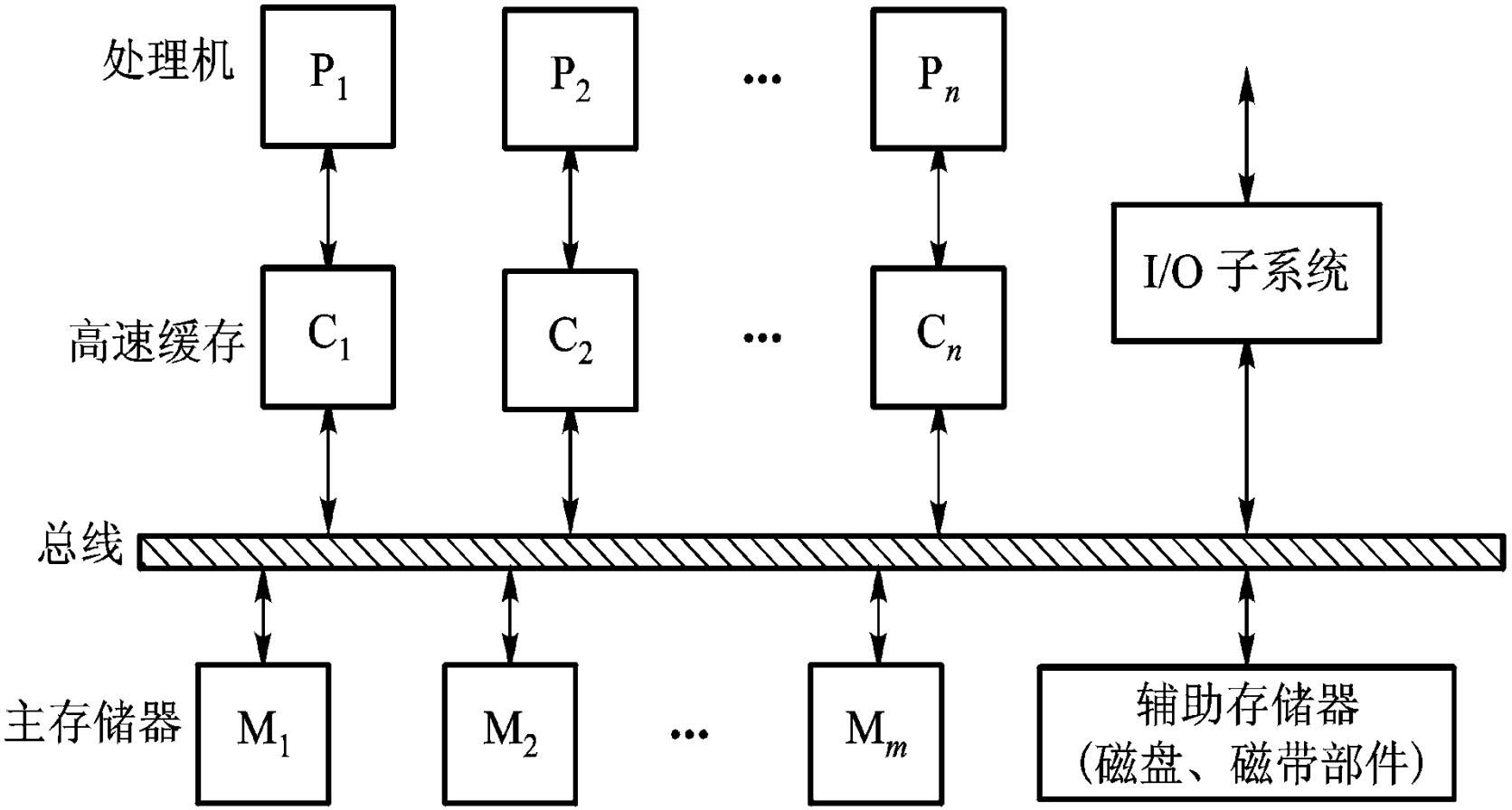
作业: 9.8, 9.9（改一字：“每步只能使用Cube0或σ一次”）

9.4 动态互连网络

9.4.1 总线网络

1. 由一组导线和插座构成，经常被用来实现计算机系统中处理机模块、存储模块和外围设备等之间的互连。
   - 每一次总线只能用于一个源（主部件）到一个或多个目的（从部件）之间的数据传送。
   - 多个功能模块之间的争用总线或时分总线
   - 特点
     - **结构简单、实现成本低、带宽较窄**

2. 一种由总线连接的多处理机系统



- 系统总线在处理机、I/O子系统、主存储器以及辅助存储设备（磁盘、磁带机等）之间提供了一条公用通路。
- 系统总线通常设置在印刷电路板底板上。处理器板、存储器板和设备接口板都通过插座或电缆插入底板。

3. 解决总线带宽较窄问题：采用多总线或多层次的总线

➢ 多总线是设置多条总线

有两种做法：

- 为不同的功能设置专门的总线
- 重复设置相同功能的总线

➢ 多层次的总线是按层次的架构设置速度不同的总线，使得不同速度的模块有比较适合的总线连接。

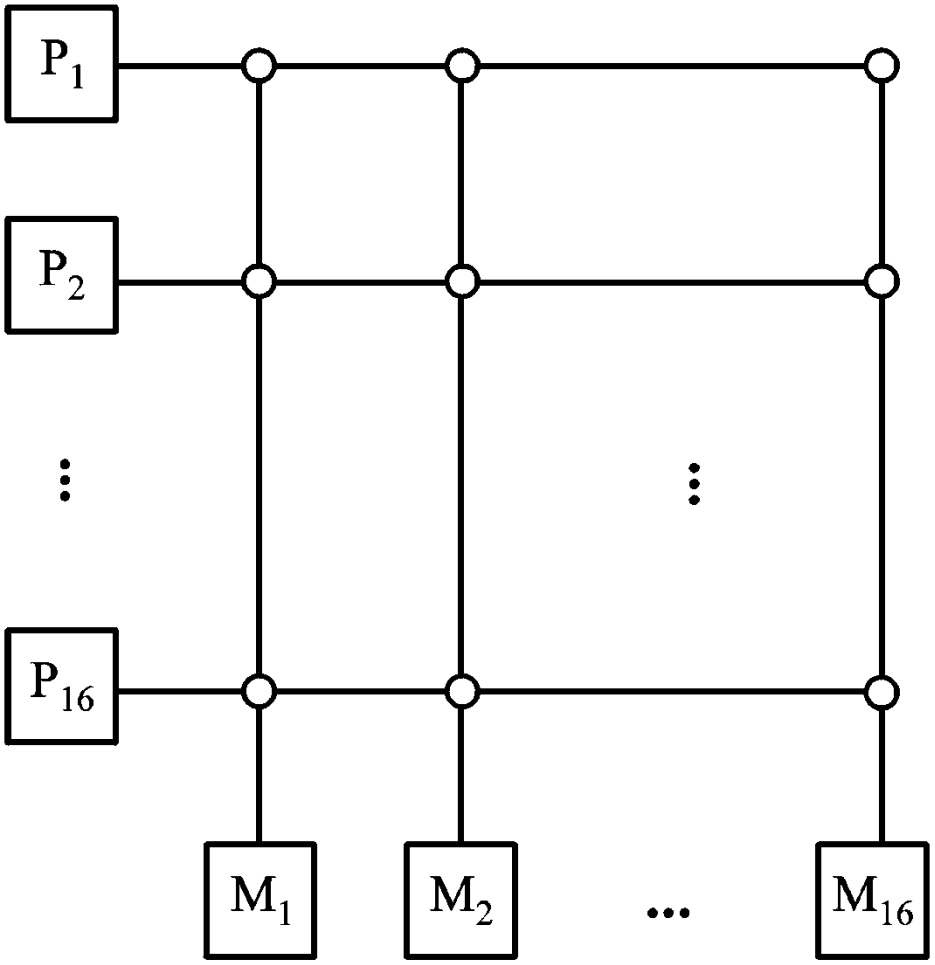
9.4.2 交叉开关网络

1. 单级开关网络

- 交叉点开关能在对偶（源、目的）之间形成动态连接，同时实现多个对偶之间的无阻塞连接。
- 带宽和互连特性最好。
- 一个n×n的交叉开关网络，可以无阻塞地实现n!种置换。
- 对一个n×n的交叉开关网络来说，需要n^2套交叉点开关以及大量的连线。
  - **当n很大时，交叉开关网络所需要的硬件数量非常巨大。**

2. C.mmp多处理机的互连结构

- 用16×16的交叉开关网络把16台PDP-11处理机与16个存储模块连在一起
- 最多可同时实现16台处理机对16个不同存储模块的并行访问
  - 每个存储模块一次只能满足一台处理机的请求
  - 当多个请求要同时访问同一存储模块时，交叉开关就必须分解所发生的冲突，每一列只能接通一个交叉点开关。
  - 为了支持并行（或交叉）存储器访问，可以在同一行中接通几个交叉点开关。

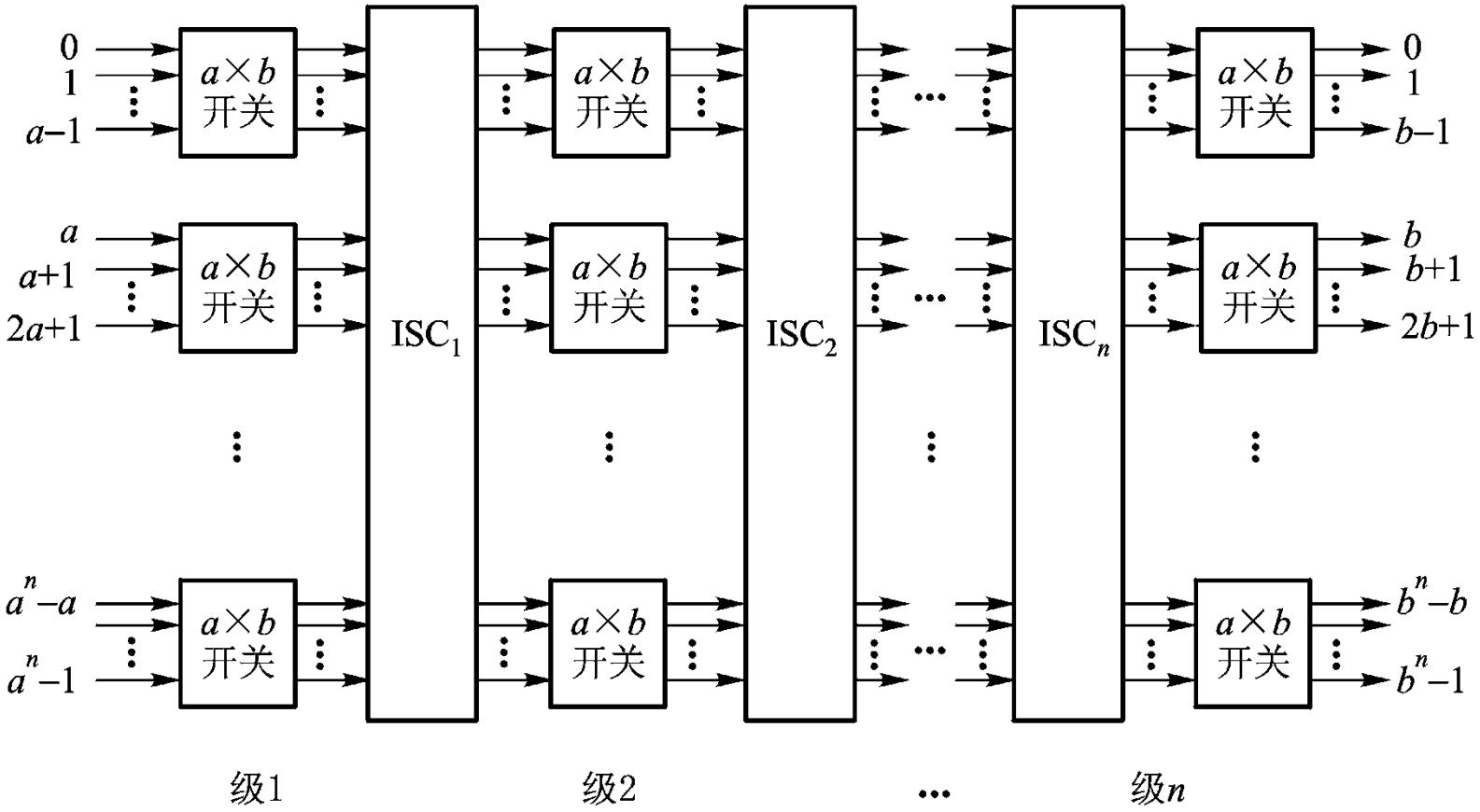

P1
P2
P16
M1
M2
...
M16

★9.4.3 多级互连网络

1. A common way of addressing the crossbar scaling problem consists of splitting the large crossbar switch into several stages of smaller switches interconnected

多级互连网络的构成

- MIMD和SIMD计算机都采用多级互连网络MIN（Multistage Interconnection Network）
- 一种通用的多级互连网络
  - 由a×b开关模块和级间连接构成的通用多级互连网络结构
  - 每一级都用了多个a×b开关
    - a个输入和b个输出
    - 在理论上，a和b不一定相等，然而实际上a和b经常选为2的整数幂，即$a=b=2^k$，$k\geqslant 1$。
  - 相邻各级开关之间都有固定的级间连接


0
1
a−1
a
a+1
2a+1
a^n−a
a^n−1
a×b 开关
ISC1
ISC2
ISCn
0
1
b−1
b
b+1
2b+1
b^n−b
b^n−1
级1
级2
…
级n

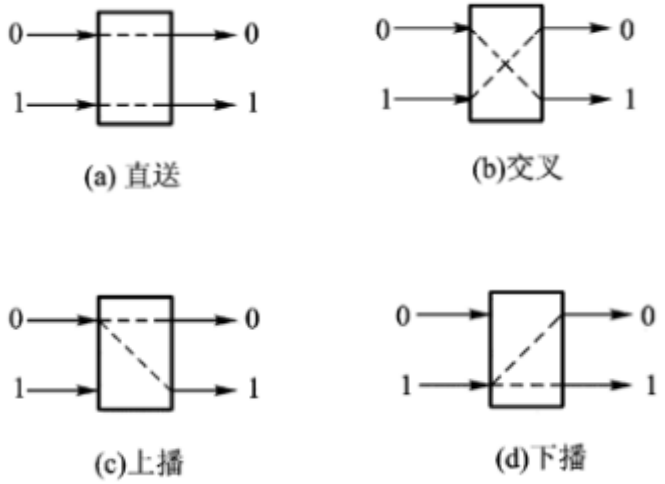
多级ICN（P281）

定义：多级ICN使用多级开关，使得数据在一次通过网络的过程中可以实现的置换种类更多。

通常在N个结点的网络中，多级ICN由n级构成（$n = \log_2 N$）。

经典的多级互连网有多级立方体网、多级混洗—交换网和多级PM2I网。我们只学习多级混洗—交换网。

多级立方体网和多级混洗—交换网不使用单级互连网中的那种多路选择开关，而是用一种2输入/2输出的二元交换开关，以减少开关总数。二元交换开关基本接通状态有“直连”、“交换”、“上播”和“下播”，进行数据置换时只能使用前2种。



模块大小	合法状态	置换连接
2×2	4	2
4×4	256	24
8×8	16 777 216	40 320
n×n	n^n	n!

- 各种多级互连网络的区别在于所用开关模块、控制方式和级间互连模式的不同。
  - 控制方式：对各个开关模块的控制方式。
    - 级控制：每一级所有开关只用一个控制信号控制，只能同时处于同一种状态。
    - 部分级控制：第i级的所有开关分别用i＋1个信号控制，0≤i≤n－1，n为级数。
    - 单元控制：每一个开关都有一个独立的制信号，可各自处于不同的状态。各开关动作独立，性能比前两种方式都更灵活，结构也更复杂。
  - 常用的级间互连模式：

均匀洗牌、蝶式、多路洗牌、纵横交叉、立方体连接等

多级混洗—交换网络（Omega网，P410/P436）

由n级构成，每一级包含一个无条件混洗拓扑线路和一列可控的二元交换开关，前后重复，便于制造。如P436图7.43所示。

各级编号是n-1，……，0，即按降序排列。

在多级混洗—交换网络中：

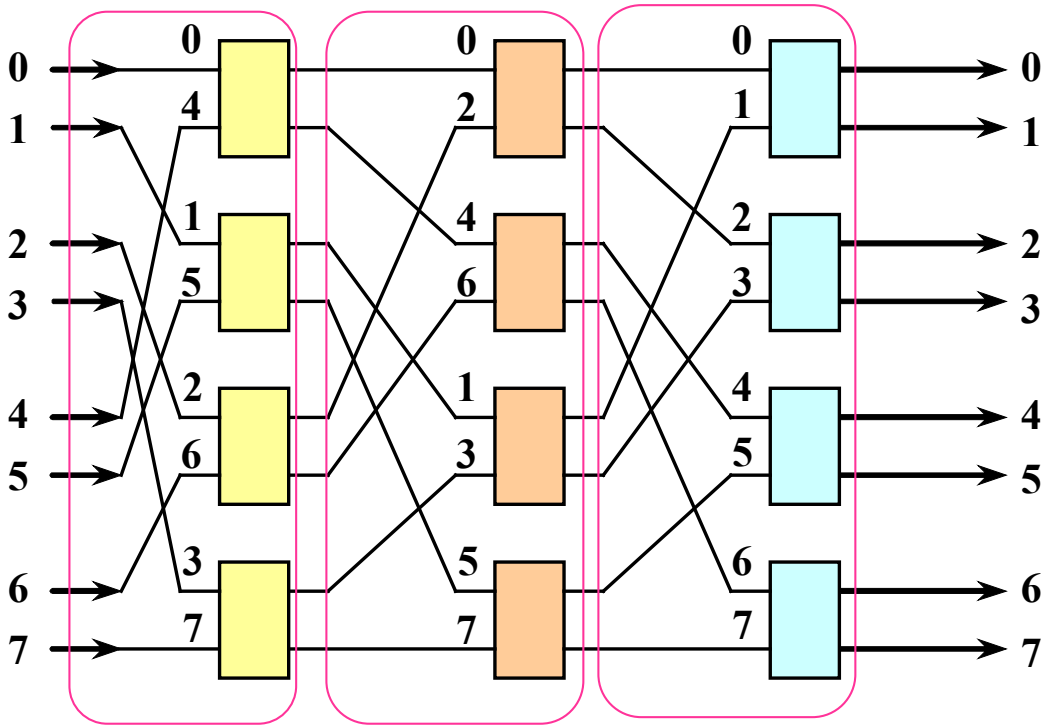
- 单独一级混洗拓扑线路可完成一次数据混洗（shuffle），
- 单独一列二元交换开关在处于“交换”状态时可完成一次交换操作（$Cube_0$）

- 如果各级二元交换开关都处于“直连”状态，N个结点的数据通过网络仅经过n次混洗操作，排列顺序最终恢复输入状态（混洗函数性质2）；
- 如果各级二元交换开关都处于“交换”状态，则N个结点的数据在每次混洗之后紧接着一次交换（$Cube_0$），也就是地址码的最低位取反，最后n位地址均被取反。

程序员根据数据置换或复制的需要，可以灵活地设置各开关的状态。

3. Omega网络

- 一个8×8的Omega网络
  - 每级由4个4功能的2×2开关构成
  - 级间互连采用均匀洗牌连接方式

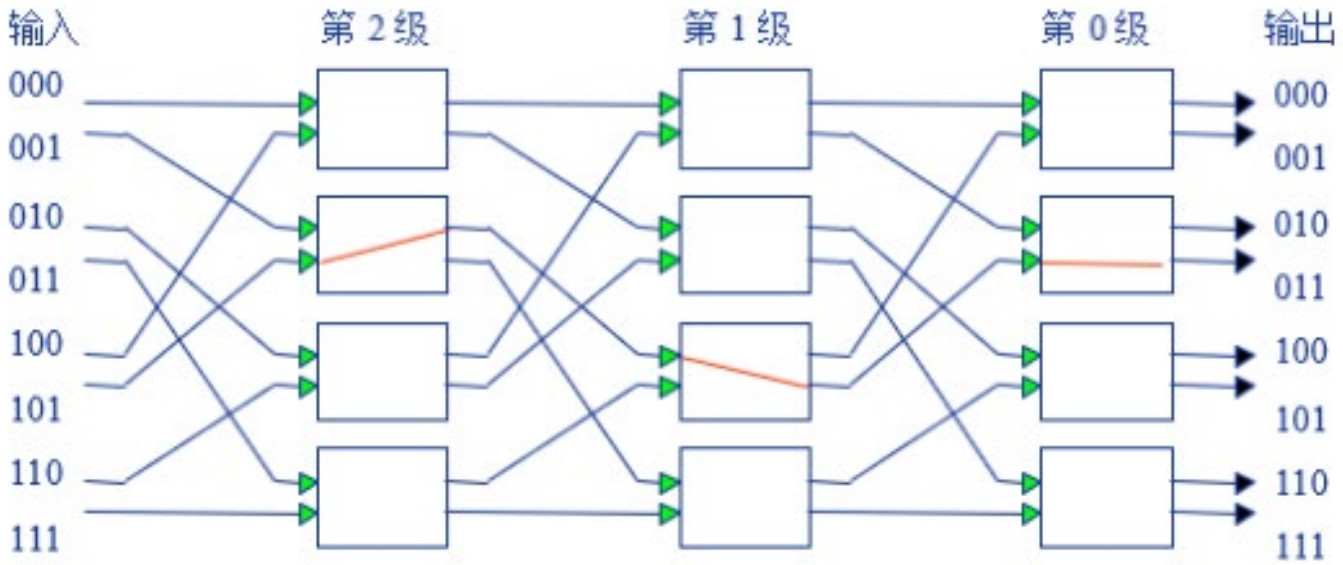


- 一个N输入的Omega网络
  - 有$\log_2 N$级，每级用$N/2$个2×2开关模块，共需要$N\log_2 N/2$个开关。
  - 每个开关模块均采用单元控制方式。
  - 不同的开关状态组合可实现各种置换、广播或从输入到输出的其它连接。

多级混洗—交换网络寻径算法

目的：根据给定的输入/输出对应关系，确定各开关的状态。
名称：源-目的地址异或法
操作：将任一个输入地址与它要到达的输出地址作异或运算，其结果的bit_i位控制数据到达的第i级开关，“0”表示“直连”，“1”表示“交换”。

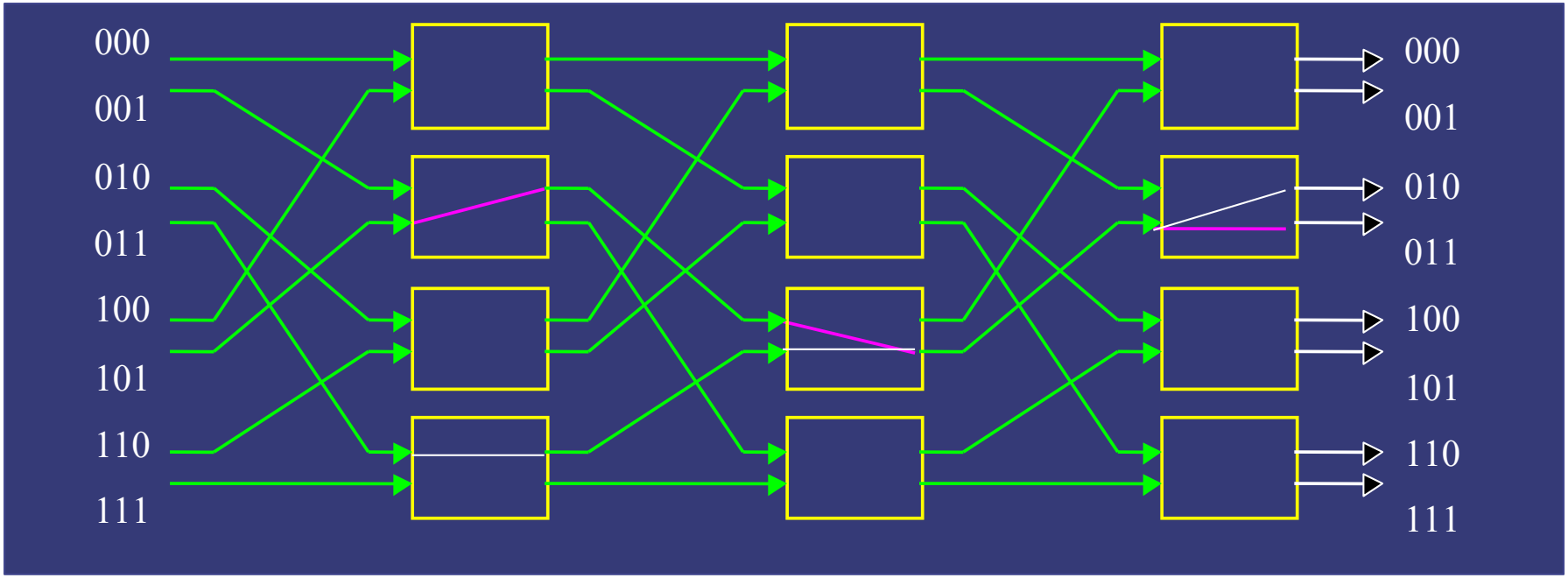
例如给定传输101B→011B，二者异或结果为110B，于是从101B号输入端开始，把它遇到的第2级开关置为“交换”，第1级开关置为“交换”，第0级开关置为“直连”。如下图红线所示。



Omega网寻径冲突

给定传输101B→011B，二者异或结果为110B，路径如下图红线所示。

给定传输011→010B，二者异或结果为001B，路径如下图白线所示。



寻径算法练习与阻塞分析

课堂练习：

寻径置换 $\pi_1=(0,7,6,4,2)(1,3)(5)$， $\pi_2=(0,6,4,7,3)(1,5)(2)$。试问：①哪个置换出现阻塞？②Omega网络一次通过可实现的非阻塞置换共有多少个

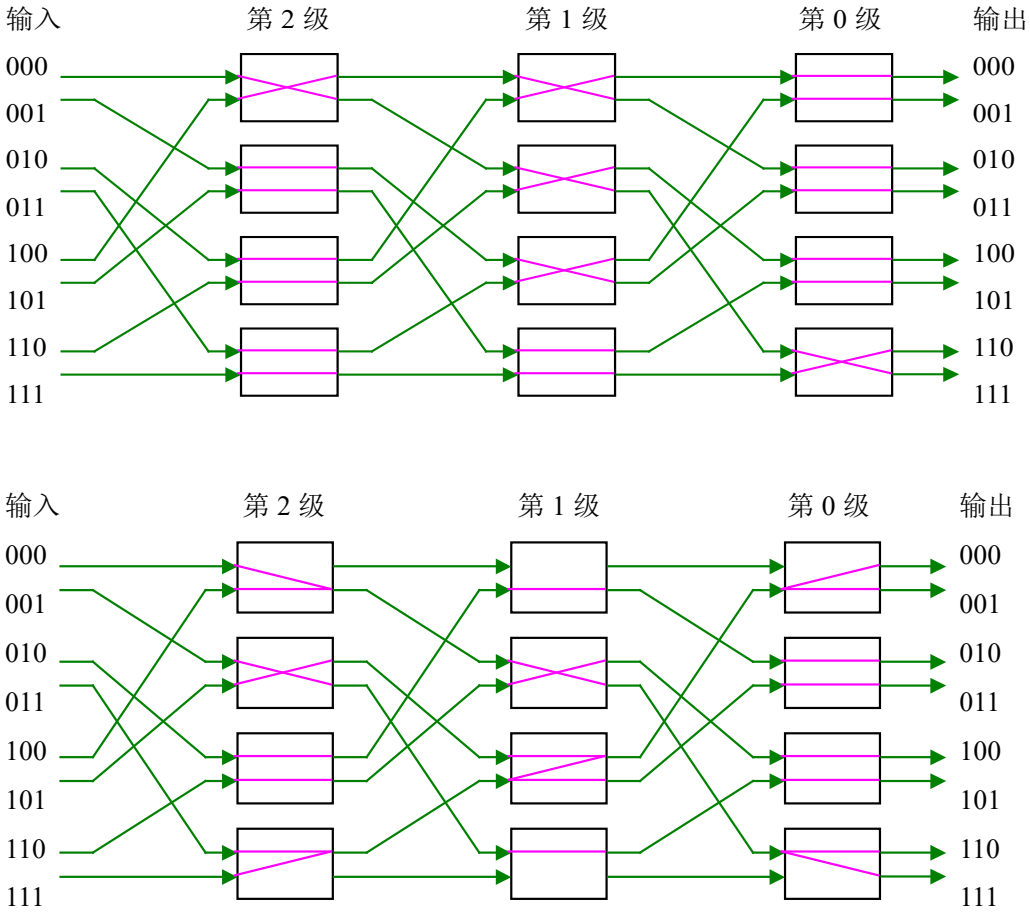
解答：按寻径算法得右图，可见 π_1没有阻塞， π_2有阻塞。

阻塞条件分析（充分条件）

①（必要条件1）：两个消息在某一级到达同一个开关。判断方法是比较它们在各级的行号，是否在第i级恰好仅第i位不同；

②（必要条件2）：两个消息在到达的同一个开关上操作相反。判断方法是逐位比较双方的控制函数。

置换数计算：



课堂练习题

(1)画出2×2开关构成的16个输入端的Omega网络；

(2)结点1011传送消息给结点0101，同时0111传送消息给结点1001，画出完成这一寻径的开关设置。这种情况会出现阻塞吗？

(3)试计算这个Omega网络一次通过实现的置换个数，一次通过实现的置换个数占全部置换的百分比是多少？

Homework:

9.8, 9.9, 9.12, 9.13

1. 从64个结点中的56号结点向3号结点发送数据，分别使用下列互连网络时，求**最少**步数，并写出依次使用的函数名称。
   - (1). 单级立方体网络；
   - (2). 单级混洗-交换网络；
   - (3). 单级加减2^i(即PM2I)网络。