第5章 并行计算机系统结构

Recommend
火龙果软件

Agenda

- 5.1 并行处理器系统
- 5.2 并行计算机的互连网络
- 5.3 芯片级并行性
- 5.4 系统级并行性

Recommend
火龙果软件

5.1 并行处理器系统

并行性的开发

- 资源重复
  - 硬件设备的多重设置
  - 阵列机、多核和多处理器系统
- 时间重叠
  - 多个处理过程在时间上错开
  - 流水线处理方式
- 资源共享
  - 分时运行
  - 多进程、多线程

Recommend
火龙果软件

并行性的层次

- 操作级
  - 流水
  - 数据并行
- 指令级
  - 相关性分析
  - 指令调度
- 循环级
  - 展开
  - 推测执行
  - 路径调度
- 线程级
  - 并行多线程、多核与多处理器
  - 数据缓存的一致性
  - 同步与互斥的有效性

Recommend
火龙果软件

多处理机系统结构

多处理机系统的结构分类

- UMA tightly coupled
  - SMP symmetrical multiprocessor
  - non SMP
- NUMA
  - DSM
  - MPP loosely coupled

Recommend
火龙果软件

UMA

Uniform memory access

☺ 可编程性

- 兼容性
- 简单的程序设计模型
- 通信开销小

☹ 可扩展性

- 共享存储器结构使得存储器成为瓶颈

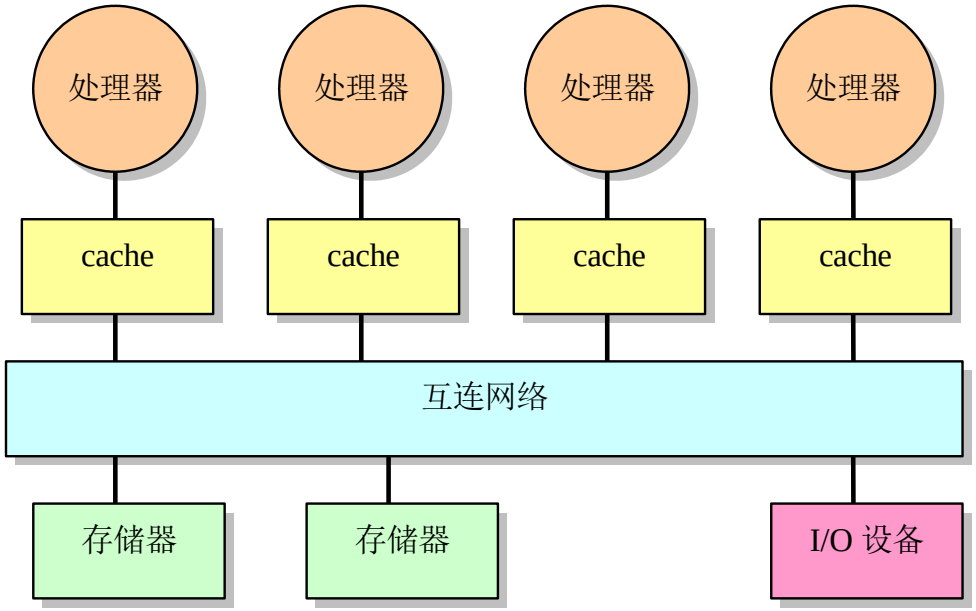
☹ 访存速度要求

- 每个数据都来自共享的存储器

Recommend
火龙果软件

UMA的实现

- 集中式存储器
- 问题
  - 访存冲突
  - 网络带宽
- 解决方案
  - 多体交叉
  - 本地指令存储器
  - 本地cache



Recommend
火龙果软件

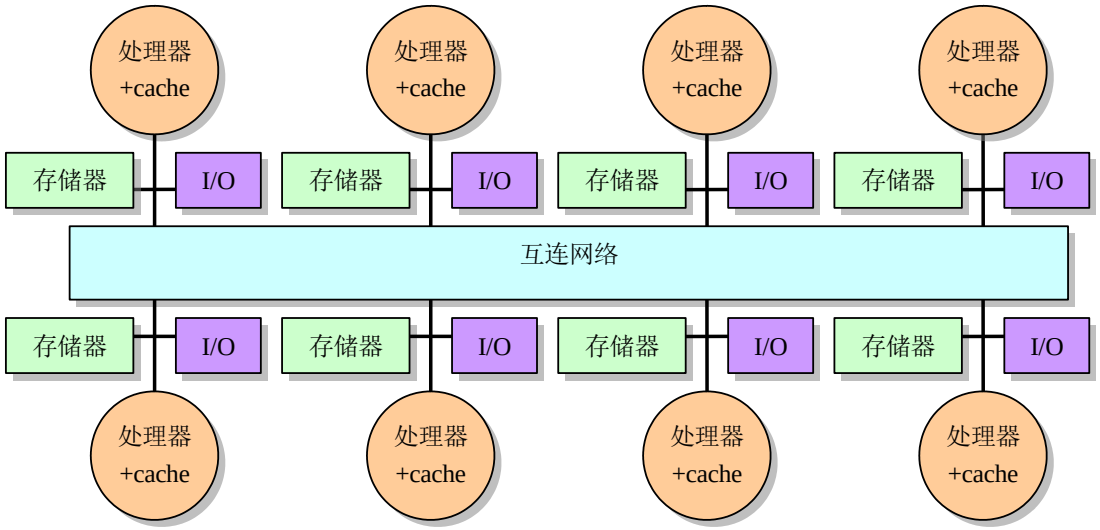
NUMA

- DSM distributed shared memory
  - 用硬件机制将分布的存储器构成一个逻辑上统一的共享存储器
  - cache及其一致性问题
  - CC-NUMA cache coherent NUMA
  - COMA cache only memory access
- MPP massively parallel processors
  - 大规模并行处理机系统
  - 消息传递
    - 显式数据交换
  - 程序设计较难

Recommend
火龙果软件

NUMA的实现

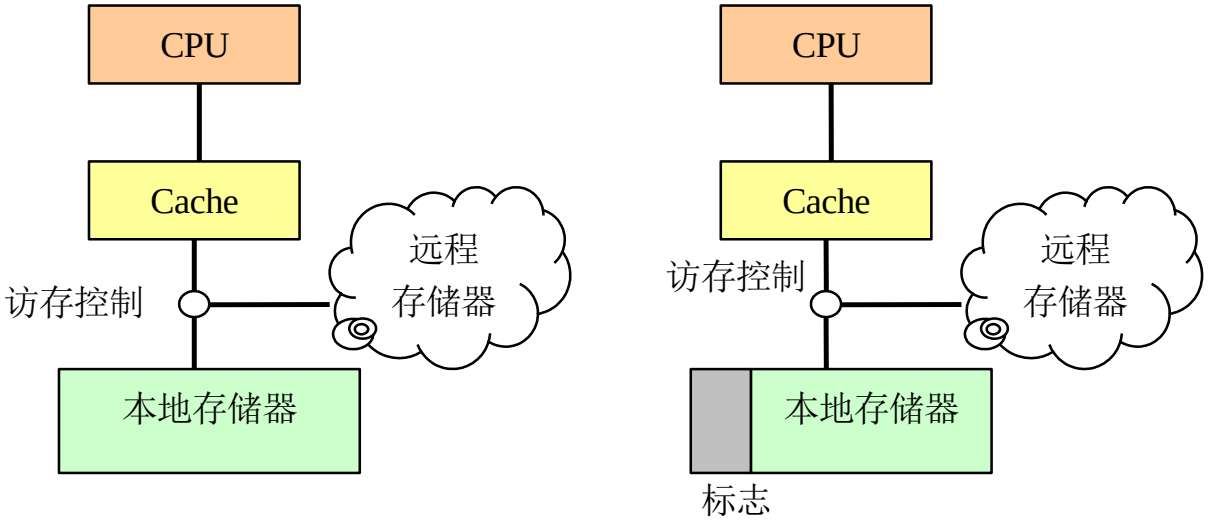
- 分布式存储器
  - 本地存储器延迟较小
  - 扩展性较好
  - 性价比高
  - 数据一致性问题
  - 编程复杂



Recommend
火龙果软件

DSM的结点

- CC-NUMA
  - 本地cache失效时数据将根据地址从本地存储器中或者远程存储器访问
  - Cache命中率低
- COMA
  - 增加每个数据块的命中标志信息和状态信息
  - 数据块可迁移

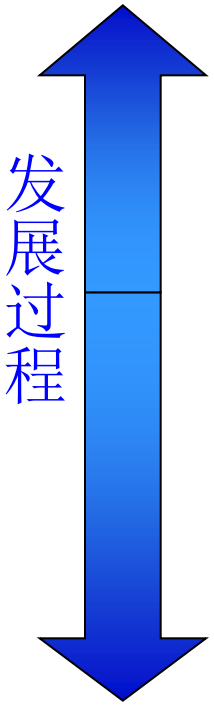


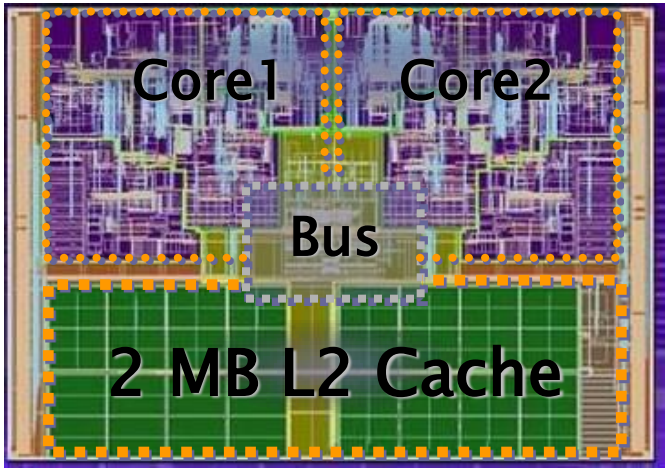
(a) CC-NUMA (b) COMA

Recommend
火龙果软件

多处理机系统的结构分类

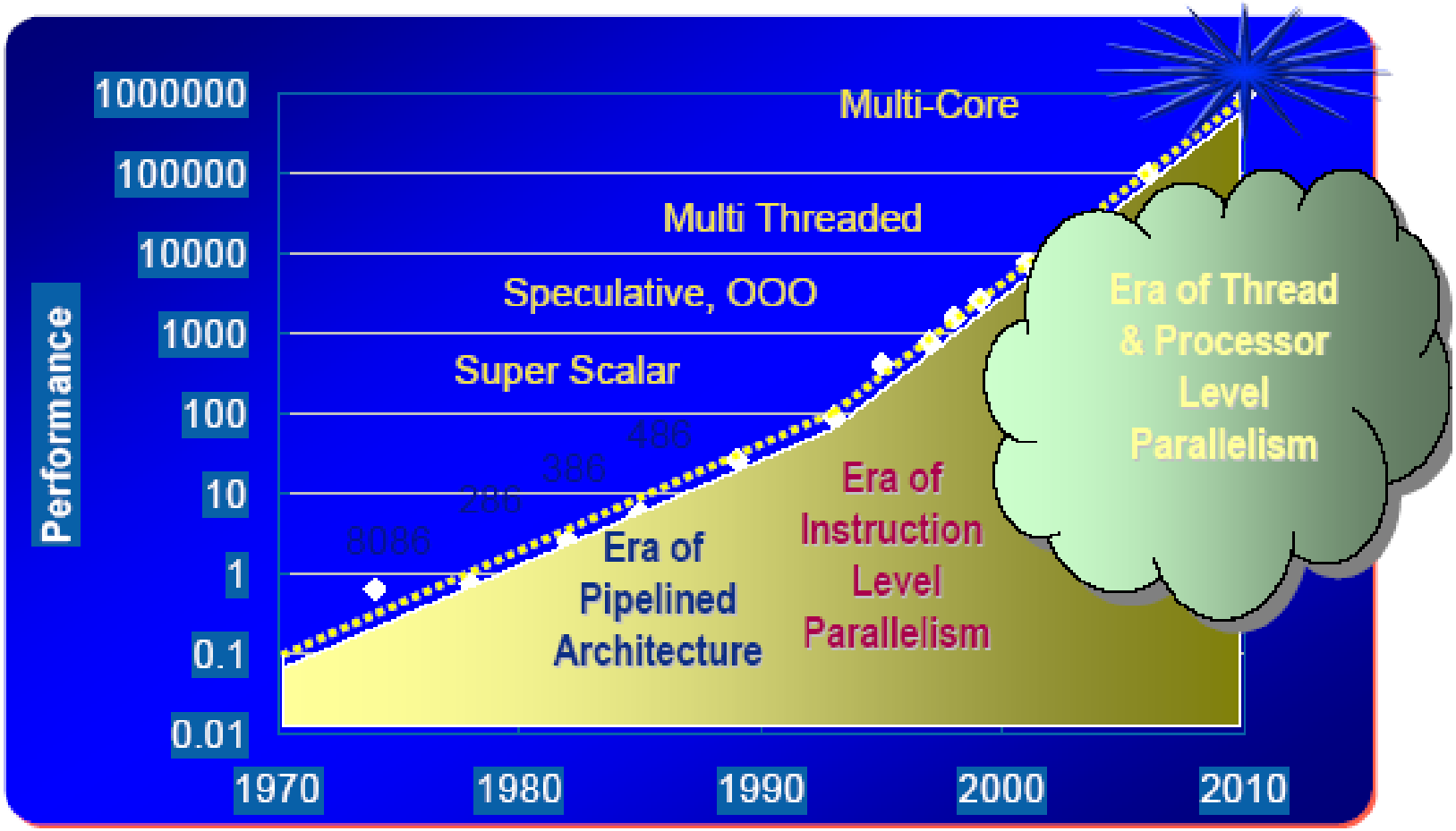
- 按硬件层次
  - 网络级
    - 网格
  - 机架级
  - 板级
    - 多CPU芯片主板
  - 芯片级
    - 芯片多处理器CMP
      - 多核





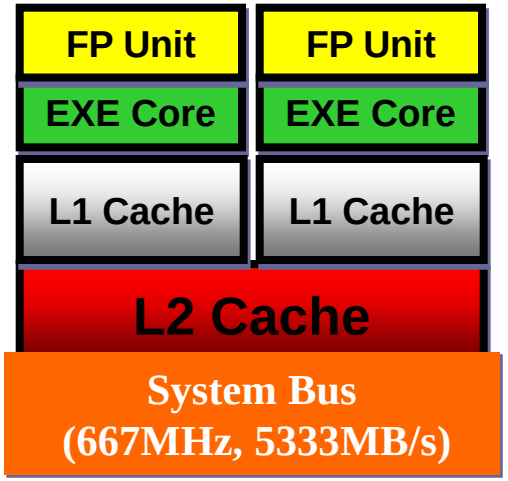
Recommend
火龙果软件

The New Era in Computing



芯片级并行性

- 特点
  - 独立的控制流
  - 分离的内部状态
  - 没有共享的功能部件
- 分类
  - 同构多核
    - Core 2 Duo
  - 异构多核
    - Cell处理器
    - 网络处理器



Recommend
火龙果软件

芯片级并行性

性能比较

- CPU芯片降低15%工作电压的结果
  - 主频降低15%
  - 功耗降低45%
  - 性能下降10%

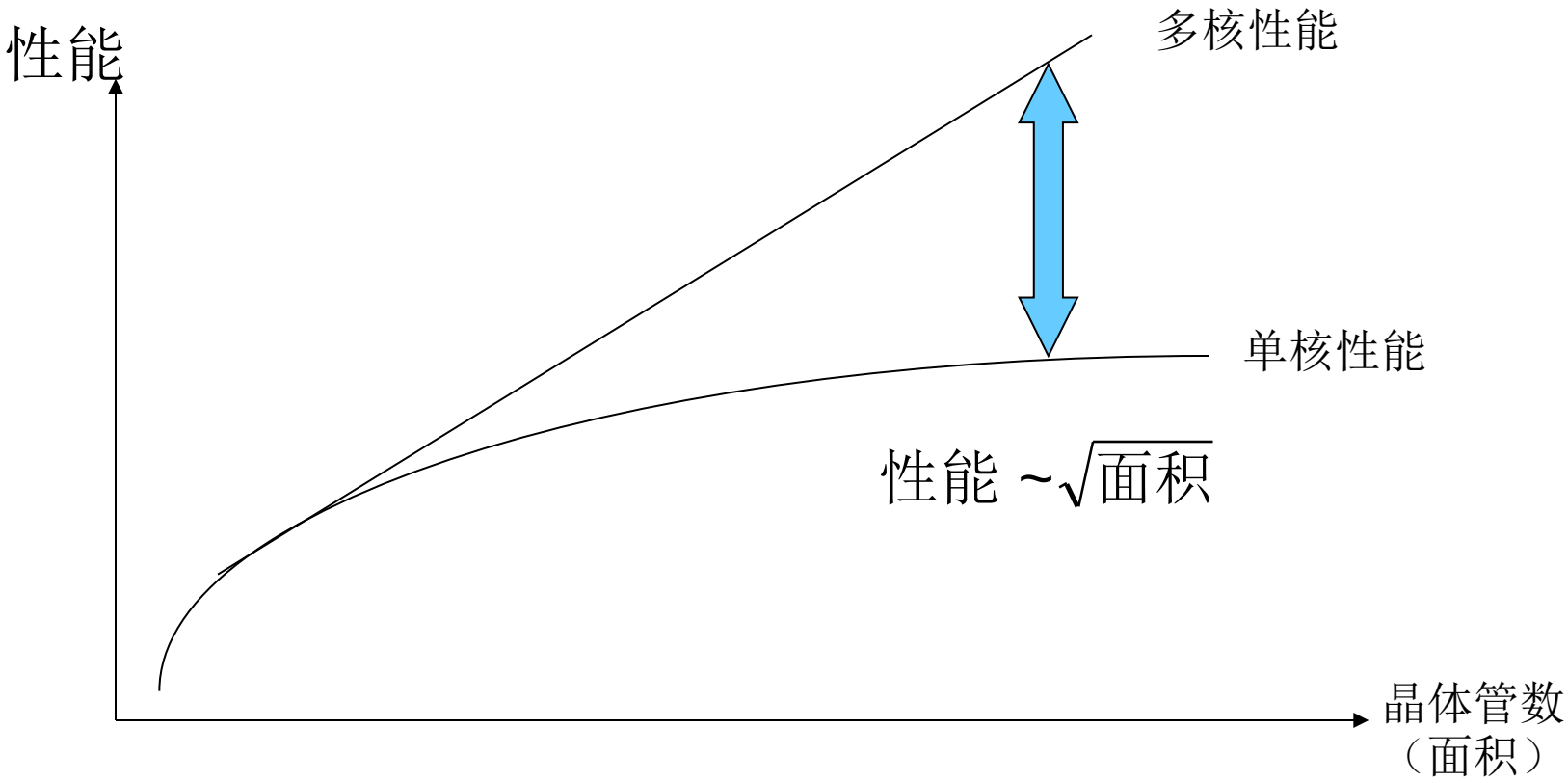
- 单核
  - 芯片面积=1
  - 电压 = 1
  - 主频=1
  - 功耗=1
  - 性能=1

- 双核
  - 芯片面积=2
  - 电压 = 0.85
  - 主频=0.85
  - 功耗=1.1
  - 性能=1.8

Recommend
火龙果软件

芯片级并行性

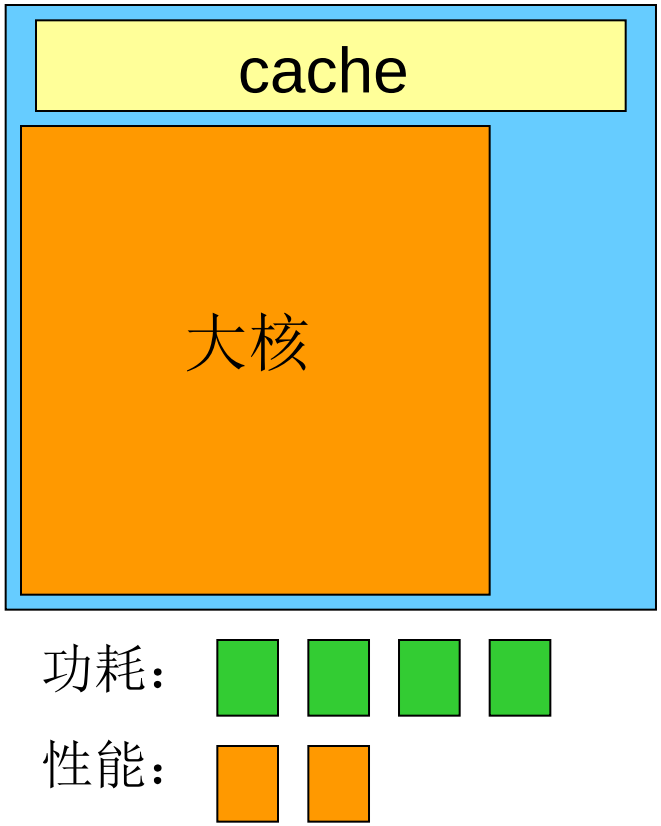
- 芯片面积与性能

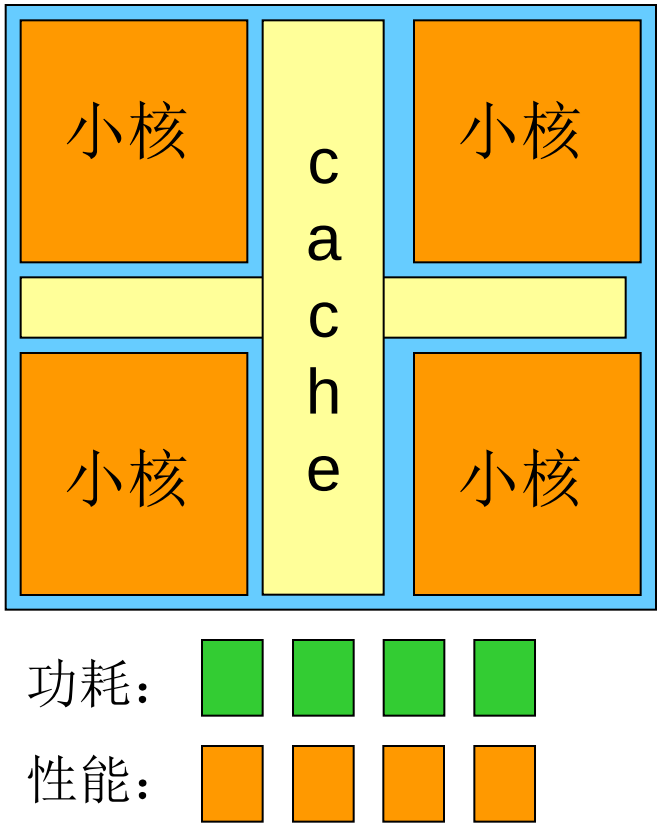


Recommend
火龙果软件

芯片级并行性

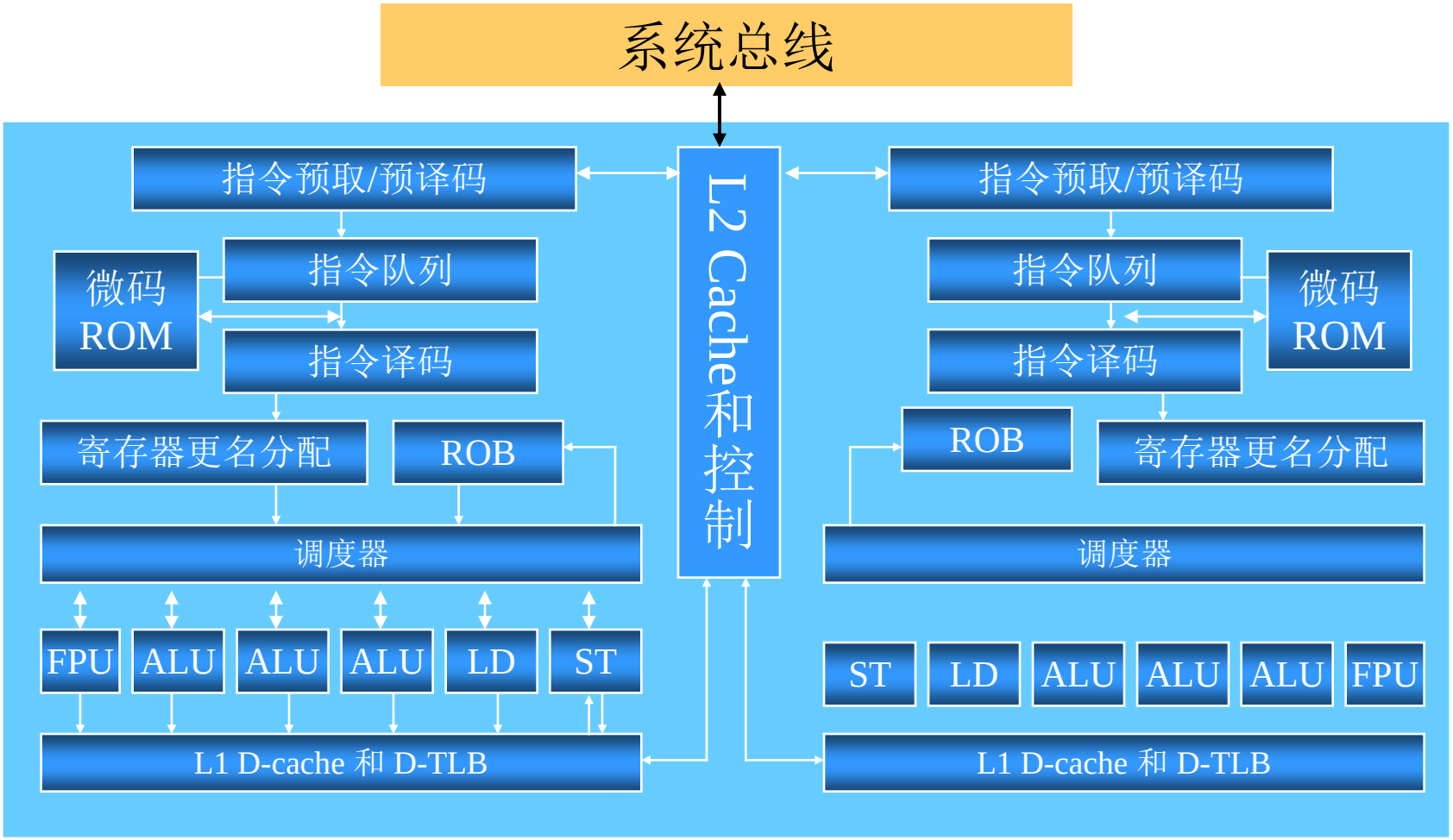
- 一个大核与多个小核的比较





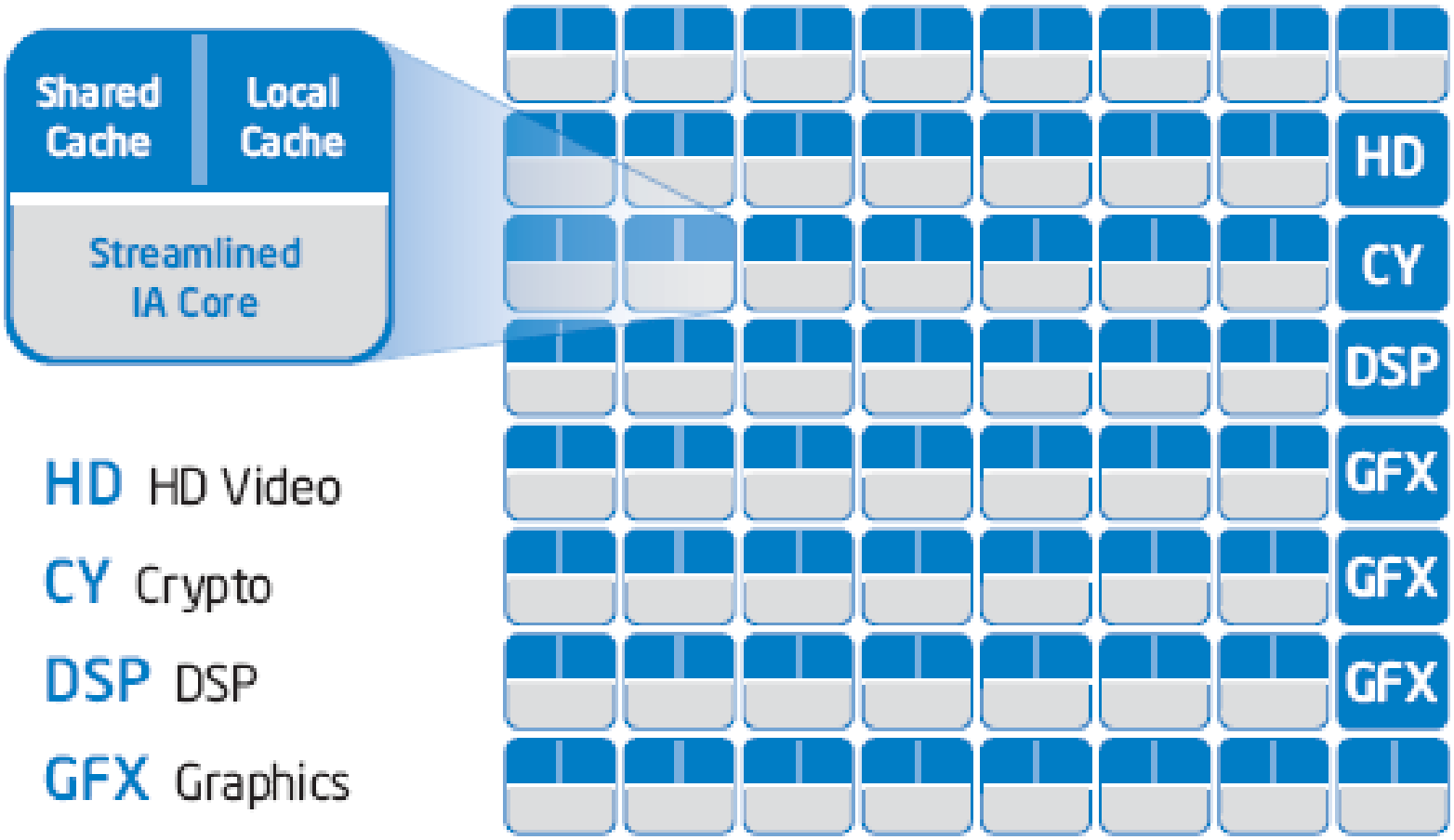
Recommend
火龙果软件

Core 2 Processor Block Diagram



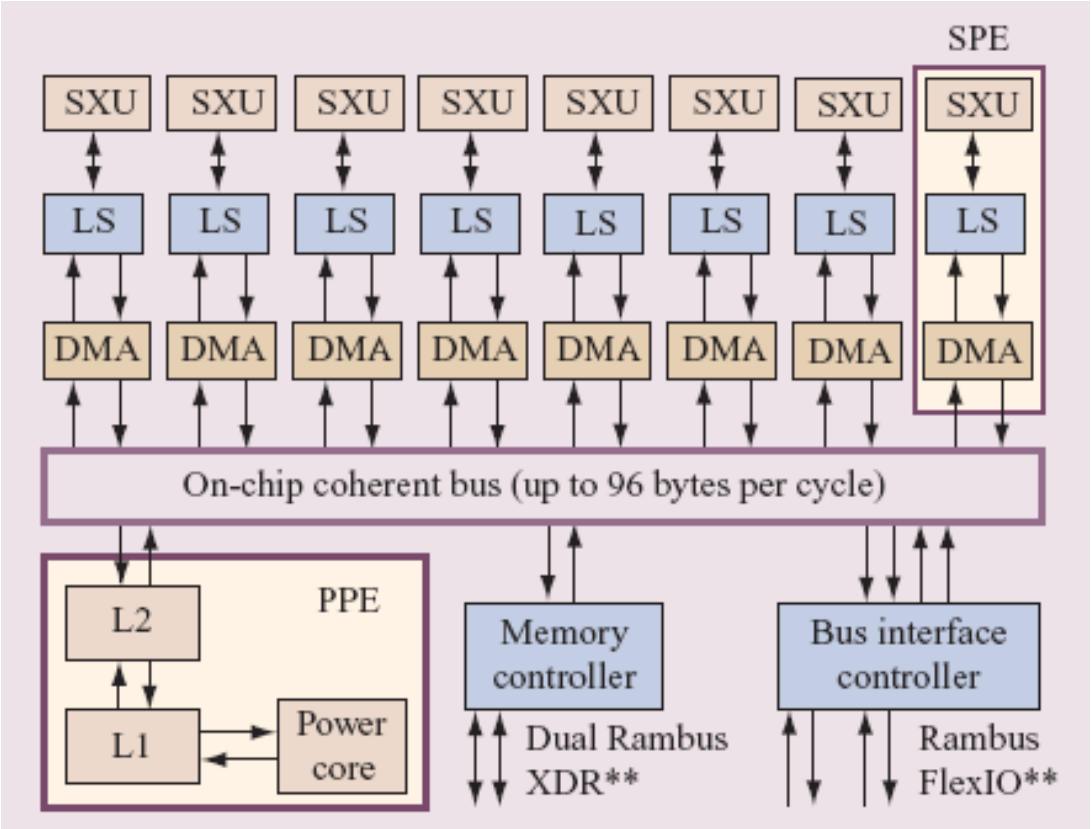
Recommend
火龙果软件

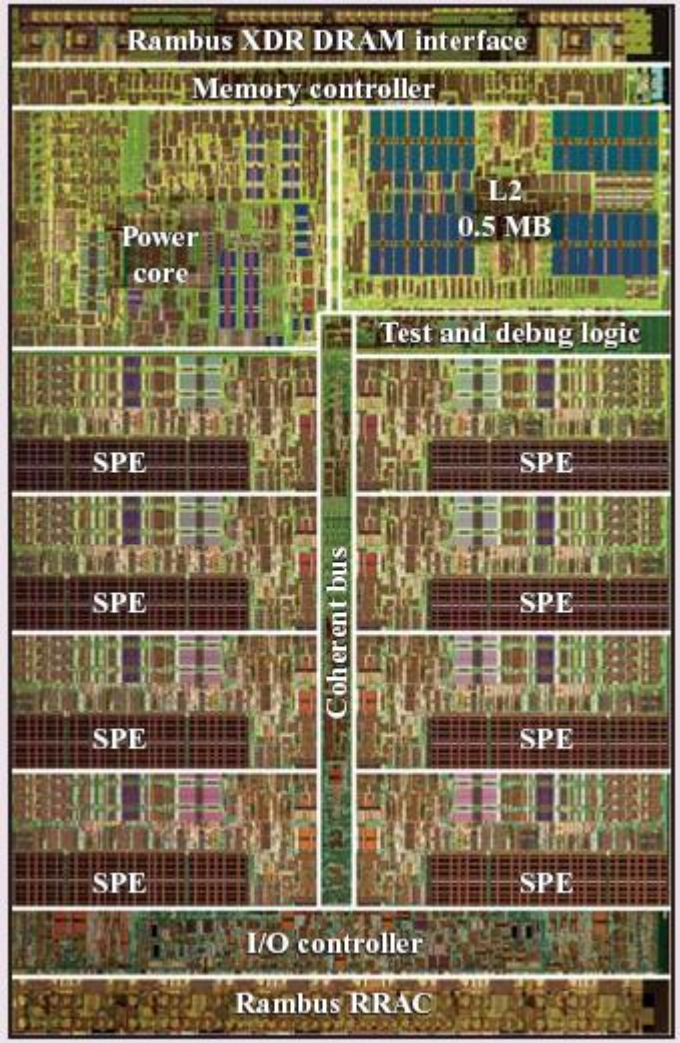
Intel's tera-scale chip



Recommend
火龙果软件

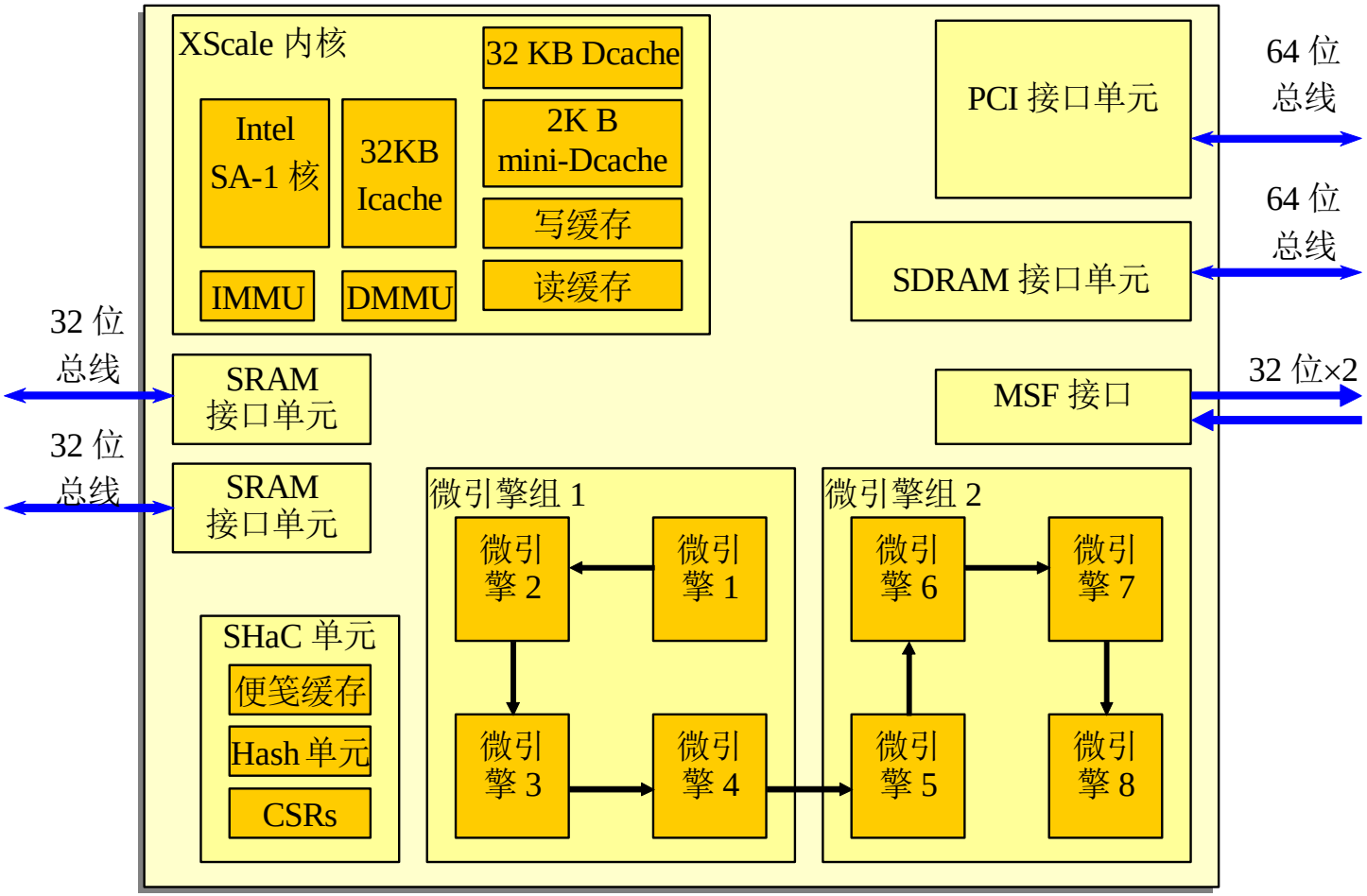
IBM-Sony Cell





Recommend
火龙果软件

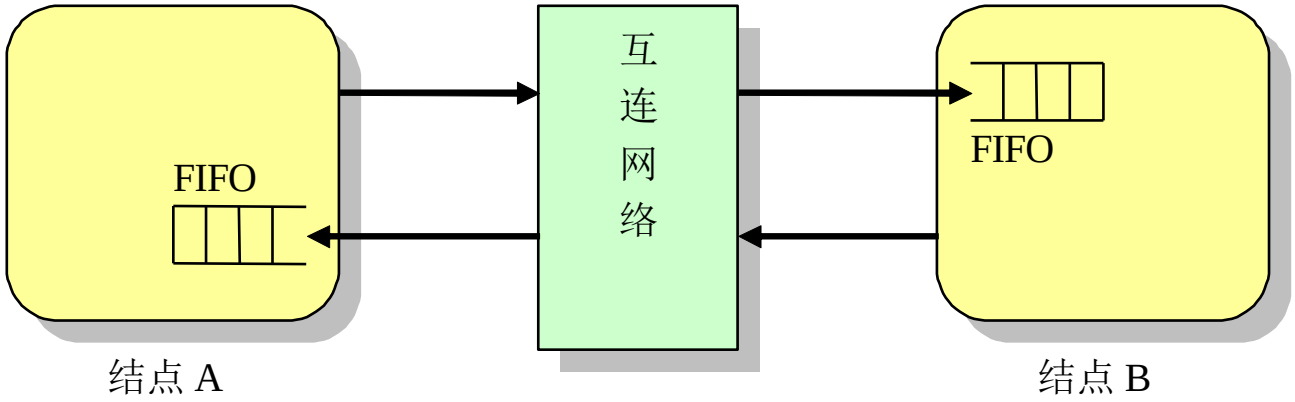
网络处理器（IXP2400）



Recommend
火龙果软件

消息传递接口

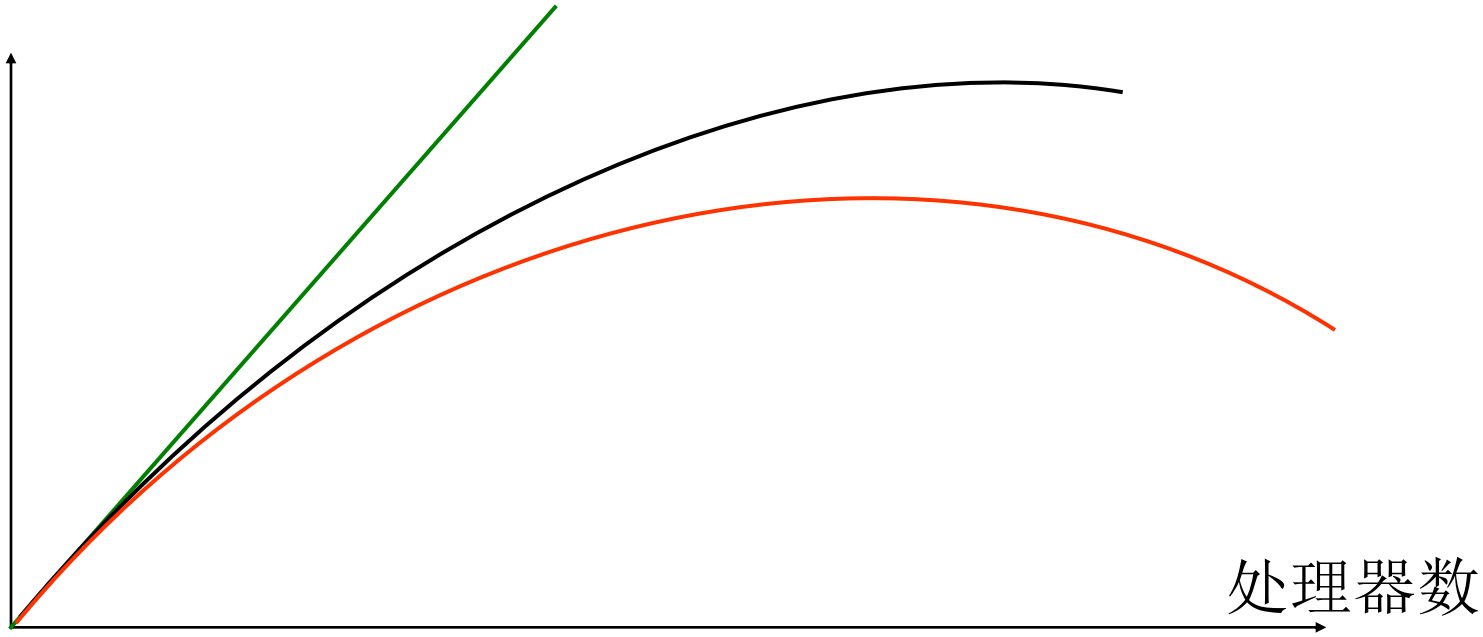
- 消息缓存
- 互连网络
- 处理器间中断（IPI）的机制



Recommend
火龙果软件

并行处理器系统的性能

- 程序的执行时间
  - 并行执行部分 + 通信时间部分 + 串行执行部分
- 加速比

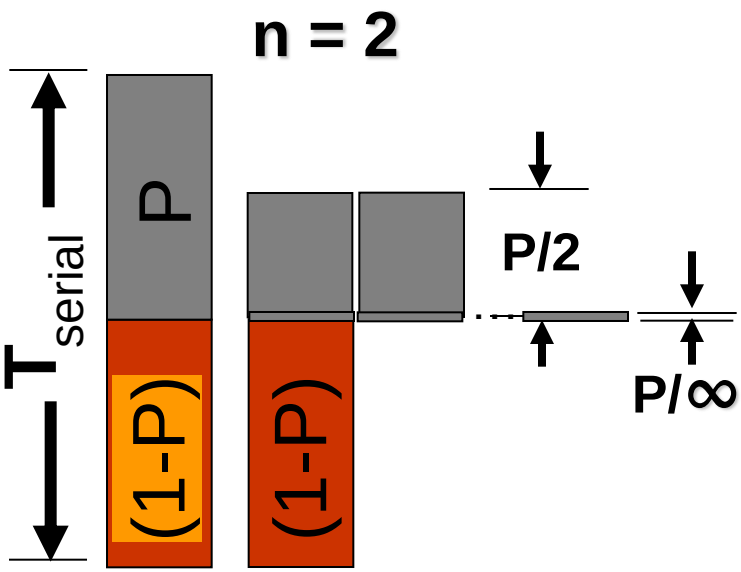


Recommend
火龙果软件

并行处理器系统的性能

- Amdahl定律
  - 通常用于分析功能部件级的性能加速比
  - 未考虑存储系统的性能影响
- Gustafson加速比模型
  - 要求在衡量并行计算机加速比的时候相应地增加问题规模
  - 使得每个处理器分得的问题规模（如数据量）保持不变
  - 设a为串行部分的执行时间比例，(1-a)为并行部分的执行时间比例，*n*为处理器数，加速比为

$$a + (1-a)n$$



Recommend
火龙果软件

并行处理器系统的性能

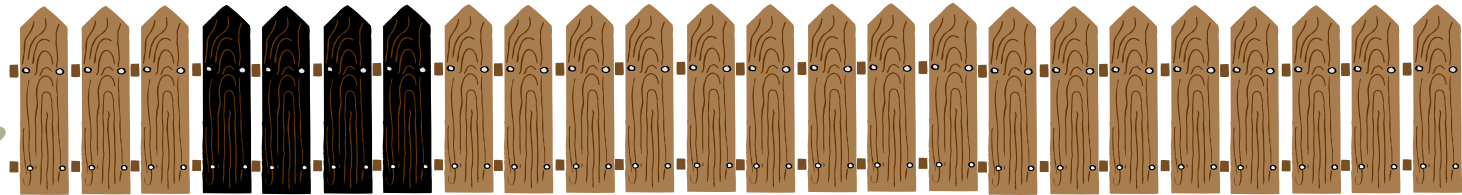
- 可伸缩性（scalability）
- 限制可伸缩性的因素
  - 互连网络的复杂性和通信延迟
  - 编程的复杂性
  - 通信和同步开销

Recommend
火龙果软件

Amdahl定律

- 加速比的例子

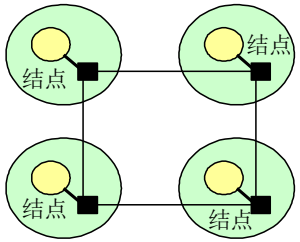
Number of painters	Time	Speedup
1	30 + 300 + 30 = 360	1.0X
2	30 + 150 + 30 = 210	1.7X
10	30 + 30 + 30 = 90	4.0X
100	30 + 3 + 30 = 63	5.7X
Infinite	30 + 0 + 30 = 60	6.0X



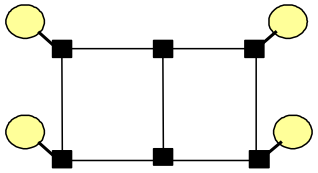
火龙果软件

5.2 并行计算机的互连网络

- 拓扑结构
  - 静态网络
    - 结点包括一个功能部件和连接网络的开关元件
    - 一维，二维，多维
  - 动态网络
    - 开关元件构成
    - 处理器等功能部件位于网络的边缘
    - 单级，多级



（a）静态网络


（b）动态网络

Recommend
火龙果软件

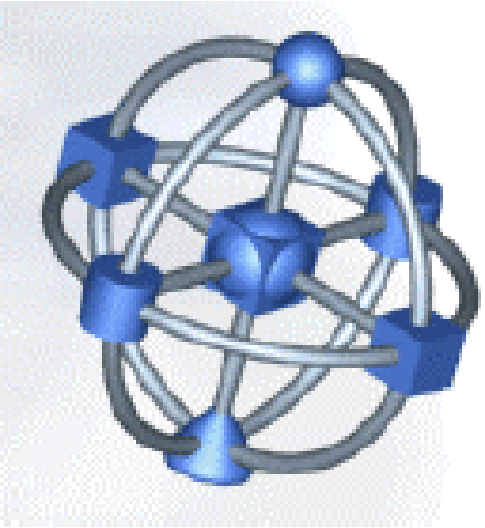
并行计算机的互连网络

- 控制方式
  - 集中控制
  - 分布控制
- 传递方式
  - 单播
  - 多播
  - 广播
- 链路类型
  - 共享链路
  - 专用链路
- 实现方式
  - 片内网络
  - 板内网络
  - 机架内网络
  - 机架间网络

Recommend
火龙果软件

并行计算机的互连网络

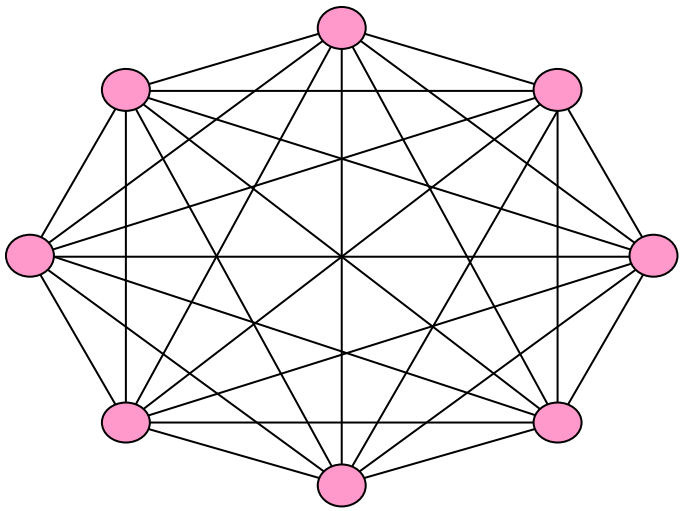
- 互连网络特性
  - 连接性
    - 阻塞，冲突
  - 规整性
- 静态网络的参数
  - 度
  - 直径
  - 带宽总和aggregate bandwidth
  - 对分带宽bisection bandwidth



Recommend
火龙果软件

静态互连网络

- 全互连网络fully connected network
  - 度=N-1
  - 直径=1
  - 链路数=$N(N-1)/2$
  - 对分带宽?
- 优点：结点间通信距离短
- 缺点：成本高，实现困难



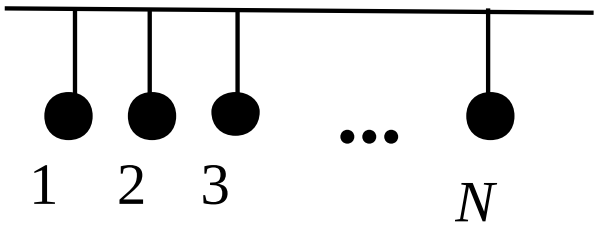
连接方式: $\forall x$，$0 \leqslant x \leqslant N-1$

$x \Leftrightarrow \forall x'$，$0 \leqslant x' \leqslant N-1$，且 $x \neq x'$。

Recommend
火龙果软件

总线型网络

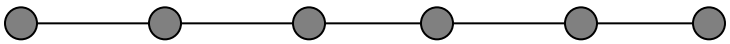
- **单总线结构**single bus
  - 度=1
    - 分时使用
  - 优点
    - 结构简单
    - 成本低廉
    - 容易实现
  - 缺点
    - 使用冲突

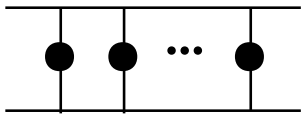


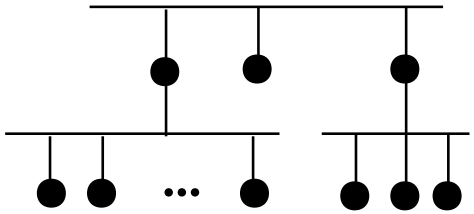
Recommend
火龙果软件

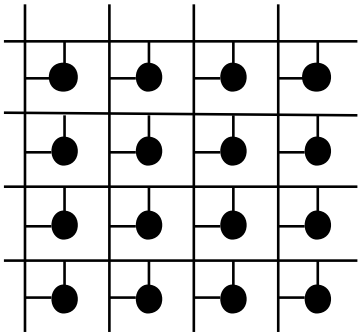
总线型网络

- **多总线结构**
  - 度=总线数
- **多级总线结构**
  - 分级的多总线结构
- **二维总线结构**
- 总线的分割





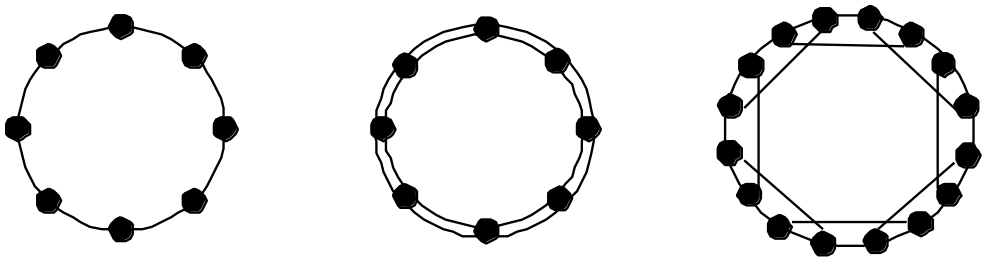




Recommend
火龙果软件

环型网络

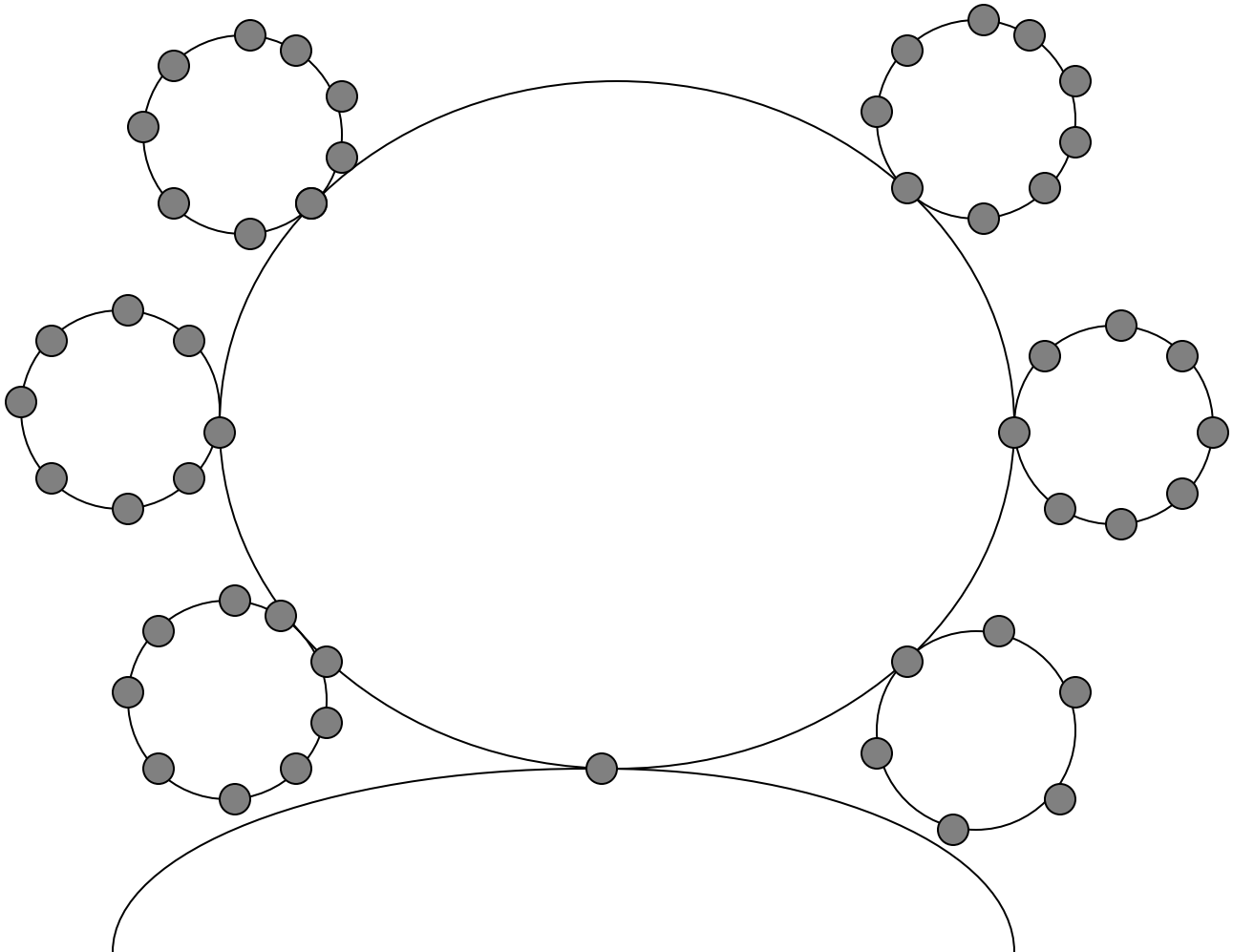
- 单环网络single ring
  - $x \leftrightarrow (x \pm 1) \bmod N$
  - 直径=? 度=?
  - 寻径算法简单，可同时传送多个信息，吞吐率比单总线高。
- 双环网络：增加吞吐率和可靠性
- 层次多环网络：可伸缩性
- 带弦环型网络chordal ring



(a) 单环结构　(b) 双环结构　(c) 带弦环结构

Recommend
火龙果软件

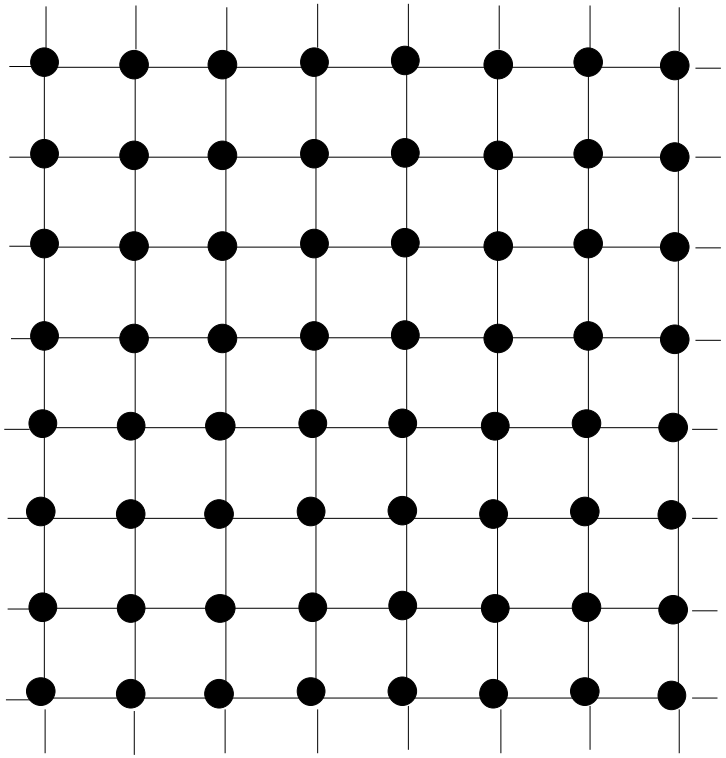
层次多环网络



Recommend
火龙果软件

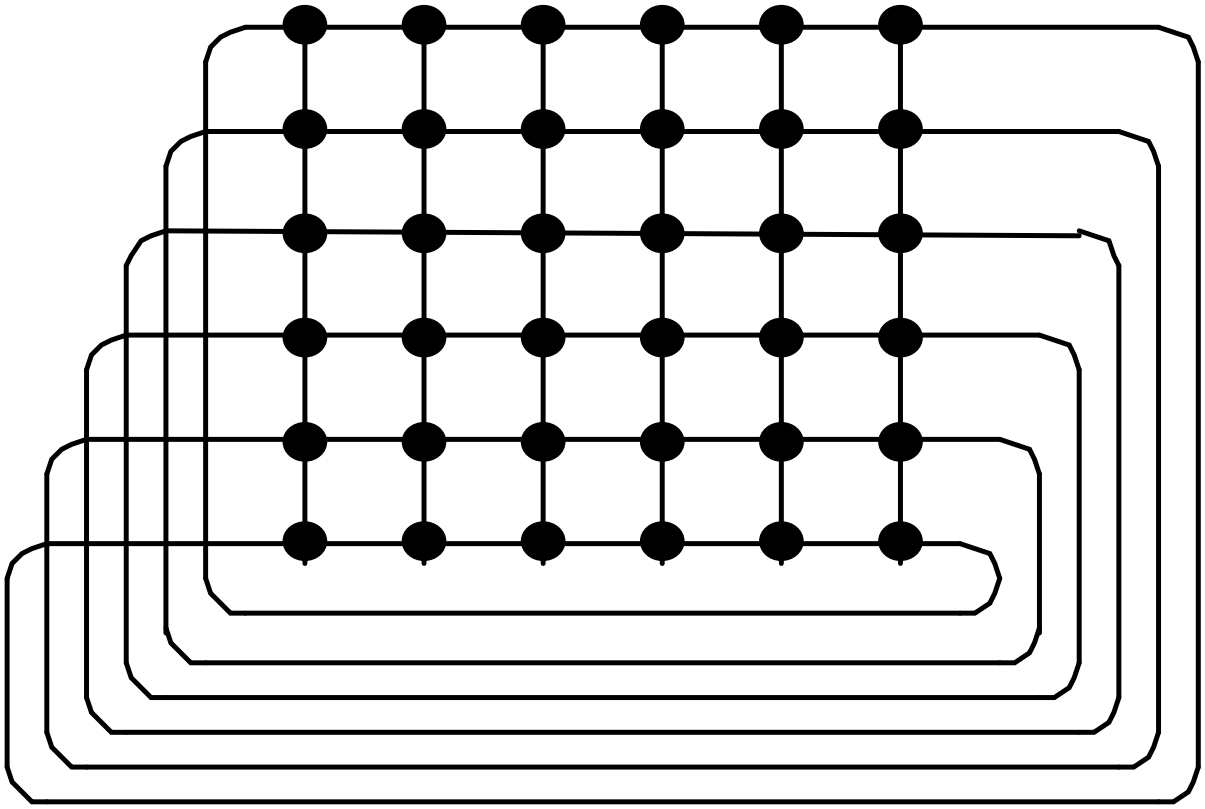
二维网格型网络mesh

- 度=4
- 直径=2(n-1)
- 对分带宽 = n
- 链路总数=?
- 优点：寻址简单，度不变
- 缺点：流量不对称，伸缩性差
- 绞带环、双绞螺面、带环网格和闭合螺面
- 网格的推广：网孔形



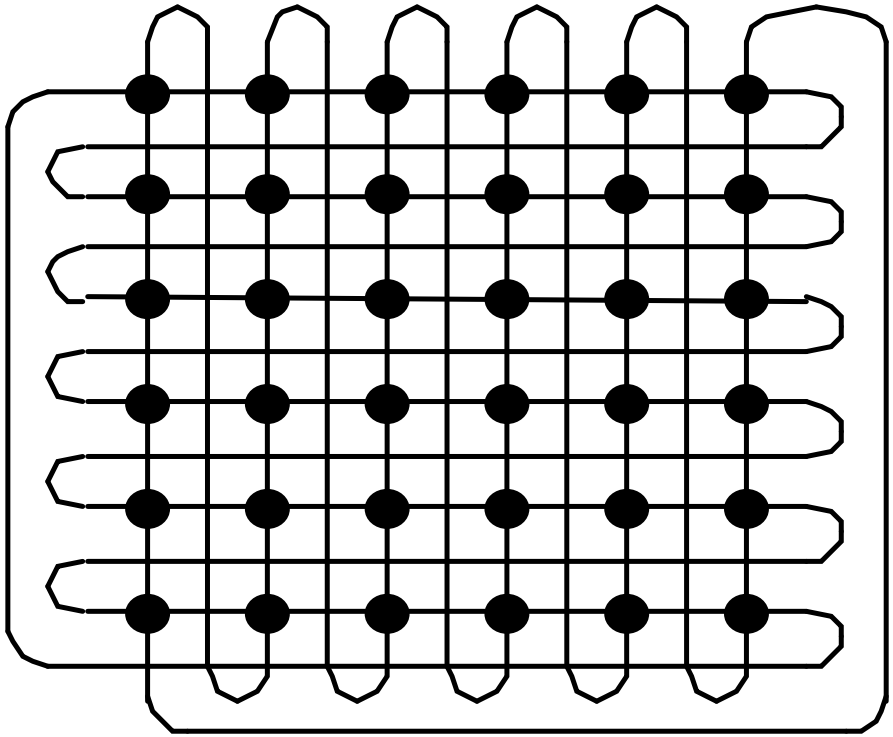
Recommend
火龙果软件

绞带环



Recommend
火龙果软件

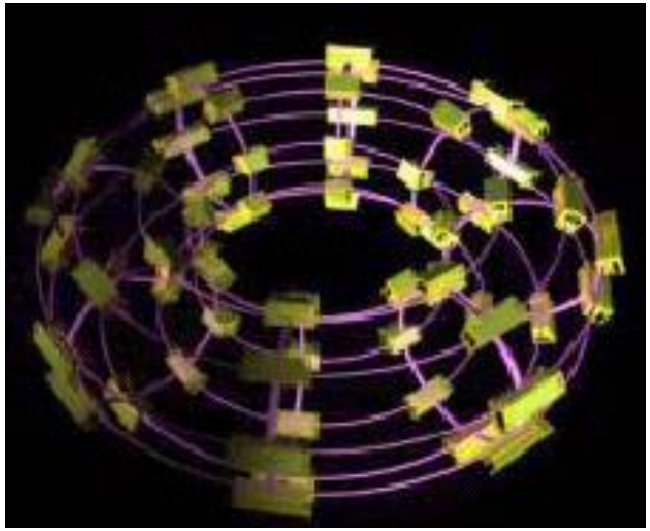
双绞螺面

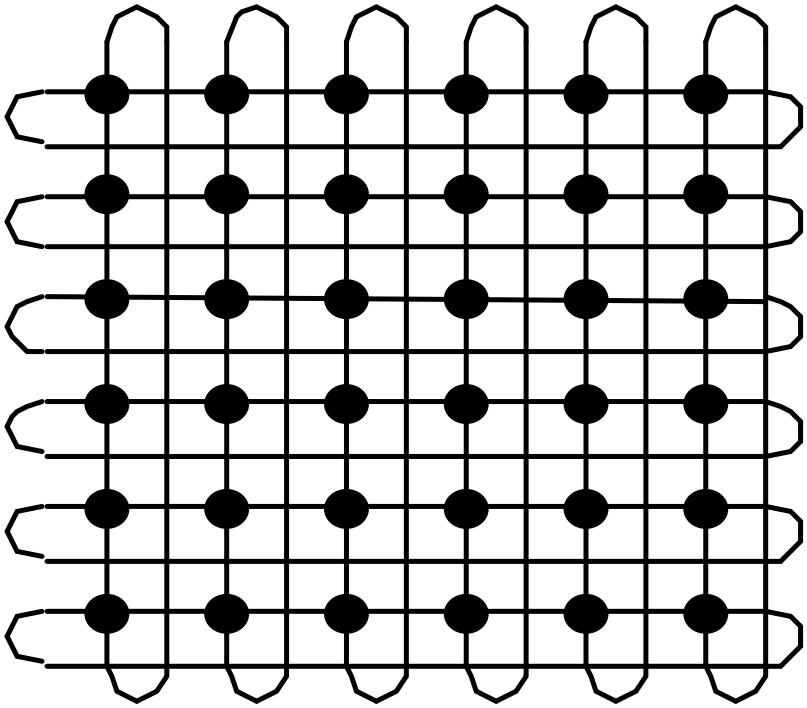


Recommend
火龙果软件

带环网格torus

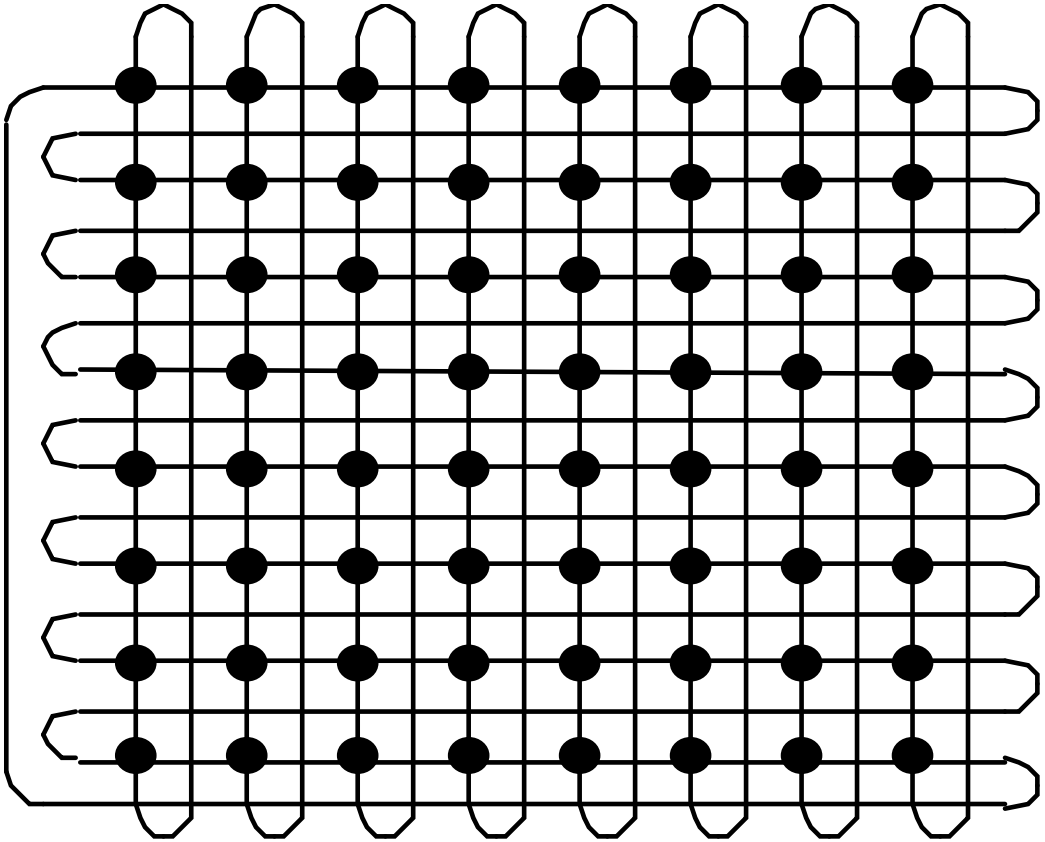
- 度=4
- 直径=?
- 对分带宽=?
- 链路总数=?





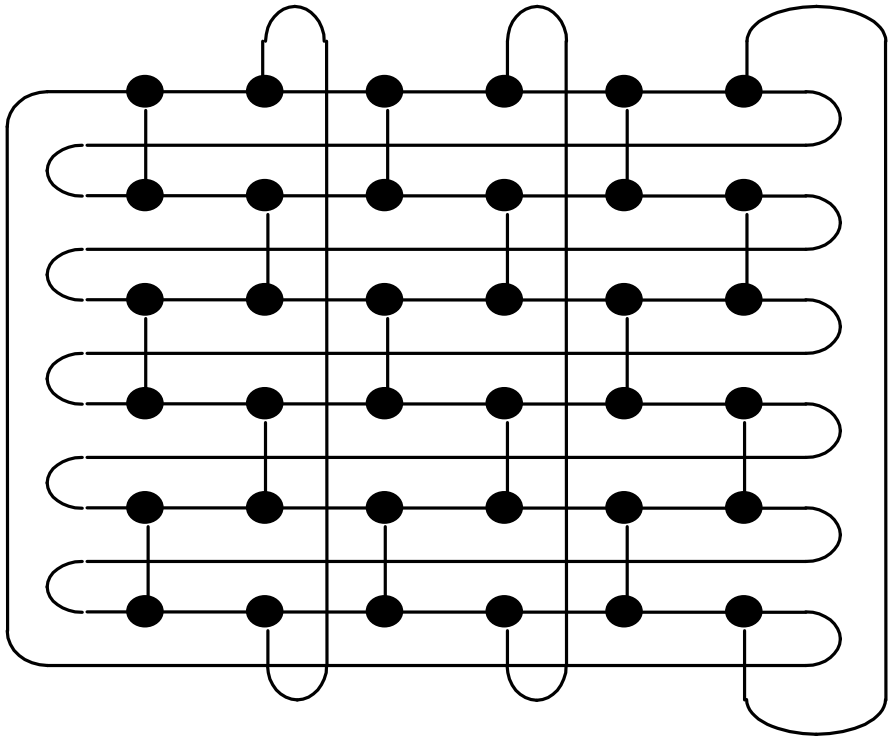
Recommend
火龙果软件

闭合螺面



Recommend
火龙果软件

这是什么**？

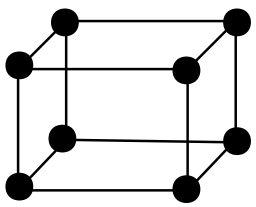


Recommend
火龙果软件

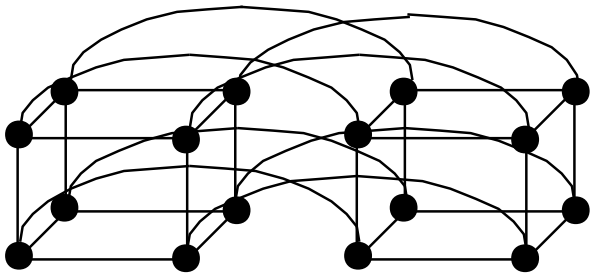
立方体网络

- 二进制超立方体binary hypercube
- 度为n
- 直径$k=n=\log_2 N$
- 优点
  - 结点间的通信距离较短
  - 寻径算法简单
- 缺点
  - 可扩充性差
  - 度随N的增加而增大

$$(a_n a_{n-1} \cdots a_i \cdots a_2 a_1) \leftrightarrow (a_n a_{n-1} \cdots \overline{a}_i \cdots a_2 a_1)$$



(a) 三维网格



(b) 四维网

Recommend
火龙果软件

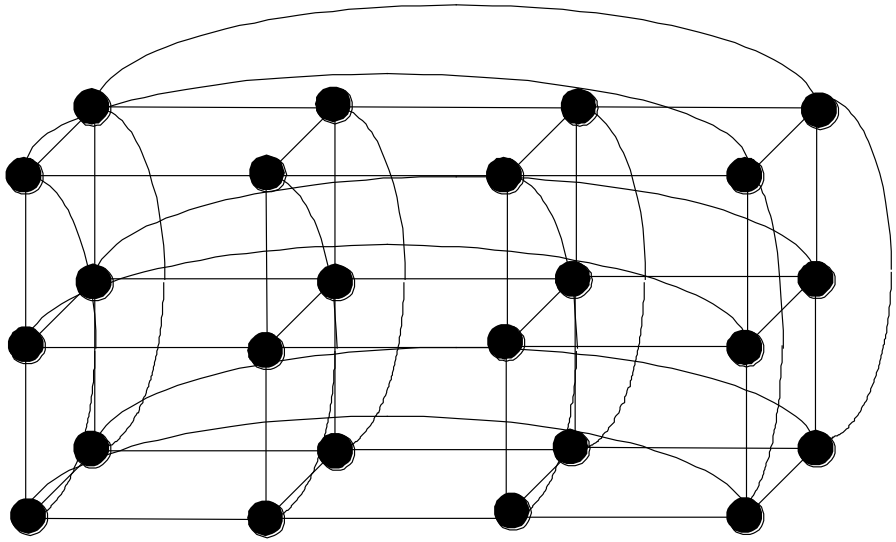
立方体网络

- ***k*-ary *n*-cube网络**
  - 每一维的基数均为k的n维正方体
  - $N=k^n$
  - 度$d=2n$
  - 链路总数$L=n*k^n=nN$
  - 对分带宽为$=2*k^{n-1}$　　（k为大于2的偶数）
  - 直径$D=n\lfloor k/2\rfloor$　　（k为大于2的数）
  - 结点$(x_n x_{n-1}\ldots x_{i+1} x_i x_{i-1}\ldots x_1)$与结点$(x_n x_{n-1}\ldots x_{i+1}((x_i \pm 1) \bmod k) x_{i-1}\ldots x_1)$连接
  - n维立方体网络是k-ary n-cube网络的特例（$k=2$）
  - 带环网格是k-ary n-cube网络的特例（$n=2$）

Recommend
火龙果软件

立方体网络

- 超矩形网络hyper-rectangular
  - $x'_i=(x_i \pm 1) \bmod M_i$
    - 每一维内环形连接
  - 度=$2n$
  - 直径$d = \sum_{i=1}^{n} \left\lfloor \frac{Mi}{2} \right\rfloor$
  - 链路总数$L=nN$
  - 结点总数$N = \prod_{i=1}^{n} Mi$



Recommend
火龙果软件

立方体网络

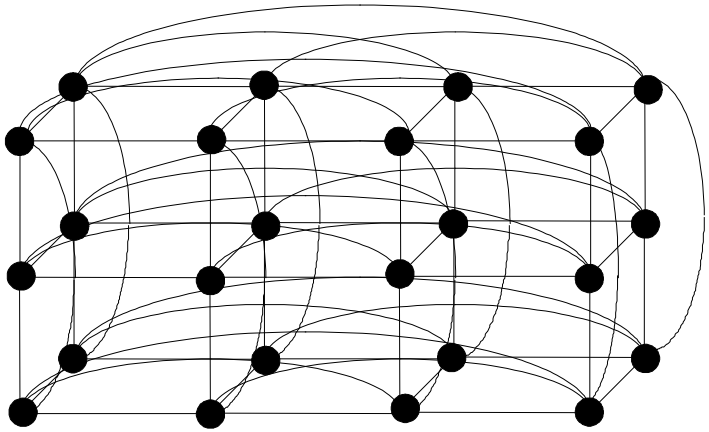
- 一般化的超立方体网络generalized hyper cube
  - 采用混合基数制表示结点的地址
  - 每一维的基数为M_i，$1 \le i \le n$，节点总数$N = \prod_{i=1}^{n} M_i$

$$(x_n x_{n-1} \cdots x_{i+1} x_i x_{i-1} \cdots x_1) \leftrightarrow \forall (x_n x_{n-1} \cdots x_{i+1} x_i' x_{i-1} \cdots x_1)$$

  - 结点的度d为各维链路数之和

$$d = \sum_{i=1}^{n} (M_i - 1)$$

  - 总的链路数为$L = N*d/2$
  - 直径$D=k=n$



（c）2×3×4 的一般化超立方体

Recommend
火龙果软件

例5-1

- 用度与直径的乘积作为性能指标，分析环、二维网格和*n*维立方体这三种静态网络在结点数为4、16、64、256、1024和4096时的性能指标，指出哪一种网络是最好的拓扑结构。

Recommend
火龙果软件

解

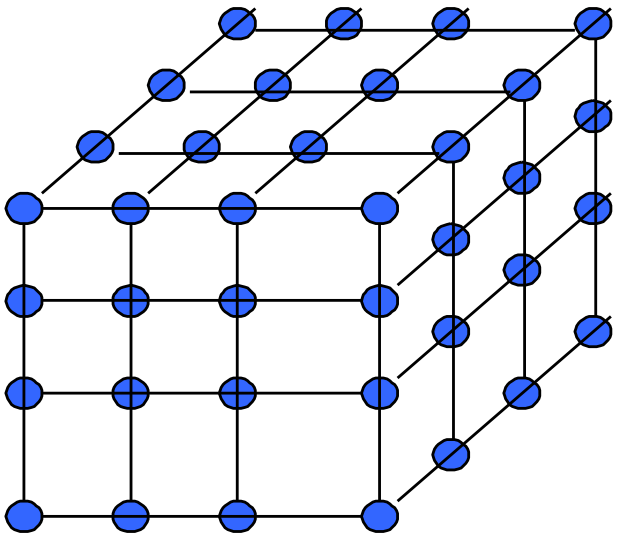
- 在环形网络中，度d=2，直径D=N/2，乘积为N。
- 在网格网络中，d = 4（N>4），D = 2 (-1)，乘积为8 (-1)，（N = 4时的网格退化为环，d=2，乘积为4）。
- 在立方体网络中，d=log2N，D=log2N，乘积为(log2N)2。

N	**4**	**16**	**64**	**256**	**1024**	**4096**
环	4	16	64	256	1024	4096
网格	4	24	56	120	248	504
立方体	4	16	36	64	100	144

Recommend
火龙果软件

例

- 对于不带环的3维网格网络，设每一维方向上有r个节点，求：
  - （1）节点总数 = r^3
  - （2）网络的直径 = $3(r-1)$
  - （3）链路总数 = $3(r-1)r^2$
  - （4）对分带宽 = r^2
  - （5）网络的度 = 6



Recommend
火龙果软件

习题

- 4
- 6
- 7

Recommend
火龙果软件

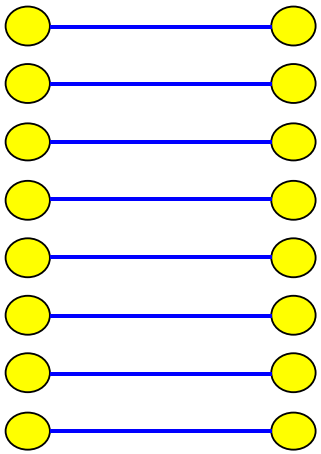
动态互连网络

- **一、网络的互连函数**
  - 互连函数
    - 端口地址的一个一到一的(bijection)映射
  - 表示方法

1. 函数表示法
   - 用$f(x)$表示互连函数
2. 表格表示法

$$\begin{bmatrix} 0 & 1 & \dots & N-1 \\ f(0) & f(1) & \dots & f(N-1) \end{bmatrix}$$

3. 循环表示法
   - 如(0 1)(2 3)(4 5)(6 7)
4. 图形表示法
   - 用连线表示映射关系



Recommend
火龙果软件

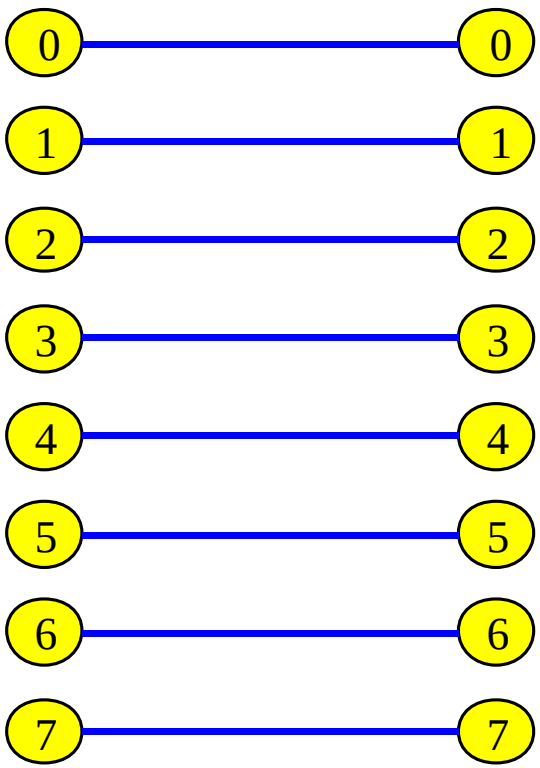
常见的基本互连函数

- **恒等置换**identity permutation

$\mathrm{I}(x) = x$

$$\begin{bmatrix} 0 & 1 & \dots & N-1 \\ 0 & 1 & \dots & N-1 \end{bmatrix}$$

$(0)(1)(2)(3)(4)\dots(N\text{-}1)$



Recommend
火龙果软件

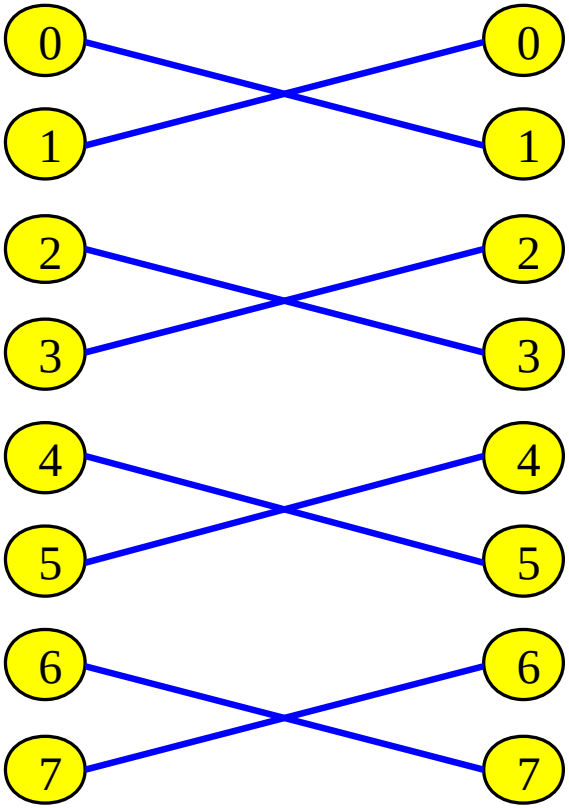
常见的基本互连函数

- **交换置换**exchange permutation
- $E(x_{n-1}x_{n-2}\ldots x_1x_0) = x_{n-1}x_{n-2}\ldots x_1$

例

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 3 & 2 & 5 & 4 & 7 & 6 \end{bmatrix}$$

(0 1)(2 3)(4 5)(6 7)



Recommend
火龙果软件

常见的基本互连函数

- **方体置换**cube permutation

$C_k(x_{n-1}x_{n-2}\ldots x_{k+1}\ x_k\ x_{k-1}\ \ldots x_1x_0) = x_{n-1}x_{n-2}\ldots x_{k+1}\ \ x_{k-1}\ \ldots x_1x_0$

C1:

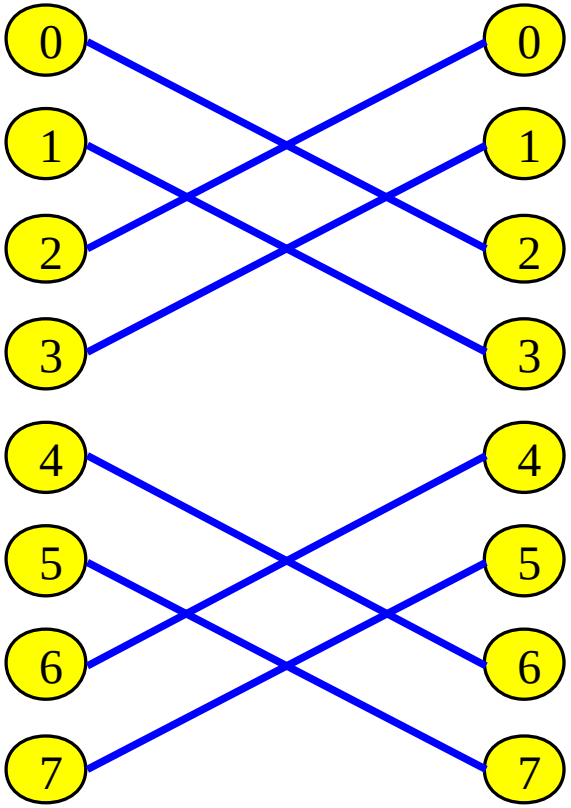
$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 0 & 1 & 6 & 7 & 4 & 5 \end{bmatrix}$$

(0 2)(1 3)(4 6)(5 7)

C2:

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 0 & 1 & 2 & 3 \end{bmatrix}$$

(0 4)(1 5)(2 6)(3 7)



Recommend
火龙果软件

常见的基本互连函数

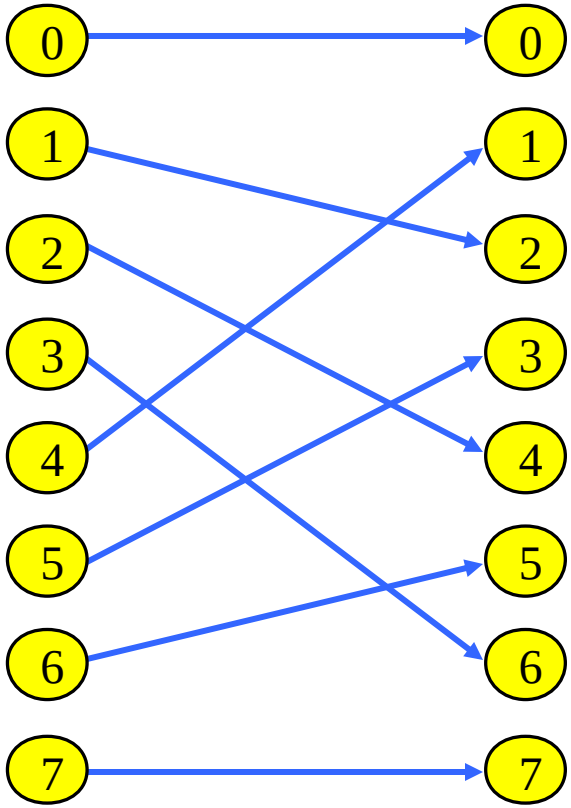
- **均匀洗牌**perfect shuffle permutation

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

例：

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 2 & 4 & 6 & 1 & 3 & 5 & 7 \end{bmatrix}$$

(0)(1 2 4)(3 6 5)(7)



Recommend
火龙果软件

常见的基本互连函数

- **逆洗牌**reverse shuffle

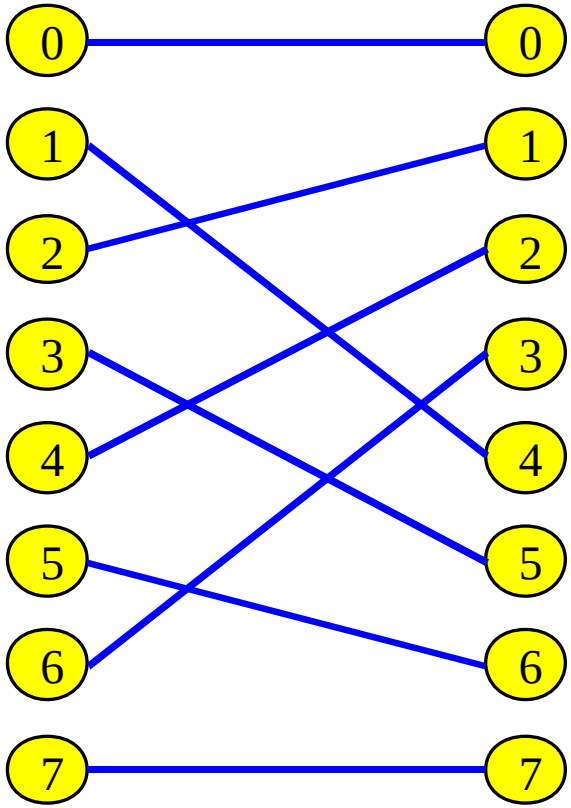
$\sigma^{-1}(x_{n-1}x_{n-2}\ldots x_1x_0)$ =

$x_0x_{n-1}x_{n-2}\ldots x_1$

例

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 4 & 1 & 5 & 2 & 6 & 3 & 7 \end{bmatrix}$$

(0)(1 4 2)(3 5 6)(7)



Recommend
火龙果软件

常见的基本互连函数

- **子洗牌**subshuffle

将端口分成若干个组，对每个组进行均匀洗牌置换。

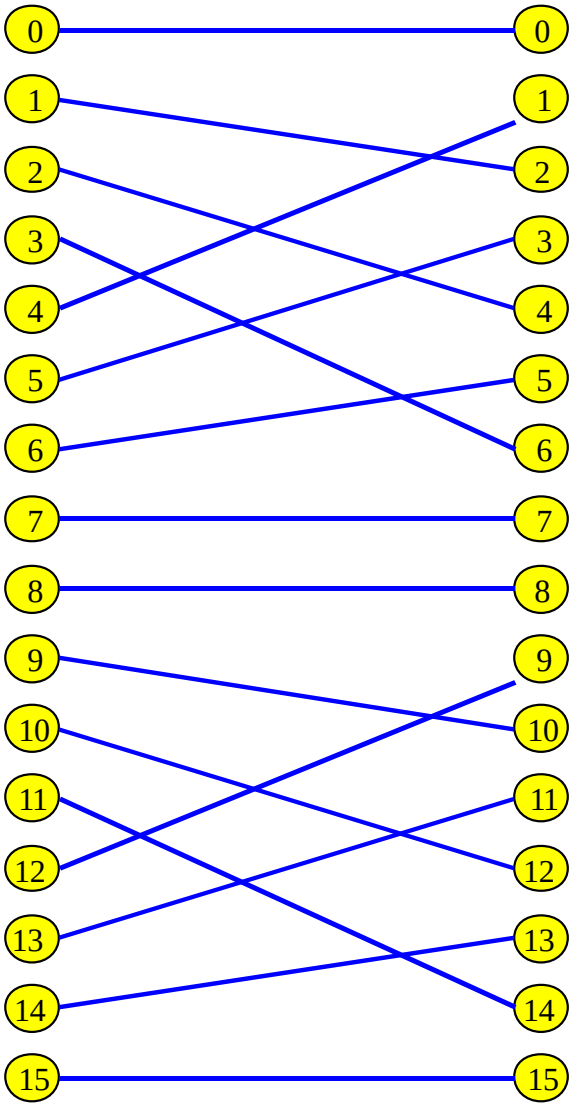
$$\sigma_{(k)}(x_{n-1}x_{n-2}\ldots x_{k+1}x_k x_{k-1}\ldots x_1 x_0) = x_{n-1}x_{n-2}\ldots x_{k+1}x_{k-1}\ldots x_1 x_0 x_k$$

$$0 \le k \le n-1$$

- 例：$\sigma_{(2)}$

(0)(1 2 4)(3 6 5)(7)(8)

(9 10 12)(11 14 13)(15)



Recommend
火龙果软件

常见的基本互连函数

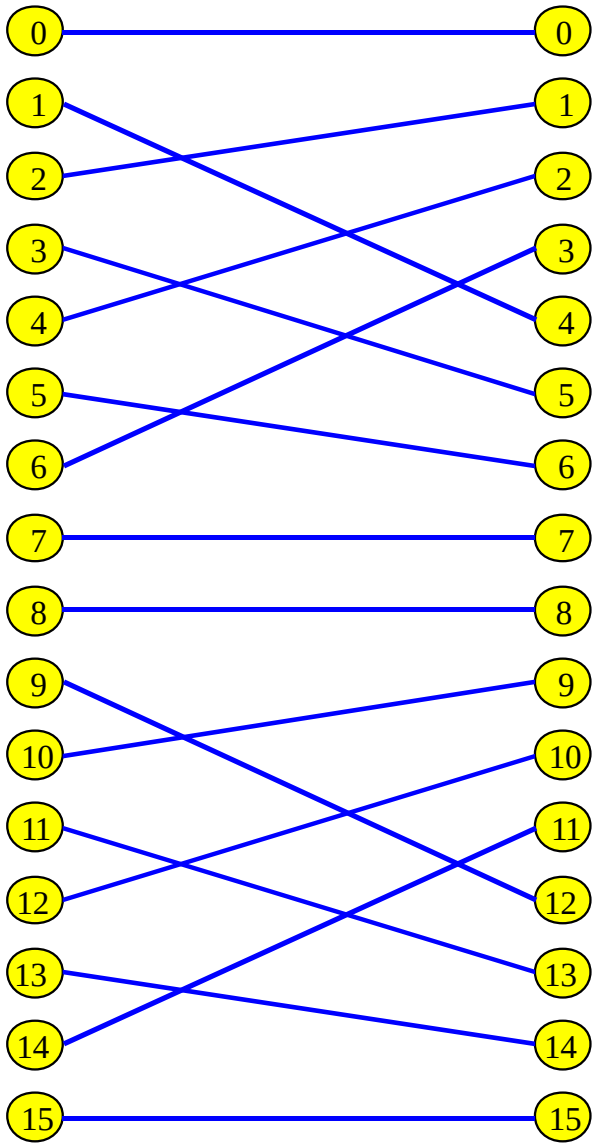
- **子逆洗牌**reverse subshuffle

将端口分成若干个组，对每个组进行逆洗牌置换。

$\sigma^{-1}_{(k)}(x_{n-1}x_{n-2}\ldots x_{k+1}x_kx_{k-1}\ldots x_1x_0) = x_{n-1}x_{n-2}\ldots x_{k+1}x_0x_kx_{k-1}\ldots x_1 \quad 0\le k\le n-1$

- 例：$\sigma^{-1}_{(2)}$

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



Recommend
火龙果软件

常见的基本互连函数

- **超洗牌**supershuffle

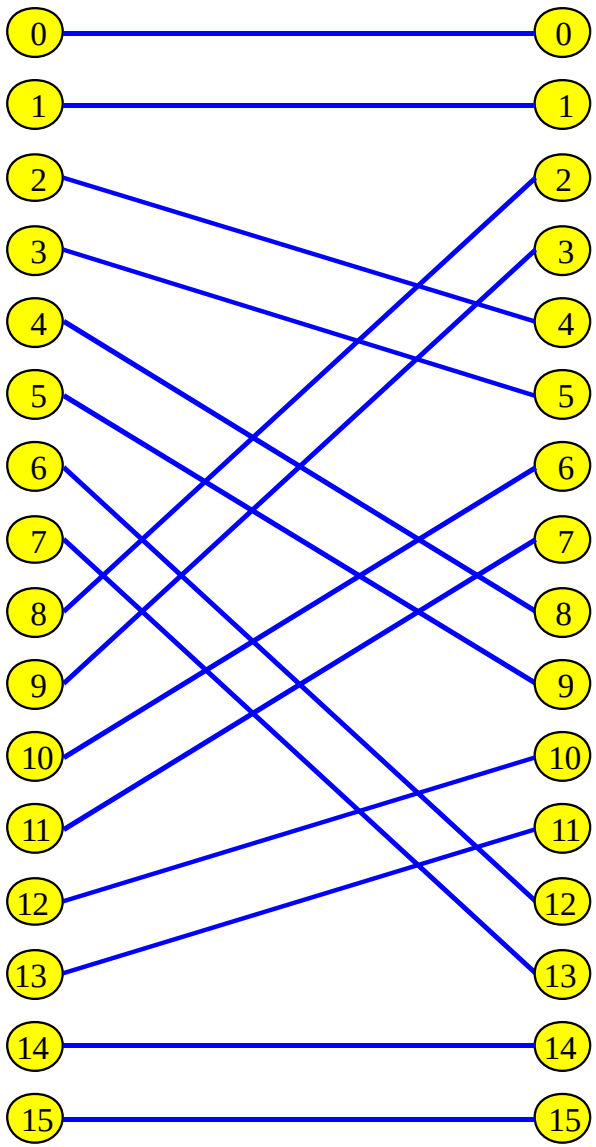
以组为单位洗牌

$$\sigma^{(k)}(x_{n-1}x_{n-2}\ldots x_{n-k}x_{n-k-1}\ldots x_1x_0) = x_{n-2}x_{n-3}\ldots x_{n-k-1}x_{n-1}x_{n-k-2}\ldots x_1x_0$$

$$0 \le k \le n-1$$

例： $\sigma^{(1)}$

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



Recommend
火龙果软件

常见的基本互连函数

- **蝶式置换**butterfly permutation

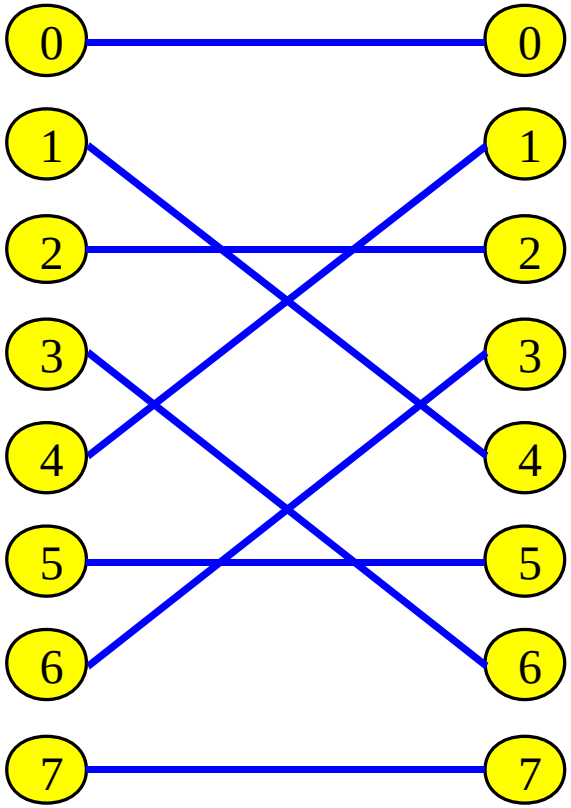
$\beta(x_{n-1}x_{n-2}\ldots x_1x_0)$

$= x_0x_{n-2}\ldots x_1x_{n-1}$

例：

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 4 & 2 & 6 & 1 & 5 & 3 & 7 \end{bmatrix}$$

- (0)(2)(1 4) (3 6)(5)(7)



Recommend
火龙果软件

常见的基本互连函数

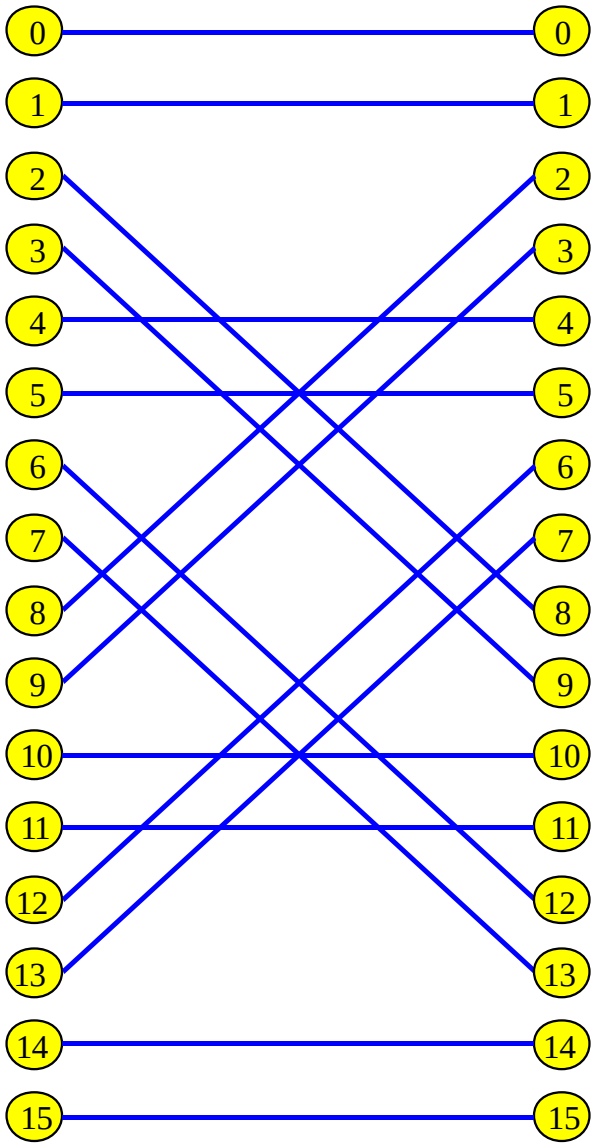
- **超蝶式置换**

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

例

(0)(1)(2 8)(3 9)(4)(5)(6 12)

(7 13)(10)(11)(14)(15)



Recommend
火龙果软件

常见的基本互连函数

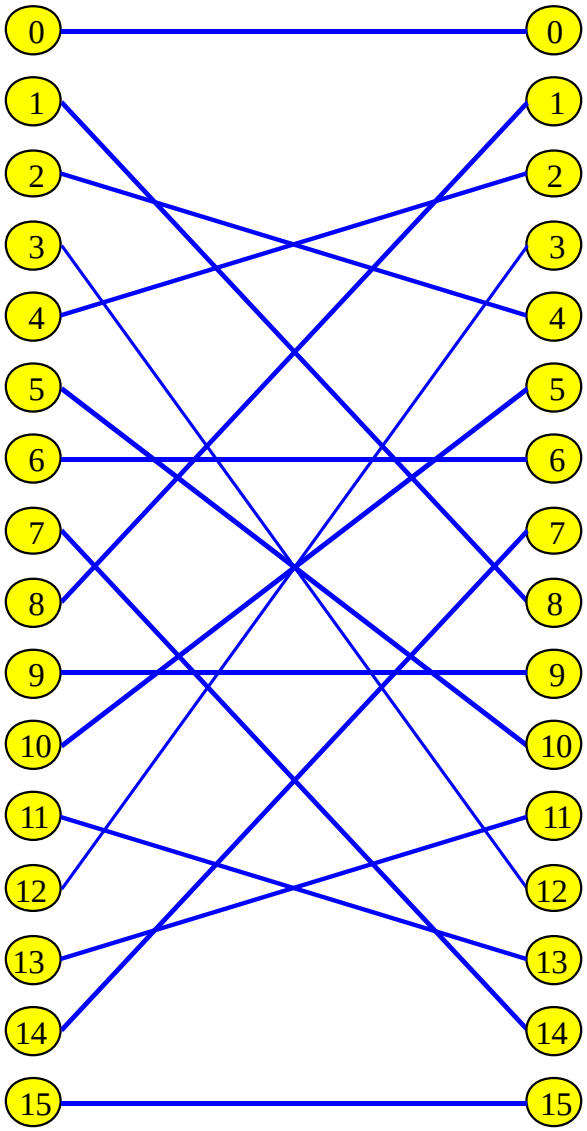
- **位序颠倒置换**bit reversal permutation

$$\rho(x_{n-1}x_{n-2}\ldots x_1x_0) = x_0x_1\ldots x_{n-2}x_{n-1}$$

例

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

- N=8时与蝶式置换相同
- 子位序颠倒置换
  - 低位段颠倒
- 超位序颠倒置换
  - 高位段颠倒



Recommend
火龙果软件

常见的基本互连函数

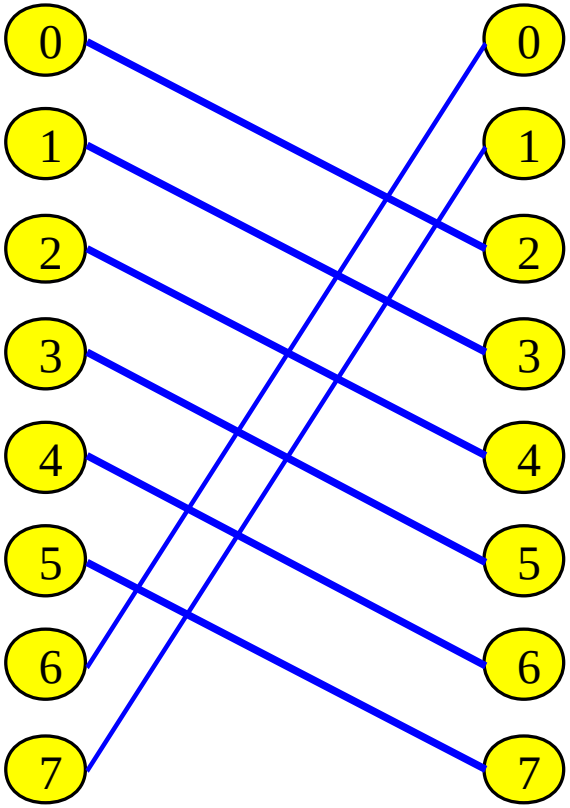
- **移数置换**shift permutation

$\alpha(x) = (x+k) \bmod N,\ 0 \le x \le N$

例：

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 0 & 1 \end{bmatrix}$$

(0 2 4 6)(1 3 5 7)



Recommend
火龙果软件

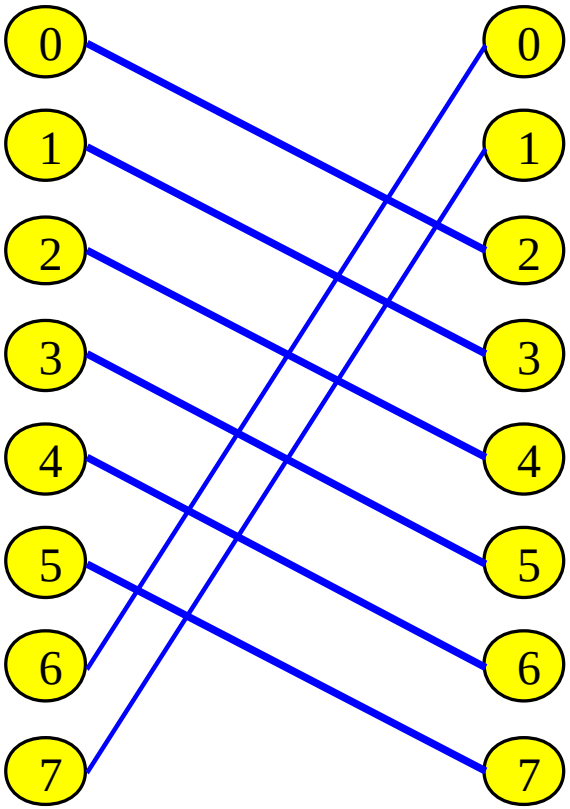
常见的基本互连函数

- 加减2^i置换plus-minus 2^i permutation

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

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

其中$0 \le x \le N-1$，$0 \le i \le \log_2 N-1$。

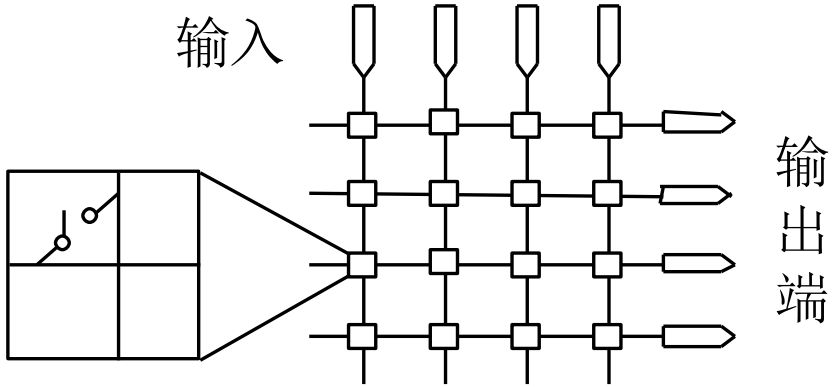


Recommend
火龙果软件

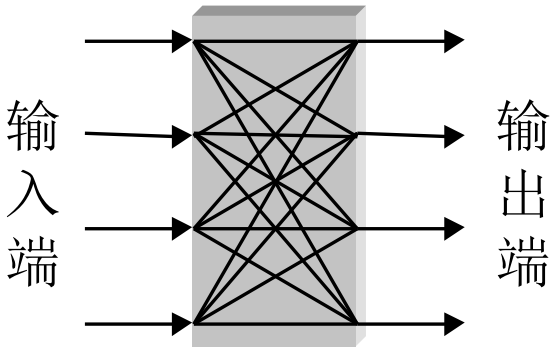
交叉开关crossbar

- 非阻塞
- 扇出：N
- 步数：1

置换数？



(a) 开关阵列模型

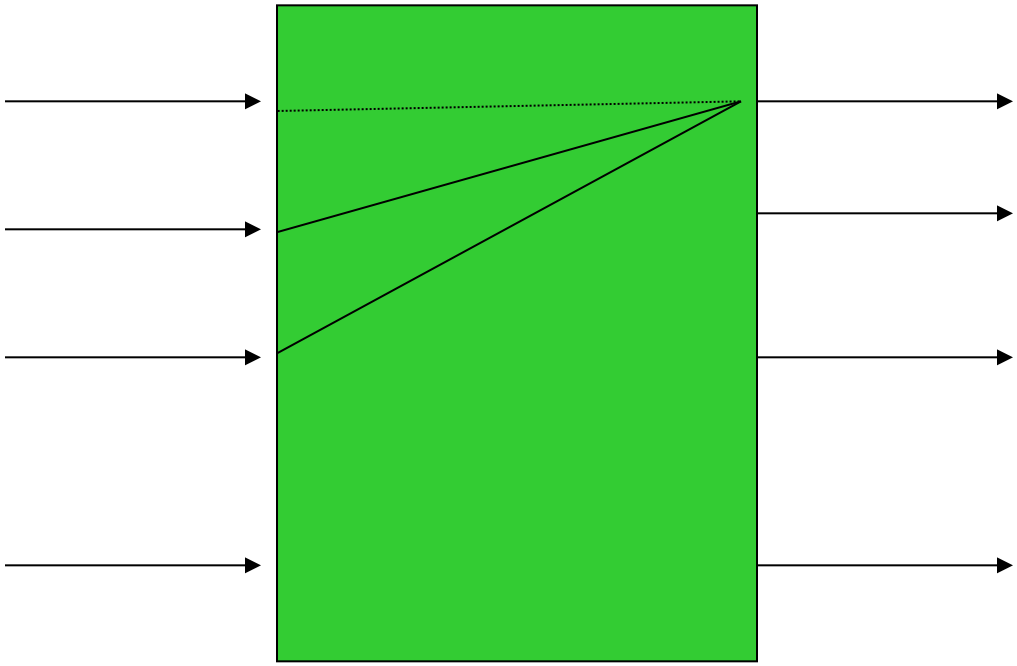


(b) 单级网络模型

Recommend
火龙果软件

交叉开关的端口冲突

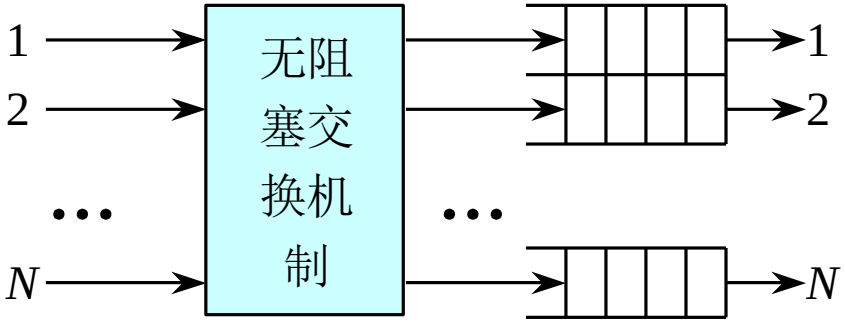
- 有多个输入端的分组需要转发到同一个输出端
  - 建立队列使分组排队等待



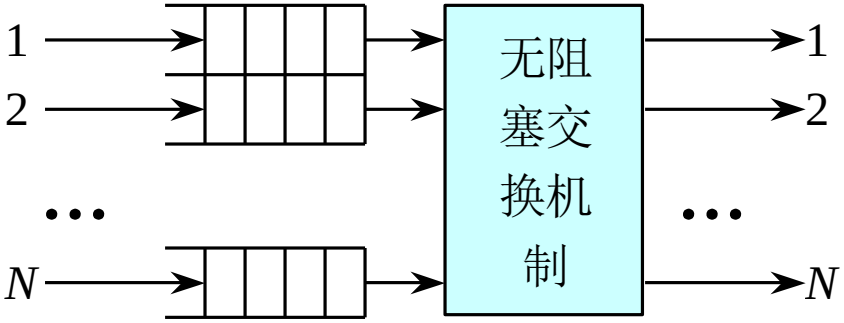
Recommend
火龙果软件

排队方式

- 输入队列(IQ)
- 输出队列(OQ)
- 虚拟输出队列(VOQ)
- 输入与开关队列的组合(CICQ)
- 输入与输出队列的组合(CIOQ)



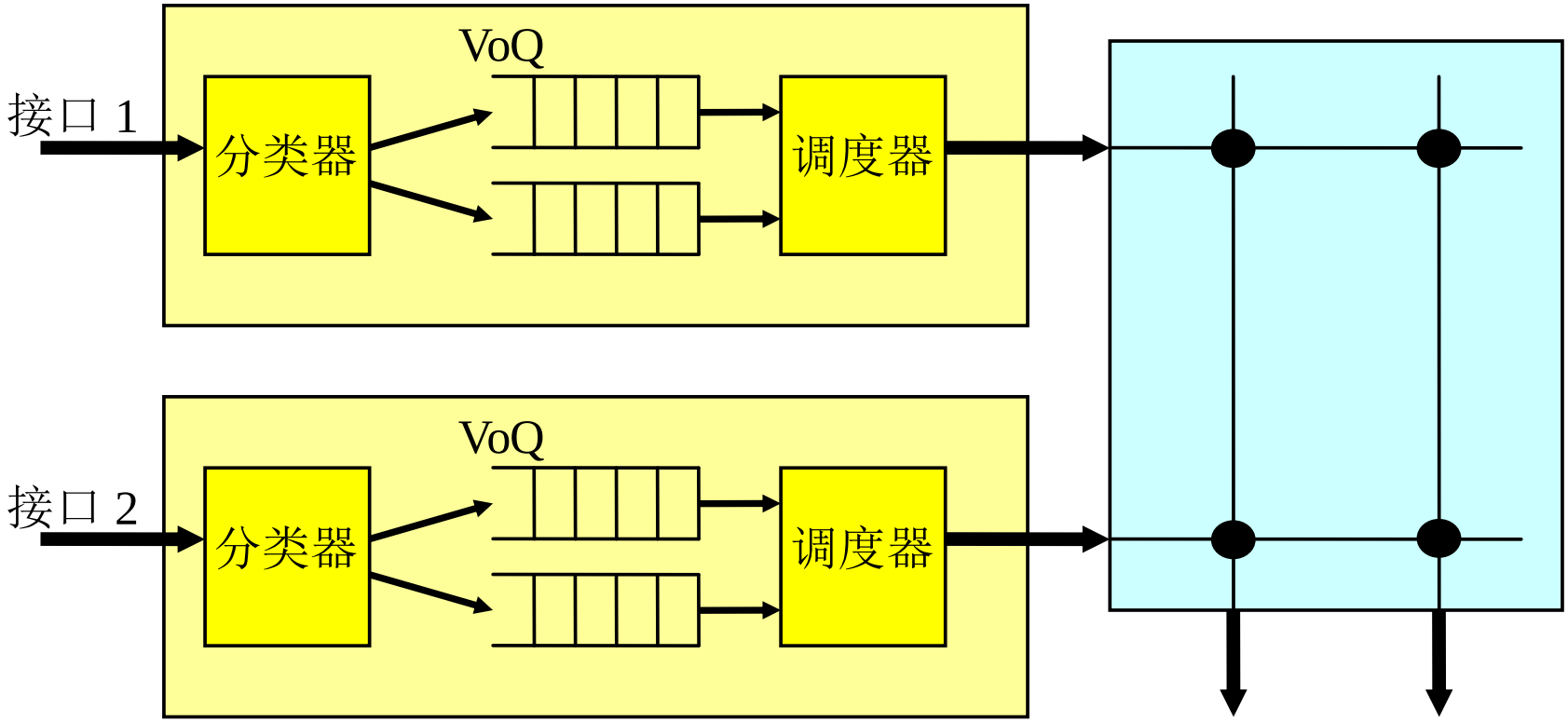
(a) 输出队列



(b) 输入队列

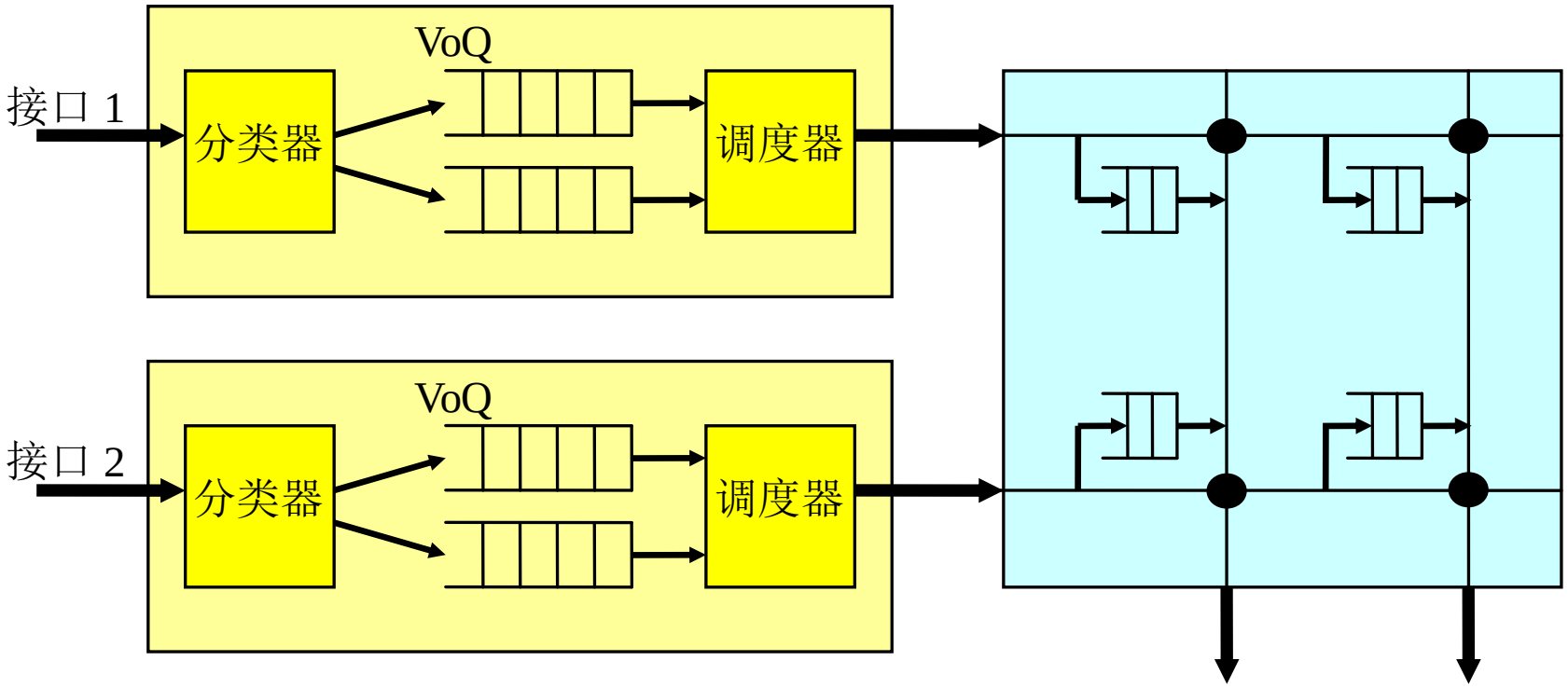
Recommend
火龙果软件

VOQ



Recommend
火龙果软件

输入与开关队列的组合(CICQ)

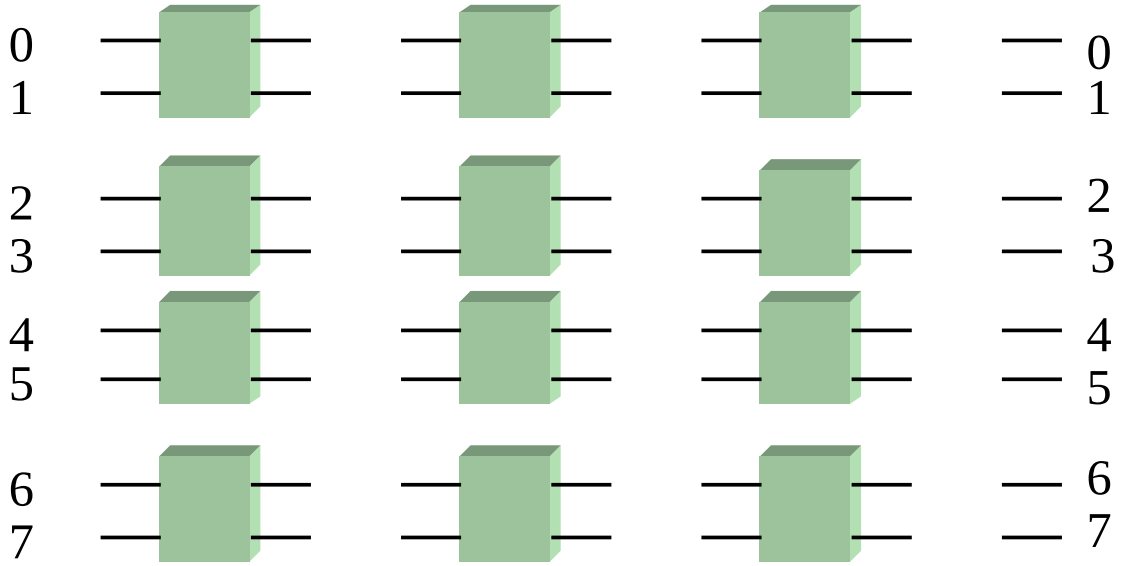


Recommend
火龙果软件

多级动态互连网络

多级动态网参数

- 开关元件
  - 2x2, 4x4, axb
- 连接模式
  - 恒等
  - 洗牌
  - 蝶式
- 控制方式
  - 级控
  - 部分级控
  - 单元控制



Recommend
火龙果软件

STARAN网络

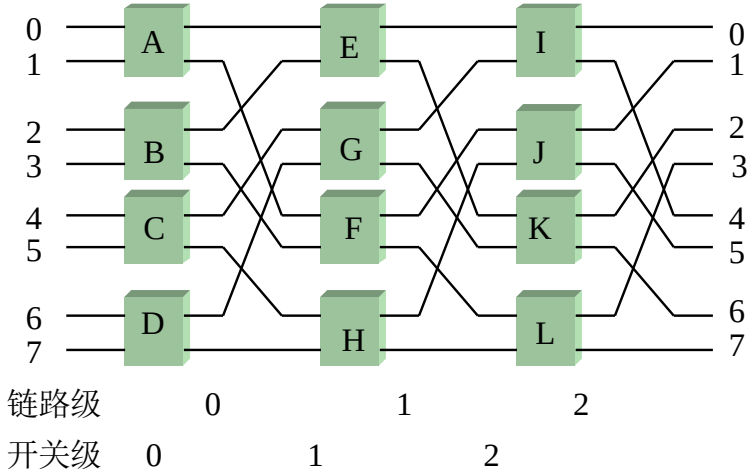
- 开关元件：两功能
- 连接模式：输入恒等，级间和输出逆洗牌
- 控制方式：级控
- 置换数：N
- 开关元件直通时实现恒等置换：

$$\sigma^{-1}\ \sigma^{-1}\ \sigma^{-1}(x_2x_1x_0) = \sigma^{-1}\ \sigma^{-1}(x_0x_2x_1) = \sigma^{-1}(x_1x_0x_2) = x_2x_1x_0$$

输入端与输出端的连通性

011→101

连通规律

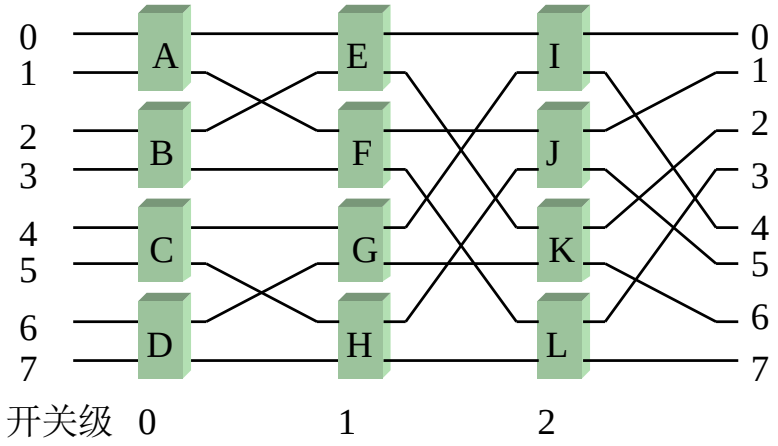


Recommend
火龙果软件

间接二进制n维立方体网络

- 开关元件：两功能
- 连接方式：输入恒等，级间子蝶式，输出逆洗牌
- 控制方式：单元控制
- 置换数：$N^{(N/2)} < N!$
- 开关元件直通时实现恒等置换：

$$\sigma^{-1}\,\beta\,\beta_1(x_2x_1x_0) = \sigma^{-1}\,\beta(x_2x_0x_1) = \sigma^{-1}(x_1x_0x_2) = x_2x_1x_0$$

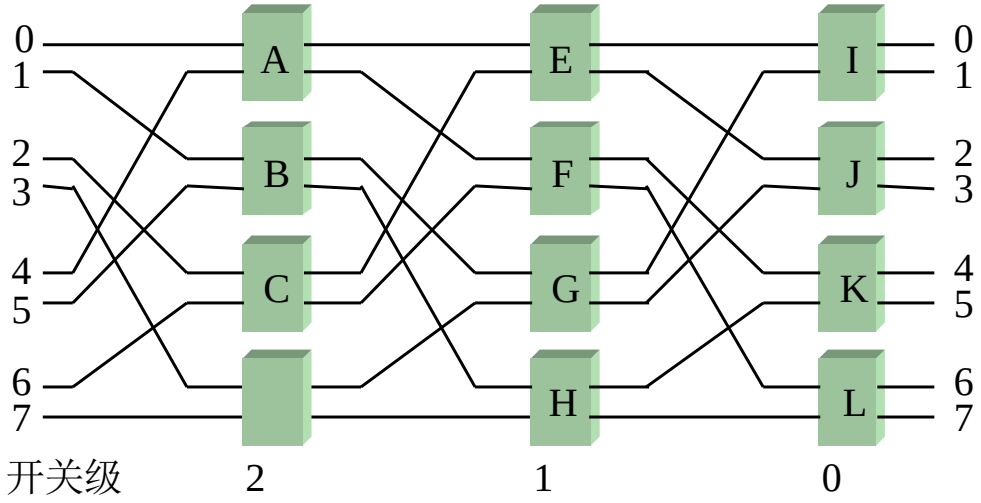


Recommend
火龙果软件

Ω网络

- 开关元件：四功能
- 连接模式：输入与级间洗牌，输出恒等
- 控制方式：单元控制
- 寻径算法：目标地址
- 开关元件直通时实现恒等置换：

$$\sigma\sigma\sigma(x_2x_1x_0) = \sigma\sigma(x_1x_0x_2) = \sigma(x_0x_2x_1) = x_2x_1x_0$$



Recommend
火龙果软件

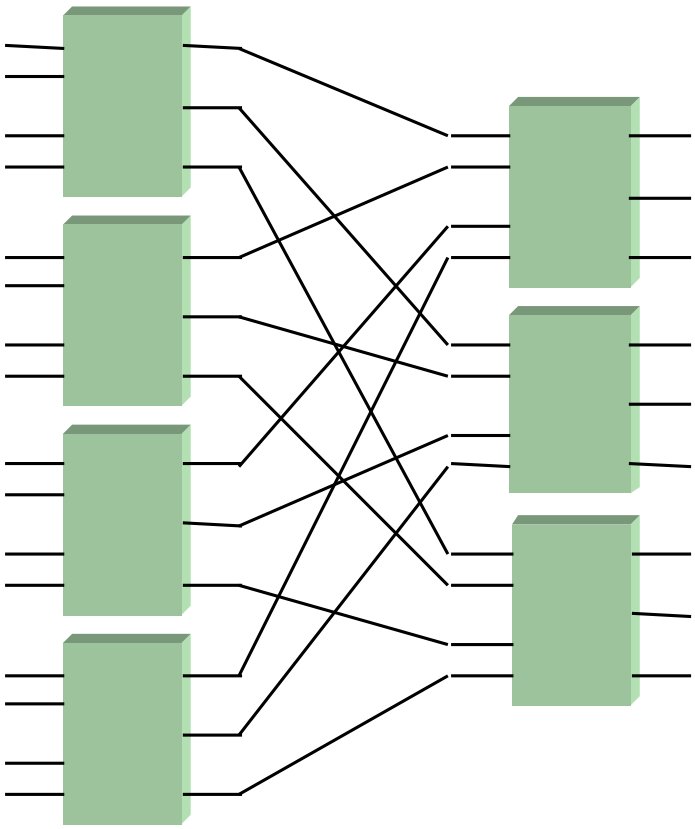
Delta网络

- 开关元件
  - $a \times b$交叉开关
- 连接模式
  - 级间a洗牌
  - 输入与输出恒等
- q洗牌：

$$S_{q \times r}(i) = (q \times i + \lfloor i / r \rfloor) \bmod q \times r$$

$$0 \le i \le q \times r - 1$$

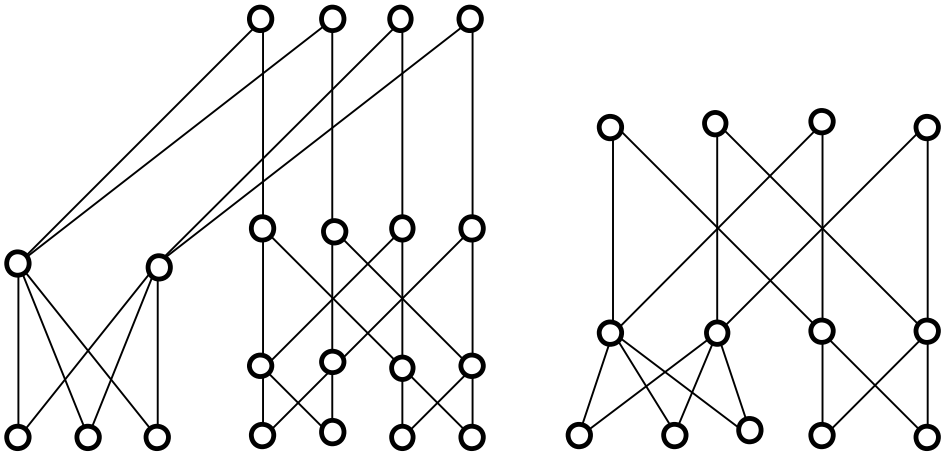
- 控制方式
  - 不限
- 输入输出端口数
  - $a^n \times b^n$



Recommend
火龙果软件

榕树网络

- 底结点base，顶结点apex，中间结点intermediates
- 特点
  - 从一个底结点到一个顶结点一定存在一条而且也只存在一条路径
  - 连接线是从下向上单向的
  - 底节点在下
  - 顶结点在上



Recommend
火龙果软件

榕树网的参数

- 级数Layer
  - 从底结点到顶结点的距离
- 展开度Spread
  - 输出数，结点向上引出的边的数量
- 扇出度Fanout
  - 输入数，结点下面引入的边的数量

榕树分类

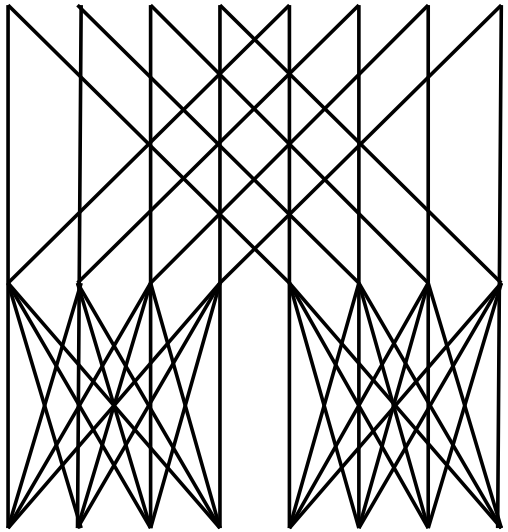
- 均匀榕树uniformed banyan

  每一级中各结点的F相同，S相同
  - 规则：每一级的F,S相同
  - 非规则
- 非均匀榕树non-uniformed banyan
- 矩形榕树rectangular banyan

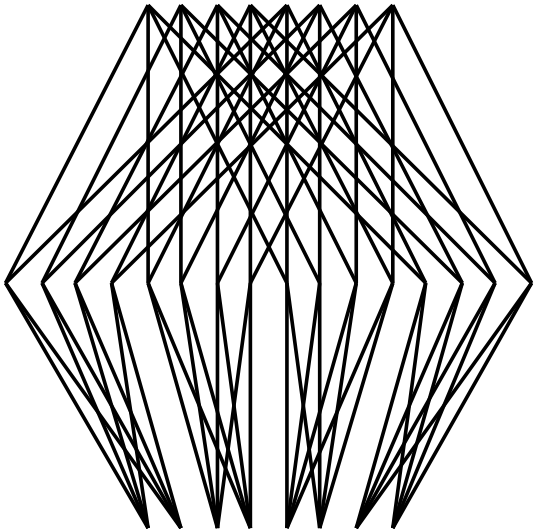
  每级结点数相同
  - 强矩形(F=S)
  - 弱矩形
- 扩展收缩形榕树
- SW榕树：递归生成

Recommend
火龙果软件

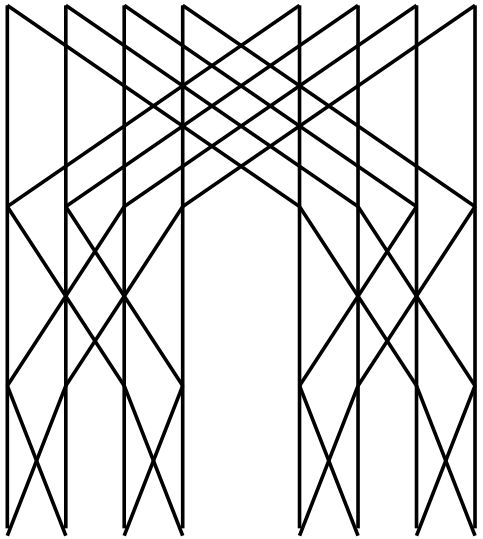
榕树网络的例子



(a) 弱矩形(非规则型)



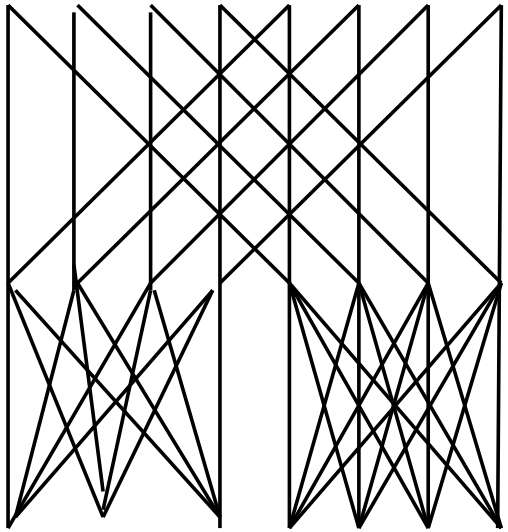
(b) 扩展收缩形



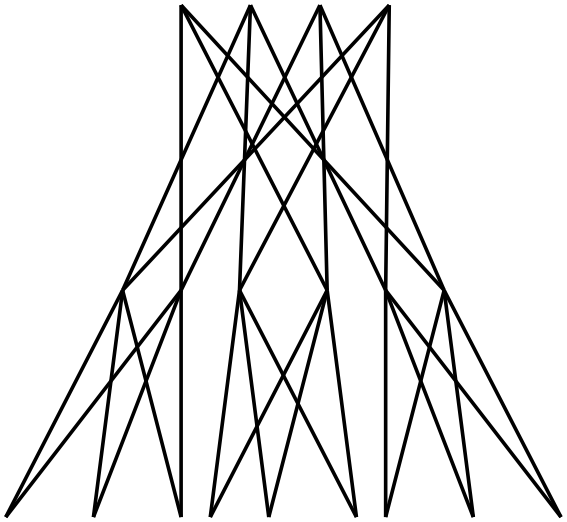
(c) 强矩形(规则型)

Recommend
火龙果软件

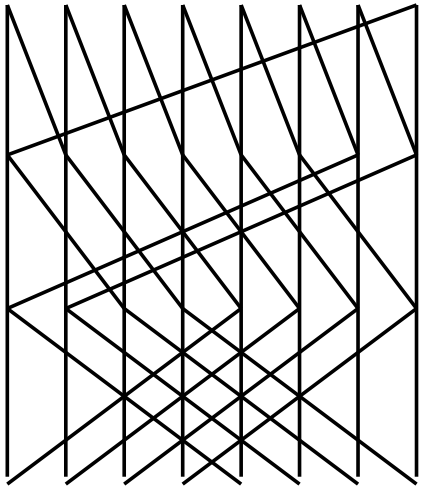
榕树网络的例子



(d) 非均匀型



(e) 规则型

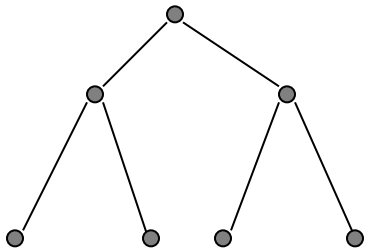


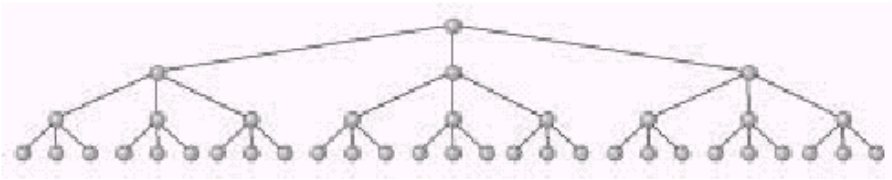
(f) 强矩形(规则型)

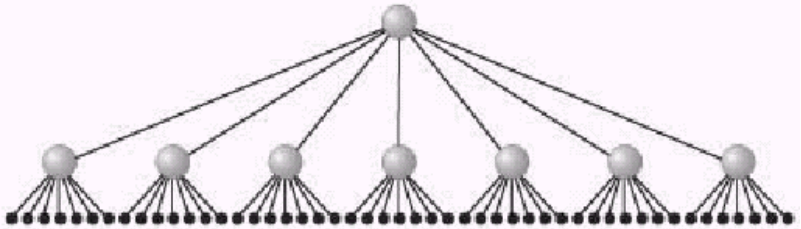
Recommend
火龙果软件

树型网络

- 特点：双向，单边
- 分类：二叉，三叉，四叉等
- 二叉数
  - 度d=3
  - 直径=$2\log_2 N$
  - 寻径简单、伸缩性好
  - 无冗余通路，容错能力差（对分带宽=1）

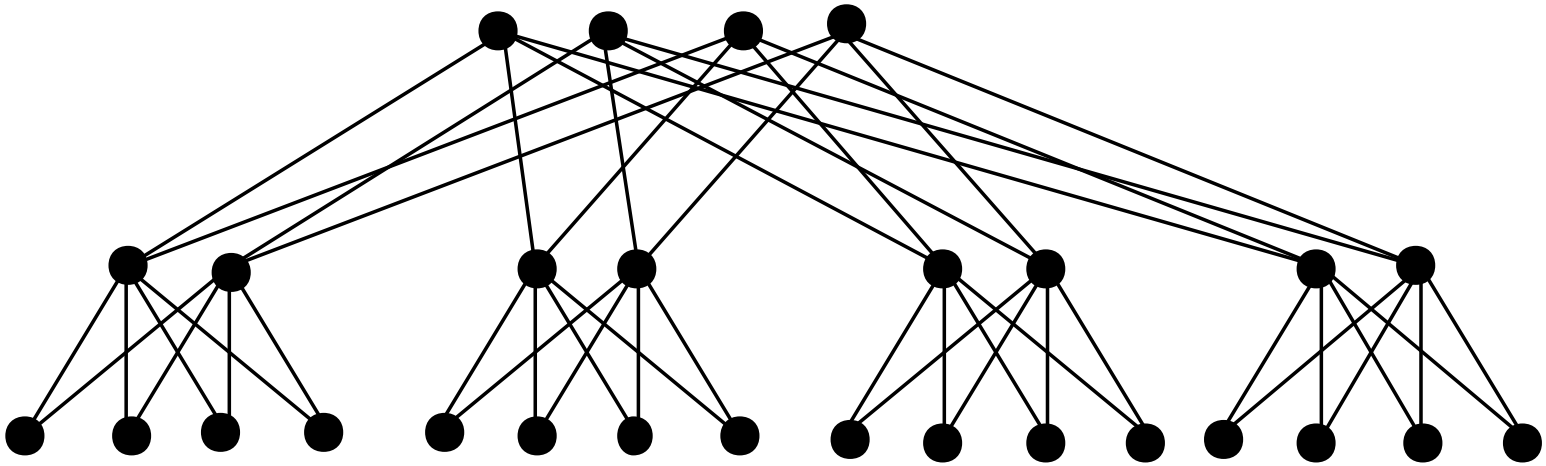






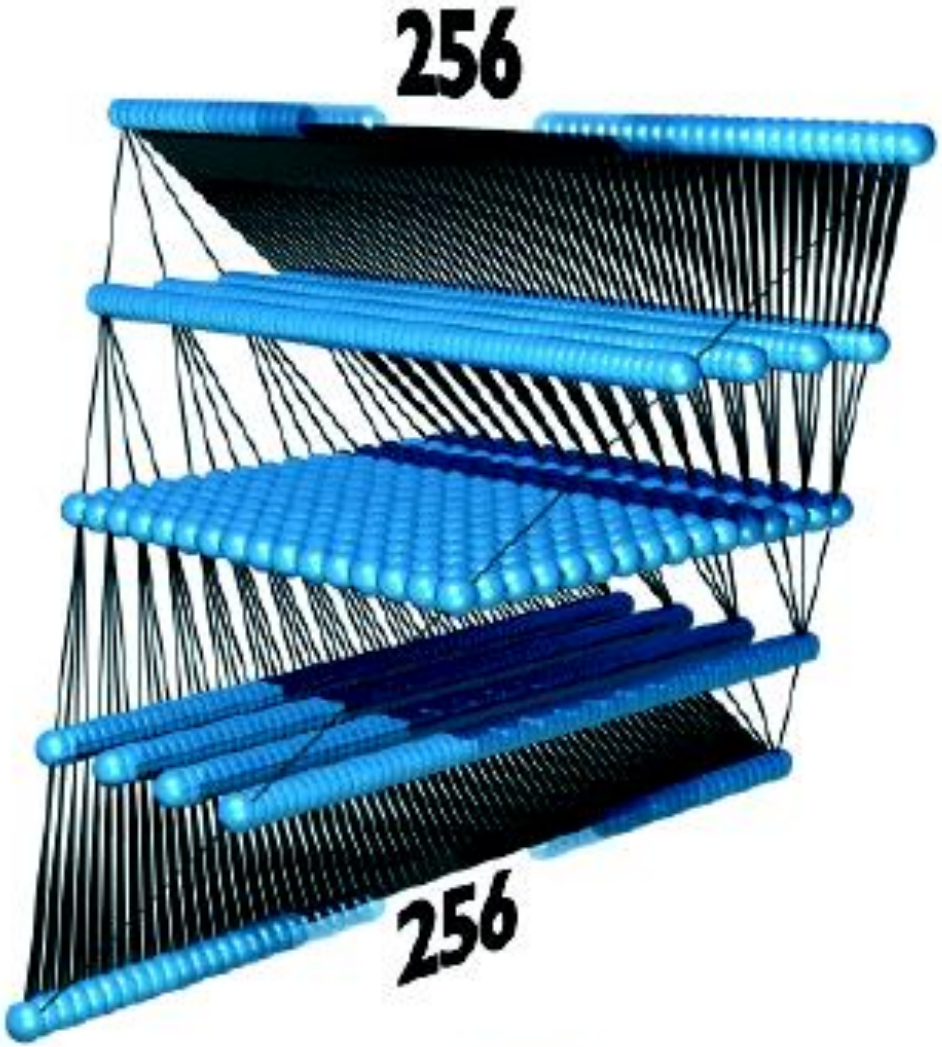
Recommend
火龙果软件

超树hypertree



Recommend
火龙果软件

超树hypertree



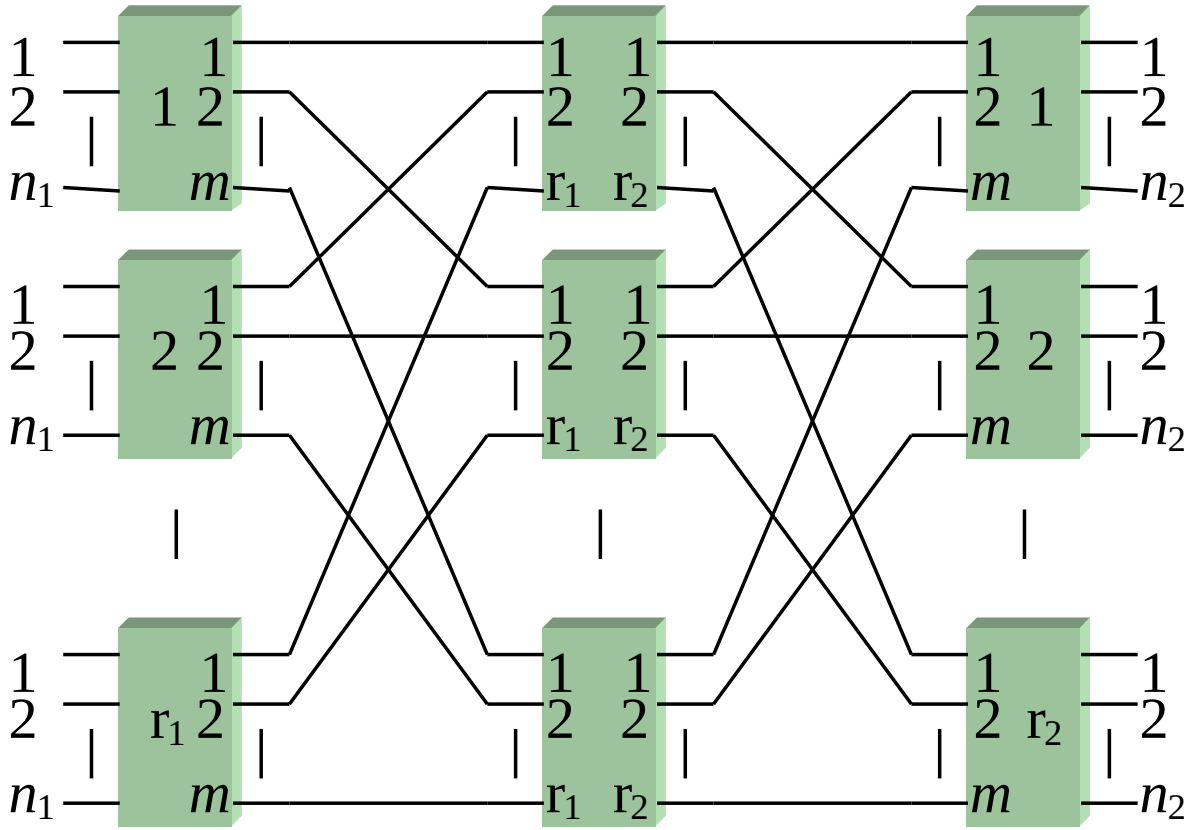
Recommend
火龙果软件

Clos网络

- 开关元件
  - 输入级r_1个$n_1 \times m$交叉开关
  - 中间级m个$r_1 \times r_2$交叉开关
  - 输出级r_2个$m \times n_2$交叉开关
- 连接方式
  - 输入与输出恒等
  - 级间r_1洗牌和m洗牌

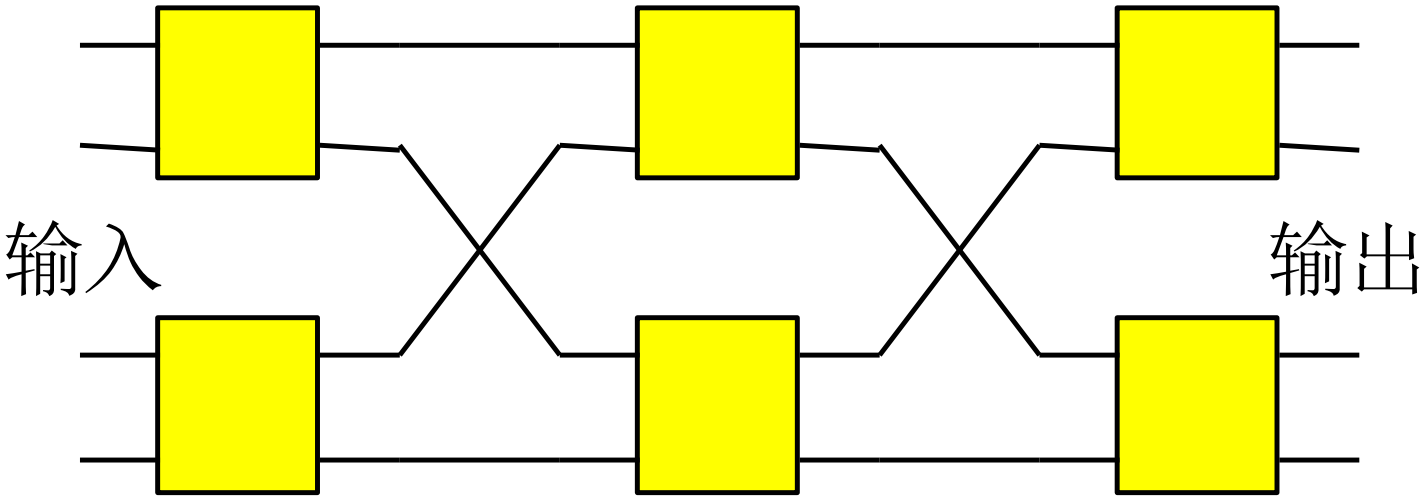
Recommend
火龙果软件

Clos网络



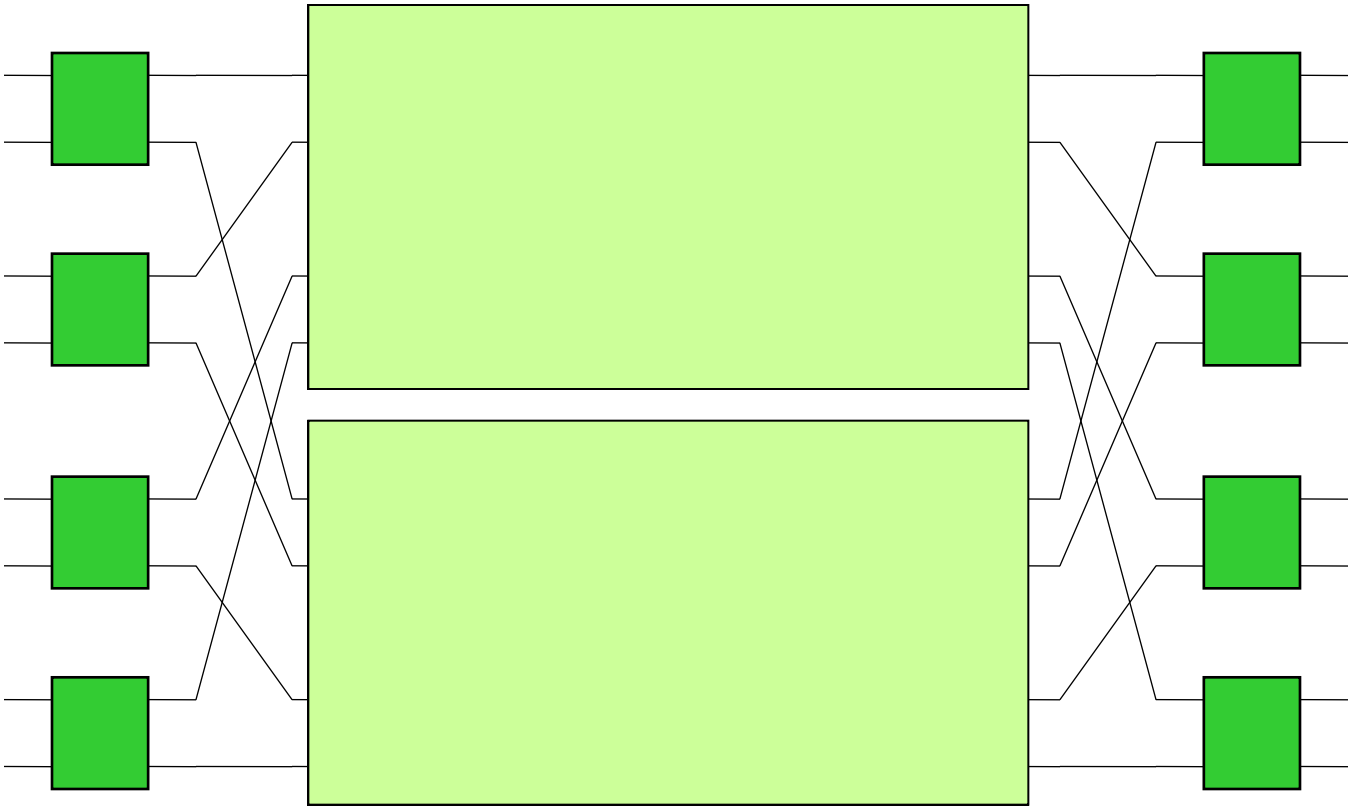
Recommend
火龙果软件

Bennus网络



Recommend
火龙果软件

Bennus网络



Recommend
火龙果软件

习题

- 9
- 10
- 12
- 13
- 14
- 15

Recommend
火龙果软件

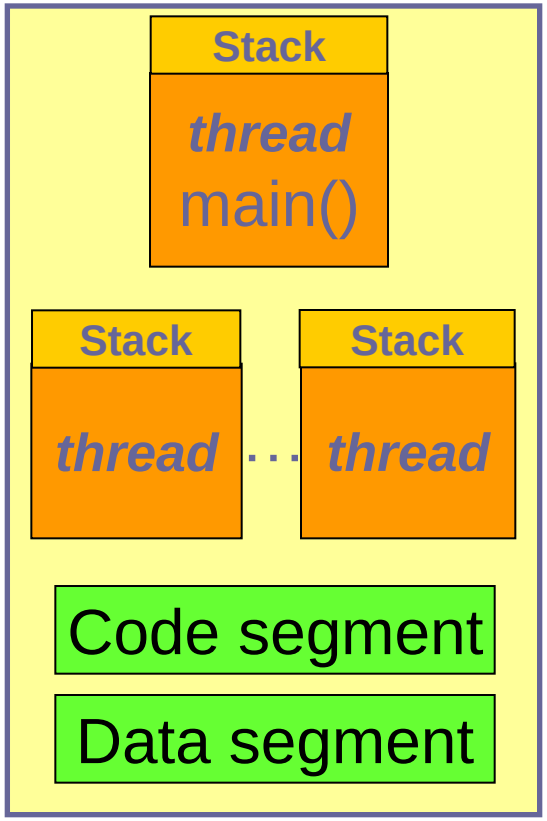
5.3 芯片级并行性开发

- 多线程并行性
- 多核并行性
- 芯片级互连网络
- 并行处理器芯片实例

Recommend
火龙果软件

多线程并行性

- 进程
  - 资源占有的单位
  - 运算的单位
- 线程
  - 进程中的一个对象
  - 状态
    - 包含指令计数器PC的值
    - 用户态的堆栈和系统态的堆栈
    - 寄存器
    - 存储区域



Recommend
火龙果软件

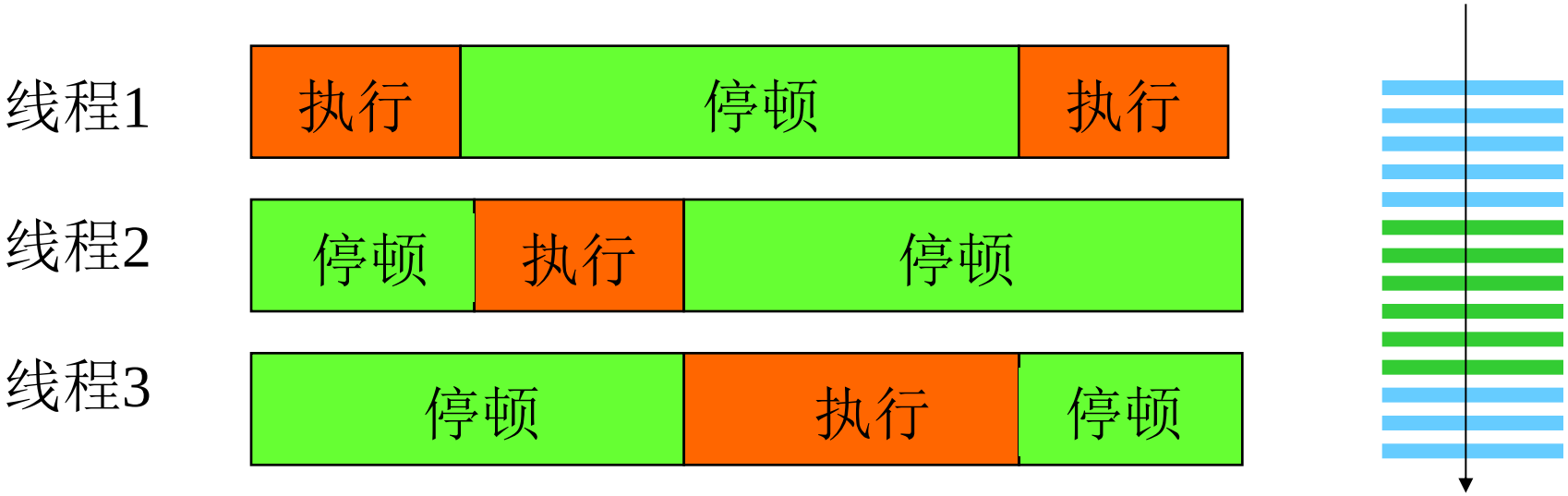
多线程处理器分类

- 串行多线程系统结构
  - 处理器在多个线程之间来回切换
    - 时间片多线程(time-slice multithreading)
    - 事件切换多线程(switch-on-event multithreading)
  - 难以充分利用处理器资源
- 并行多线程系统结构PMA
  - parallel multithreaded architecture
  - 多个线程的指令同时执行，构成多个逻辑处理器

Recommend
火龙果软件

串行多线程系统结构

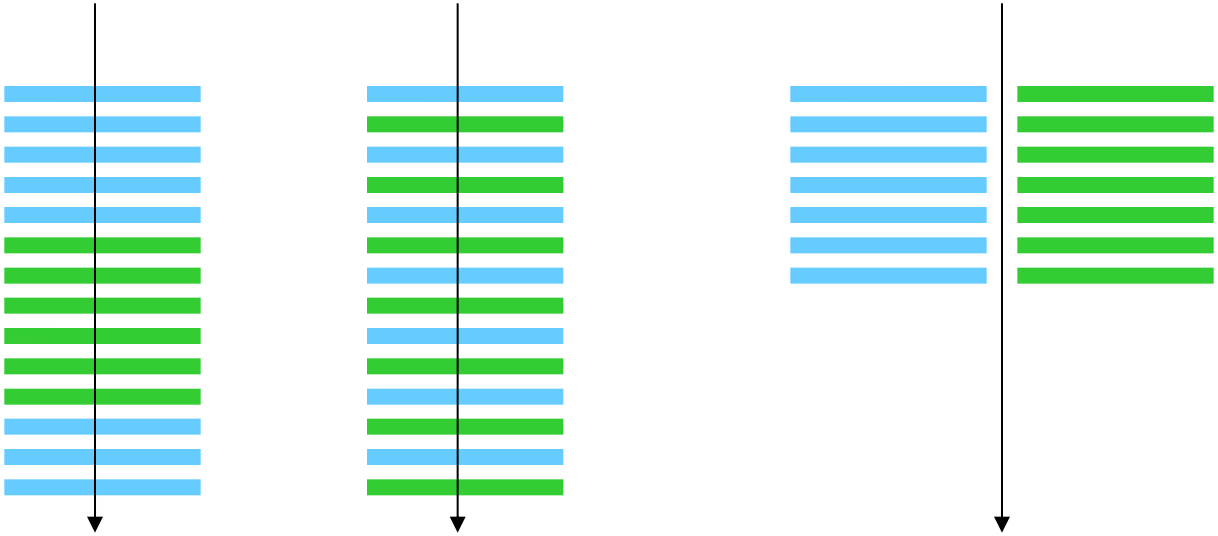
- 线程轮流执行
- 线程停顿时进行切换
  - 访存、同步操作、I/O



Recommend
火龙果软件

PMA结构

- 同时启动来自不同线程的指令
- 同时从多个线程中取出多条指令
- 同时进行指令译码和调度
- 同时在执行部件中执行



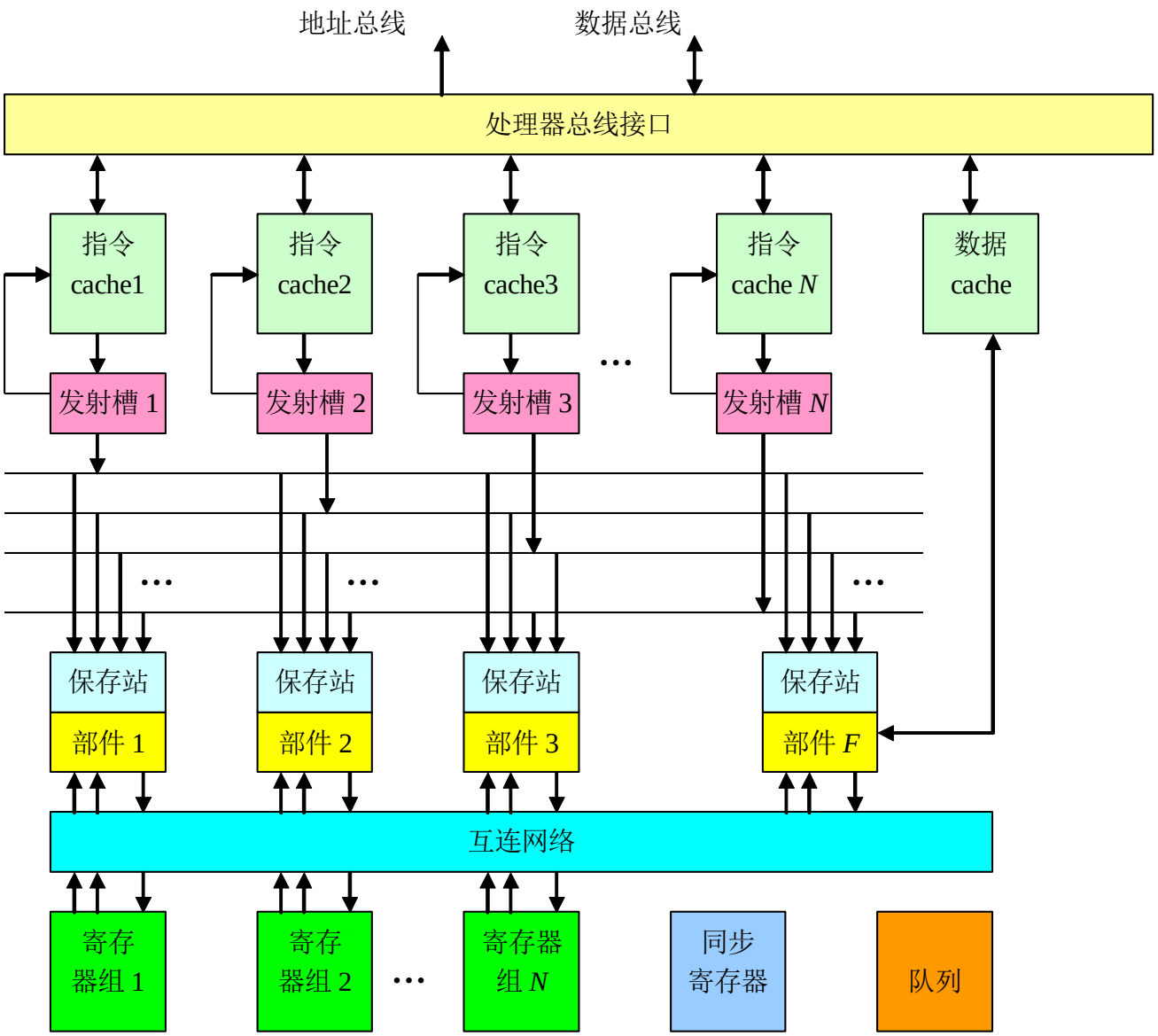
Recommend
火龙果软件

PMA的特点

- 充分利用处理器的计算资源
- 加快进程和线程切换的速度
  - PMA不用中断方式切换进程和线程
- 访存局部性受到影响
  - 采用分离的数据cache
- 指令预取准确性下降
  - 采用分离的指令cache和预取

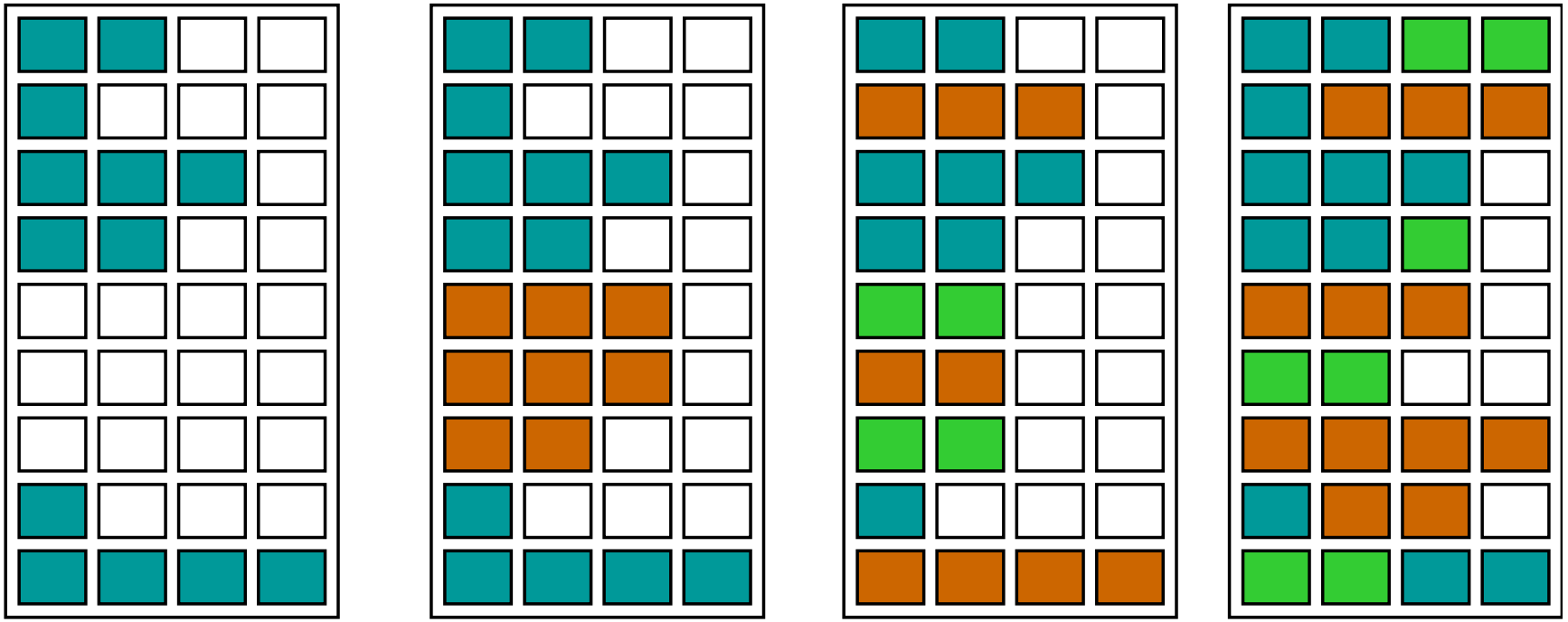
Recommend
火龙果软件

PMA结构


地址总线
数据总线
处理器总线接口
指令 cache1
指令 cache2
指令 cache3
指令 cache N
数据 cache
发射槽 1
发射槽 2
发射槽 3
发射槽 N
保存站
部件 1
保存站
部件 2
保存站
部件 3
保存站
部件 F
互连网络
寄存器组 1
寄存器组 2
寄存器组 N
同步寄存器
队列

单线程与多线程的时空图

- 指令流水中Ex阶段的时空描述



(a) 单线程　(b) 粗颗粒多线程　(c) 细颗粒多线程　(d) 同时多线程

Recommend
火龙果软件

例5-4

- 在一个有序执行的4路超标量的计算机中运行两个线程，其指令序列分别如下：

线程A：		线程B：	
1	sub r1, r0, r5	1	mult r4, r6, r0
2	add r1, r1, r0	2	add r6, r3, r0
3	mult r2, r5, r1	3	mult r3, r3, r1
4	mult r4, r6, r0	4	sub r7, r1, r2
5	add r6, r4, r0	5	div r4, r1, r2
6	add r3, r3, r1	6	mult r1, r0, r5
7	mult r7,r1,r2	7	add r1, r1, r0
8	div r4, r1, r3	8	mult r2, r5, r2
9	sub r7, r4, r7	9	sub r7, r4, r7
10	add, r8, r5, r7	10	add, r8, r5, r7

- 假设每条指令的执行阶段都只需要1个时钟周期。指令流水线中有2个加减法功能部件和2个乘除法功能部件，有相关专用通路。
- （1）试分别分析两个线程的数据相关性，画出两个线程分别运行时的功能部件时空图。
- （2）如果采用SMT执行方式，画出两个线程同时运行时的功能部件时空图。

Recommend
火龙果软件

解

- （1）相关性

线程A：

1	sub r1, r0, r5
2	add r1, r1, r0
3	mult r2, r5, r1
4	mult r4, r6, r0
5	add r6, r4, r0
6	add r3, r3, r1
7	mult r7,r1,r2
8	div r4, r1, r3
9	sub r7, r4, r7
10	add, r8, r5, r7

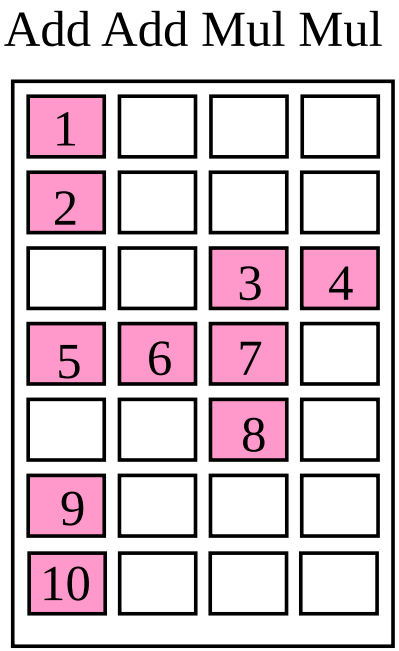
线程B：

1	mult r4, r6, r0
2	add r6, r3, r0
3	mult r3, r3, r1
4	sub r7, r1, r2
5	div r4, r1, r2
6	mult r1, r0, r5
7	add r1, r1, r0
8	mult r2, r5, r2
9	sub r7, r4, r7
10	add, r8, r5, r7

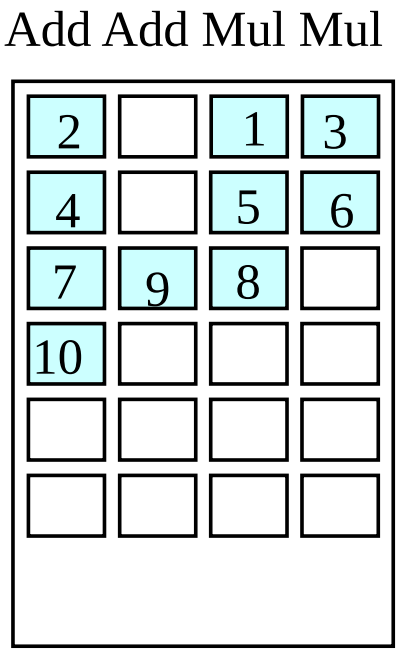
Recommend
火龙果软件

解

- 分别运行时空图



(a) 线程 A



(b) 线程 B

Recommend
火龙果软件

解

- （2）多线程运行部件时空图

Add Add Mul Mul

1	2	1	3
2	4	5	6
7		3	4
5	6	7	8
9		8	
9	10		
10			

Recommend
火龙果软件

习题

- 17

Recommend
火龙果软件

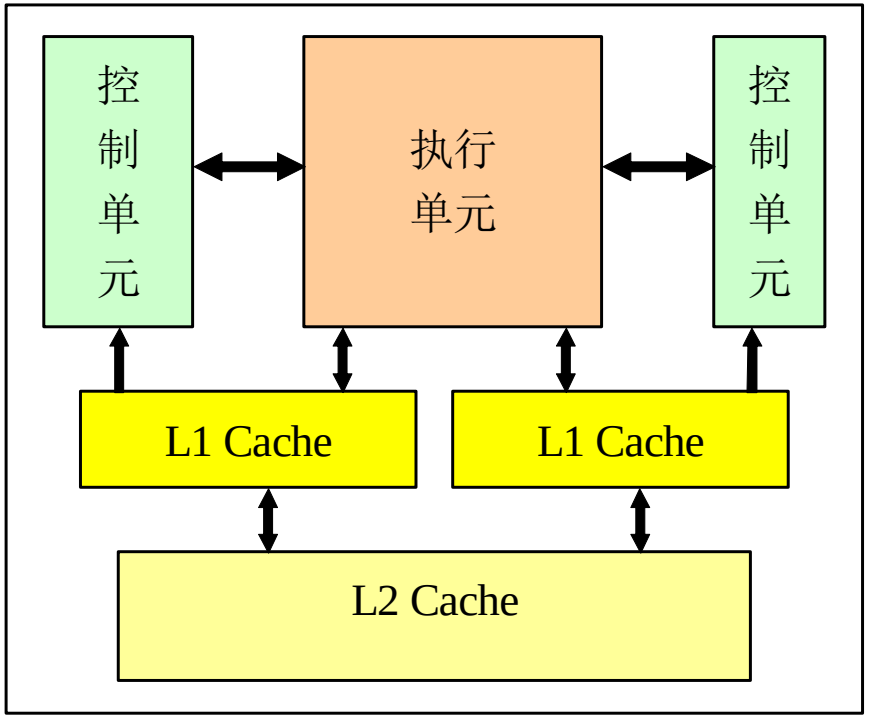
SMT处理器核中需要解决的问题

- 指令启动优先级问题
  - 有大量可供选择的指令启动时选择优先启动哪些线程的指令
  - 保证重要的线程的运行
    - 以及具有实时性要求的线程的运行
- cache冲突问题
  - 多个线程同时运行时争用共享的cache空间

Recommend
火龙果软件

多线程处理器的逻辑模型

- 多个控制器和一个共享的运算器
- 每个控制器控制一个线程的运行
- 多个线程共享使用一个运算器中的多个功能部件



Recommend
火龙果软件

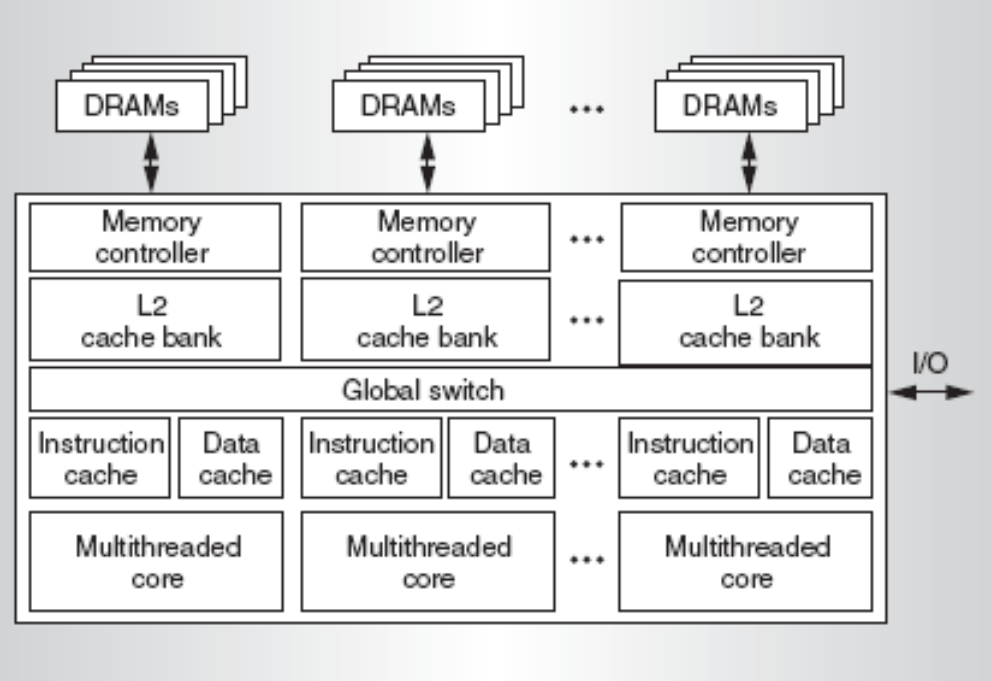
多核并行性开发

- 核
  - 独立的指令执行和控制单元
    - 独立的功能部件
    - 独立的控制器
  - 完整的指令流水线
  - 独立的或者共享的cache
    - 通常L1 cache独立
  - 单线程或SMT

Recommend
火龙果软件

多核处理器分类

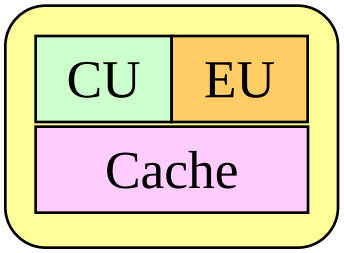
- 单核多线程处理器
  - 单核CPU构成
- 多核处理器
  - 多核芯片构成
- 多核多线程处理器
  - 每个核都是多线程的



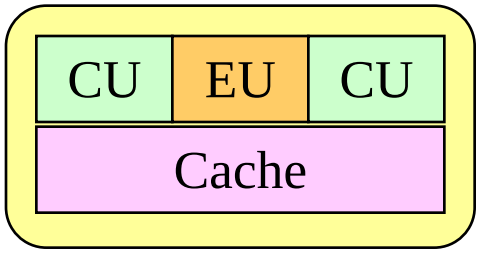
Recommend
火龙果软件

多核处理器的分类

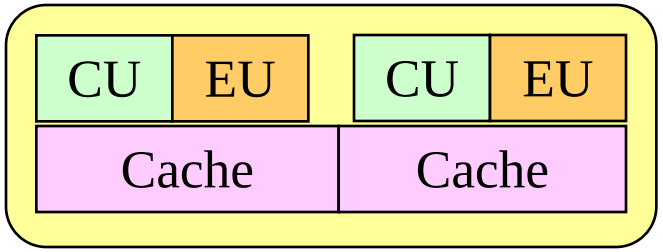
- CU: CPU state + Interrupt logic



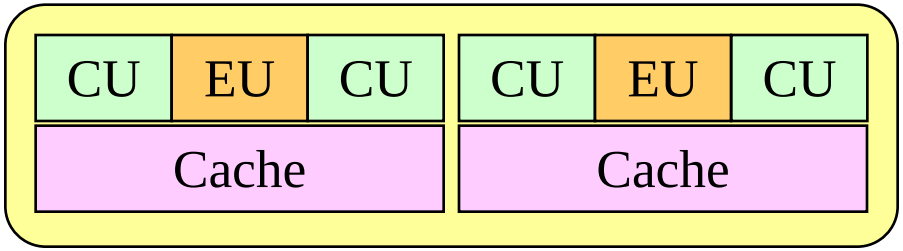
单核单线程处理器



单核多线程处理器



多核处理器

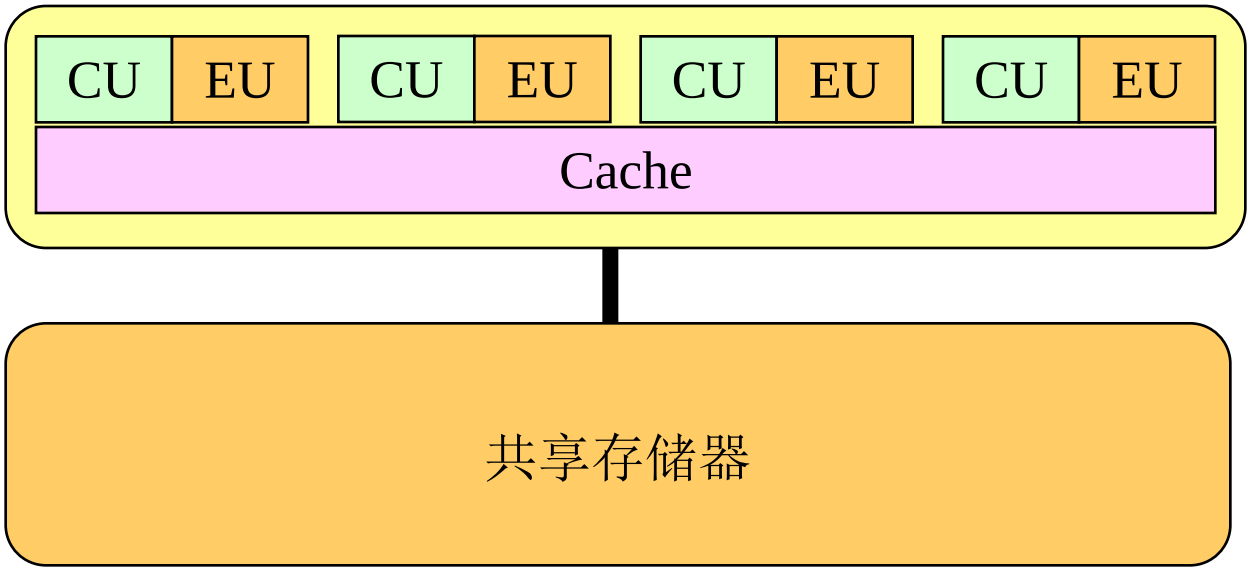


多核多线程处理器

Recommend
火龙果软件

多核处理机

- 单芯片多处理器CMP
  - 均匀cache访问UCA
  - 非均匀cache访问NUCA



Recommend
火龙果软件

多核处理机

- 均匀cache访问UCA
  - 共享L2 cache
  - 有利于提高总体cache命中率
  - cache成为处理器核之间的通信渠道
- 非均匀cache访问NUCA
  - 分布式L2 cache
  - 有利于提高二级cache的访问速度
  - 伸缩性较好

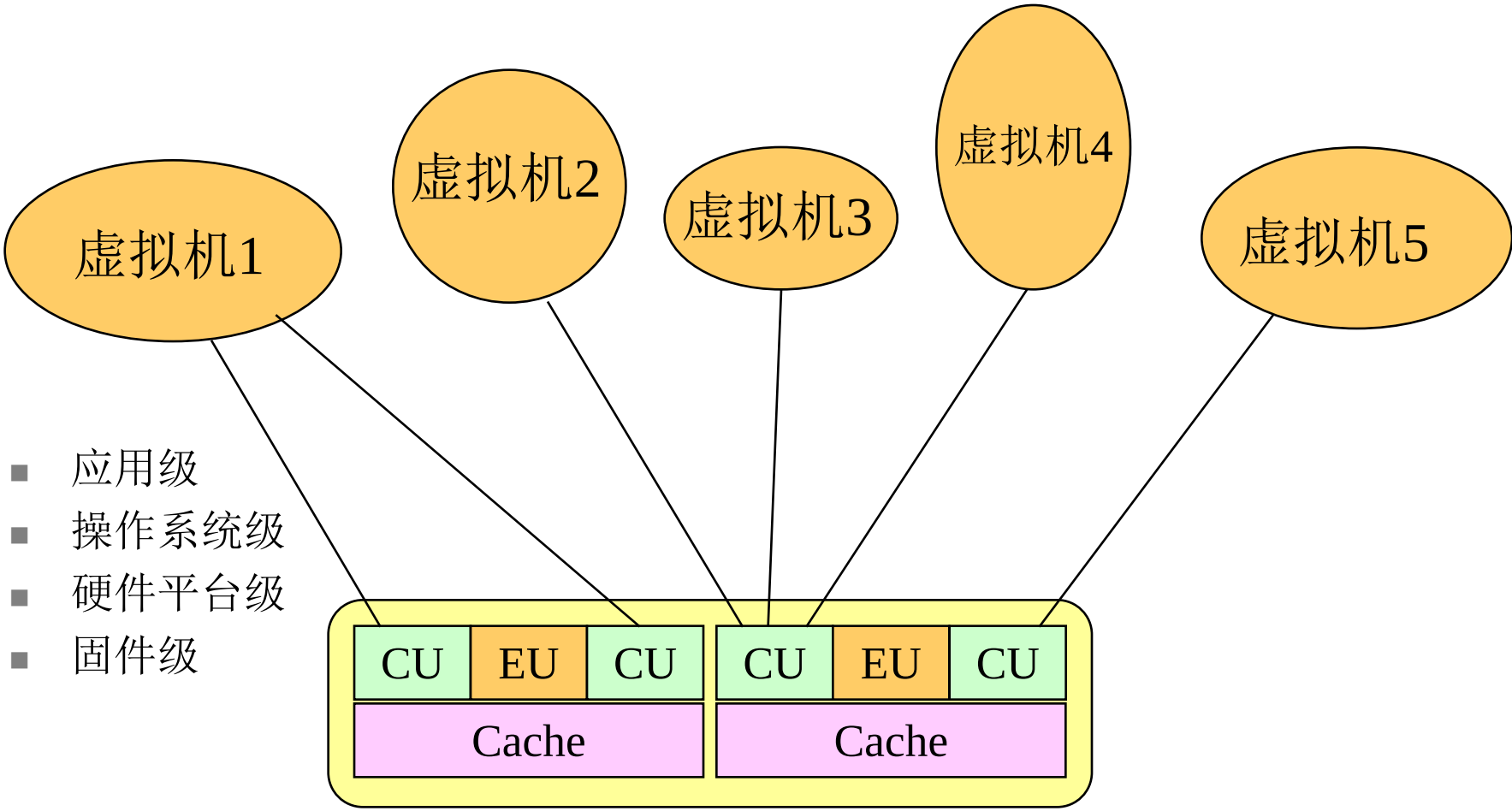
Recommend
火龙果软件

多核并行性开发面临的问题

- 功耗和散热问题
  - 需要进一步降低每一个处理器核的功耗
- 片内互连网络的设计
  - 片内互连网络随着核的数量的增加将占据越来越大的芯片面积比例
- 大容量多端口的共享cache设计
  - 可采用分体式共享cache
    - NUCA
- 存储器墙（memory wall）
- 高效同步机制
- 虚拟化

Recommend
火龙果软件

多核中的虚拟机



- 应用级
- 操作系统级
- 硬件平台级
- 固件级

Recommend
火龙果软件

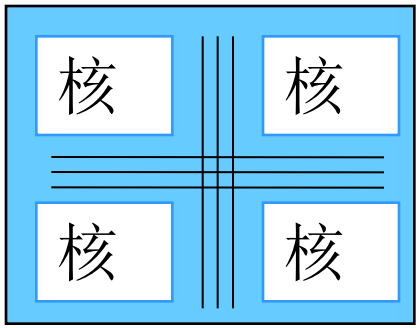
芯片级互连网络

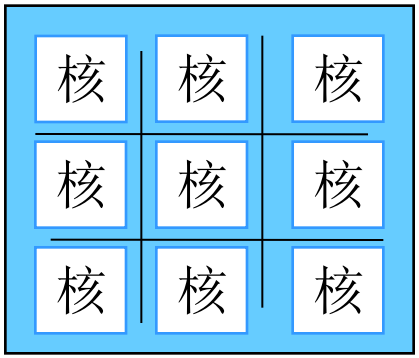
- 片内网络（NoC）
  - 连接片内的多个处理器核
- NoC的优势
  - 通信速度高得多
  - 成本要低得多
  - 连线资源多得多
- NoC的局限
  - 2维芯片表面
  - 适合于2维连接的网络
    - 二维网格
    - 树型网络

Recommend
火龙果软件

NoC需要解决的问题

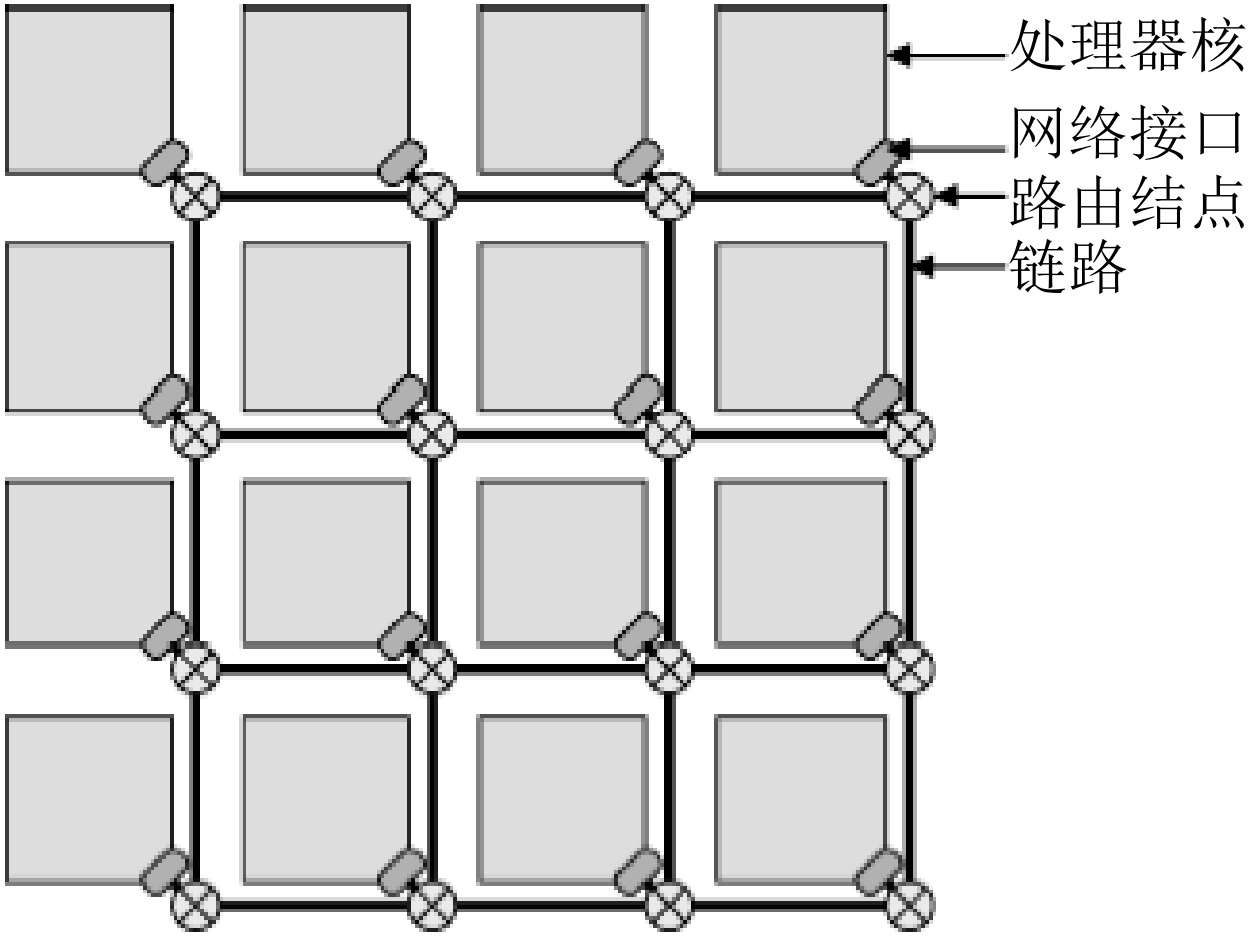
- 长线信号延时
  - 跨越整个芯片的信号传输线路
  - 传输线效应
- 解决方法
  - 增加中继开关
    - 使得传输线变成传输流水线
  - 增加路由结点
    - 多端口中继转发开关





Recommend
火龙果软件

NoC的典型例子



Recommend
火龙果软件

并行处理器芯片实例

- 超线程处理器
  - Intel Hyper-Threading Pentium 4
- 多核芯片
  - Pentium D
  - Core 2 Due
  - Cell

Recommend
火龙果软件

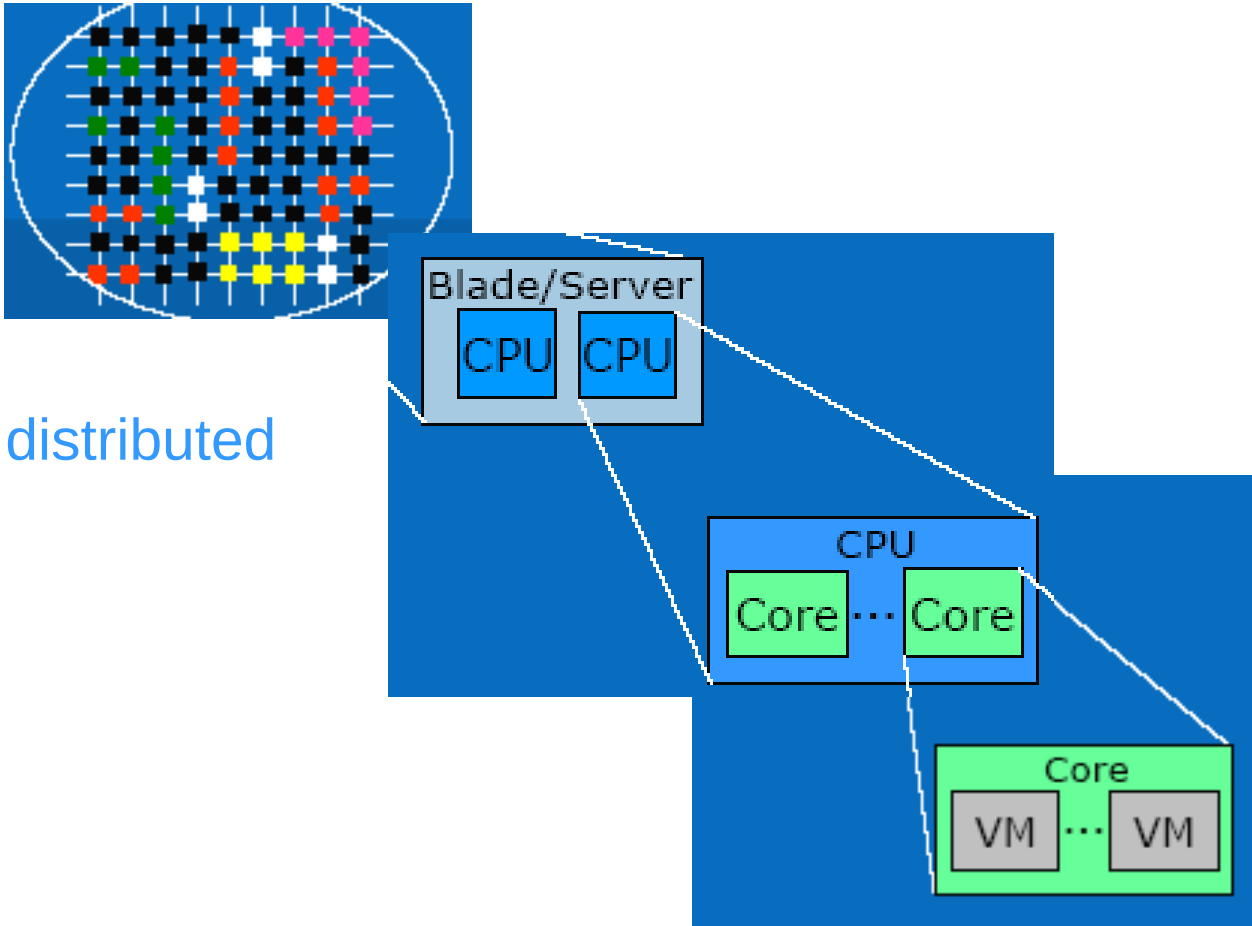
5.4 系统级并行性

- 板级并行系统
  - 一个电路板上有多个处理器芯片和存储器模块
  - 每个处理器芯片可以是单核或多核
  - 可建立起多种层次的虚拟机

Recommend
火龙果软件

多处理机中的虚拟机

- Intra Core
- Inter Core
- Inter Board
- Inter Rack
- Grid
  - Virtualization of distributed processors



Recommend
火龙果软件

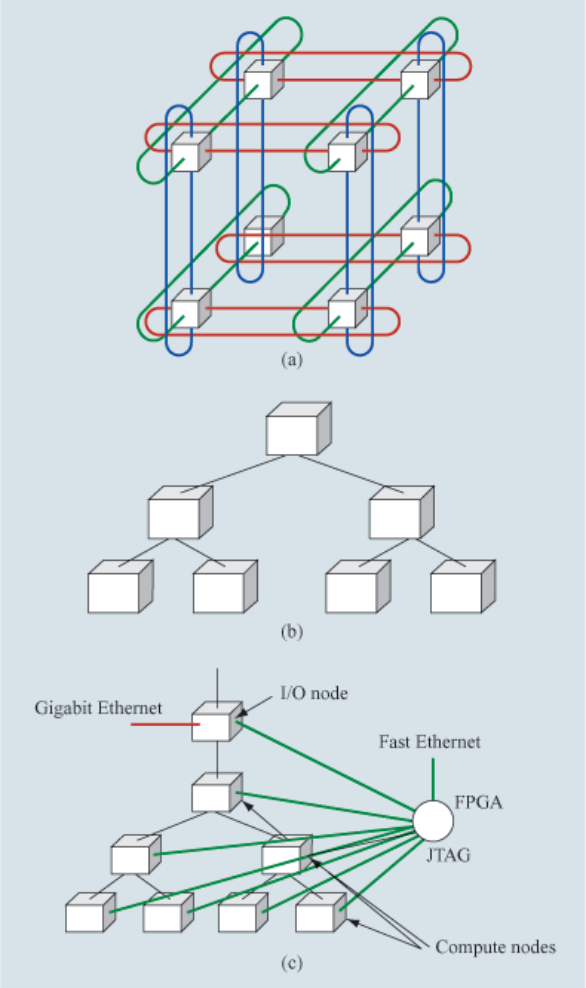
IBM BlueGene/L @server system

- No.1 of TOP500 in Nov. 2005
  - Installed at DOE's Lawrence Livermore National Laboratory (LLNL)
    - Livermore California
- System specification
  - 131072 PowerPC processors
  - 32TB main memory
  - 64 rack
    - 1024 node per rack
- Linpack performance of 280.6 TFlop/s
  - Peak rate 367 TFlop/s

Recommend
火龙果软件

IBM BlueGene/L

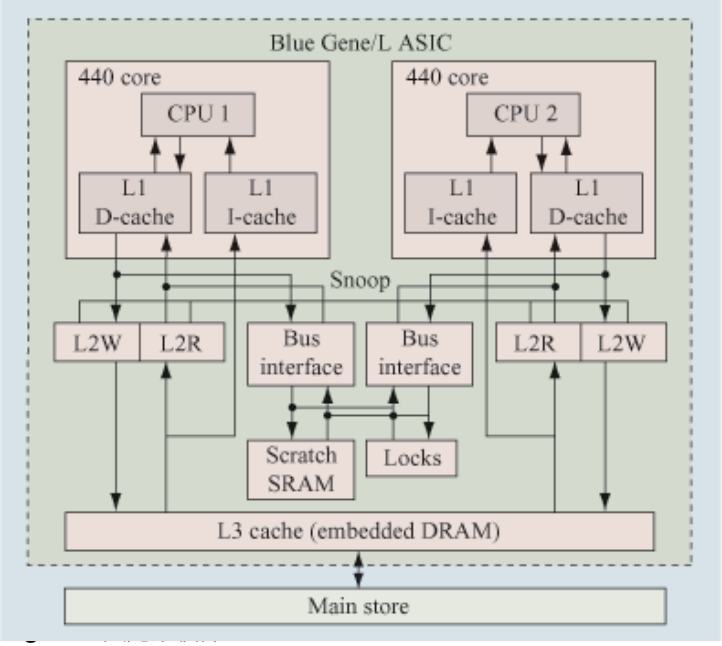
- Network
  - 3D Torus
    - 175 MB/sec in each direction
    - supports cut-through routing
    - 4 virtual channels supported
  - Collective Network
    - For data broadcast
    - 350 MB/sec
    - 1.5 usec latency
  - Global Barrier/Interrupt
    - Reduce latency of global operations
    - Control (system boot, debug, monitoring)
    - Gigabit Ethernet (I/O & connectivity)

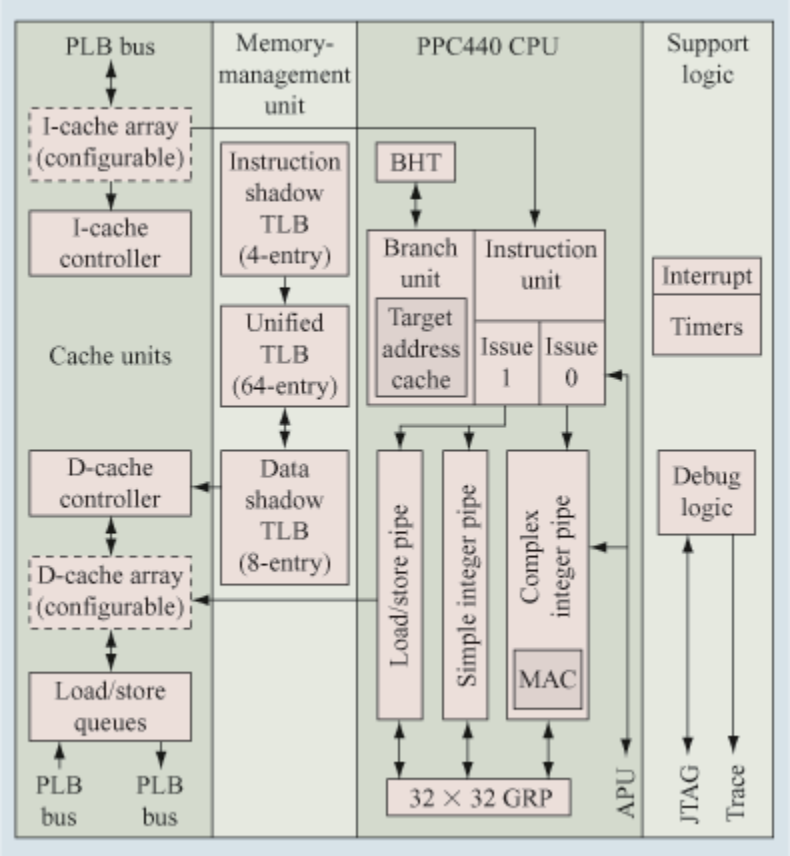


Recommend
火龙果软件

IBM BlueGene/L

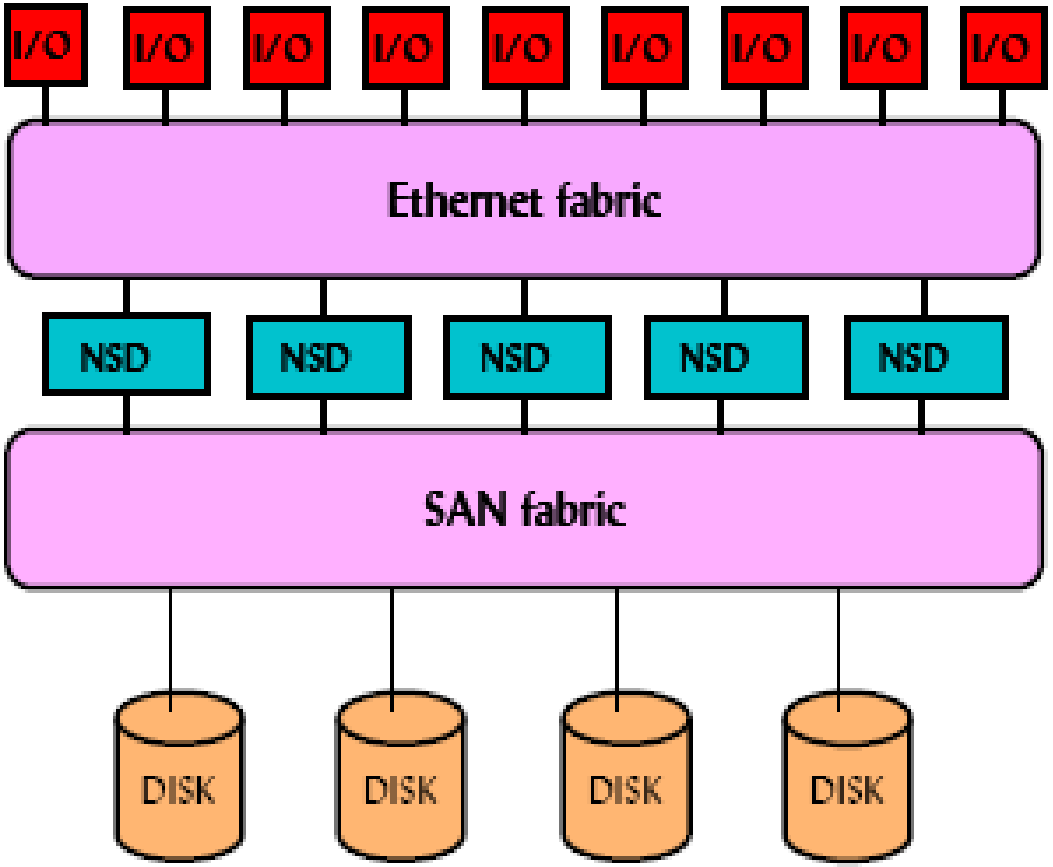
- Computing Node
  - 2 PowerPC 440 700Mhz (2.8 GFlops in all)
    - 7-stage pipeline with dual issue
    - Dual processor fully symmetric
    - 130nm CMOS
  - 512 MB SDRAM-DDR





BlueGene/L的文件子系统

- 第一个层次
  - 用一个以太网将I/O结点连接起来
- 第二个层次
  - 网络共享磁盘（NSD）服务器阵列
- 第三个层次
  - 一个SAN网络和实际磁盘。



Recommend
火龙果软件

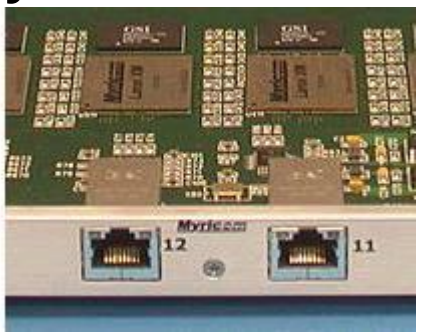
集群并行系统

- 特点
  - 结点的独立性
  - 通过标准通信网络进行协作计算
  - 采用网络操作系统或中间件
    - 建立统一的用户操作界面
  - 可以是建立起共享虚拟存储器（SVM）
  - 容错性好
    - 主机间传递心跳信号
  - 扩展性好

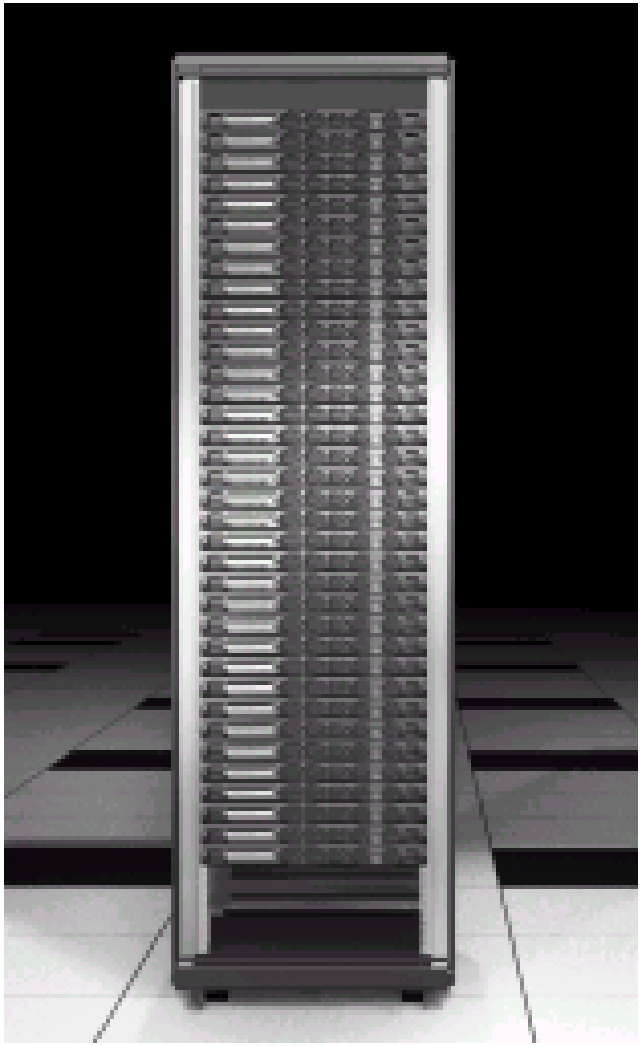
Recommend
火龙果软件

集群并行系统

- 集群系统的网络
  - 以太网
  - Myrinet网



  - Infiniband
  - QsNet



Recommend
火龙果软件

分布式系统

- 更大范围的并行计算机系统
  - 由多个通过计算机广域网络相互连接的独立的计算机组成
- 可实现大型问题的并行计算
  - 充分利用网络上的计算资源和存储资源
- 成本低
- 灵活性强
- 为不同用户提供不同视图

Recommend
火龙果软件

分布式系统

- 透明性

  (1) 位置透明

  (2) 迁移透明

  (3) 复制透明

  (4) 同时性透明

  (5) 并行性透明

- 目前的主要实现方式
  - 网格

Recommend
火龙果软件