



高性能计算技术

第二讲 并行计算机体系结构（1）

何克晶

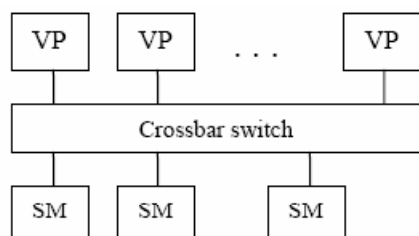
kjhe@scut.edu.cn

华南理工大学计算机学院

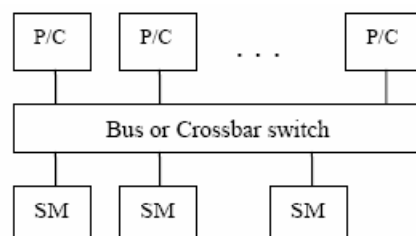
复习：并行计算机体系结构

- Single Instruction Stream Over Multiple Data Streams (SIMD)
- Parallel Vector Processor (PVP)
- Symmetric Multiprocessors (SMP)
- Massively Parallel Processors (MPP)
- Distributed Shared Memory (DSM)
- Cluster of Workstation (COW)

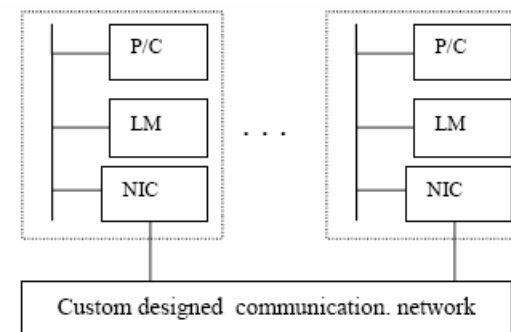
复习：并行计算机体系结构



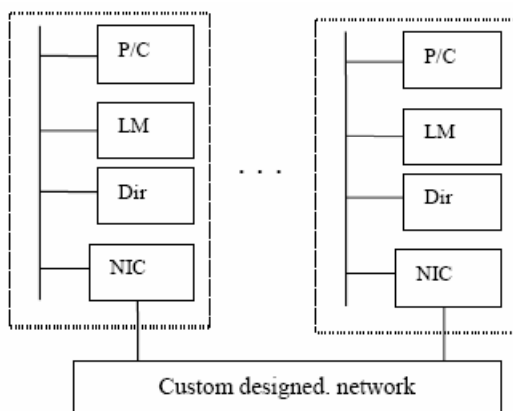
PVP - Parallel Vector Processor



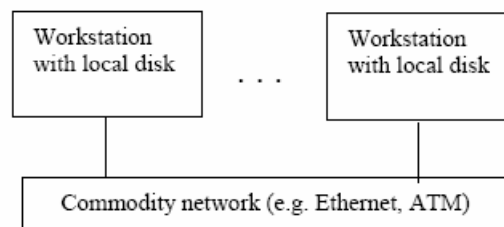
SMP - Symmetric Multiprocessors



MPP - Massively Parallel Processors



DSM - Distributed Shared Memory



COW - Cluster of Workstation

SM - shared memory module
LM - local memory
NIC - network interface circuitry
VP - vector processor
P/C - scalar processor and cache
Dir - address directory/translation

计算机的发展

- 40年代开始的现代计算机发展历程可以分为两个明显的发展时代：串行计算时代、并行计算时代。每一个计算时代都从体系结构发展开始，接着是系统软件（特别是编译器与操作系统）、应用软件，最后随着问题求解环境的发展而达到顶峰
- 创建和使用并行计算机的主要原因是因为并行计算机是解决单处理器速度瓶颈的最好方法之一



Cray X1

Latest Cray
Supercomputer

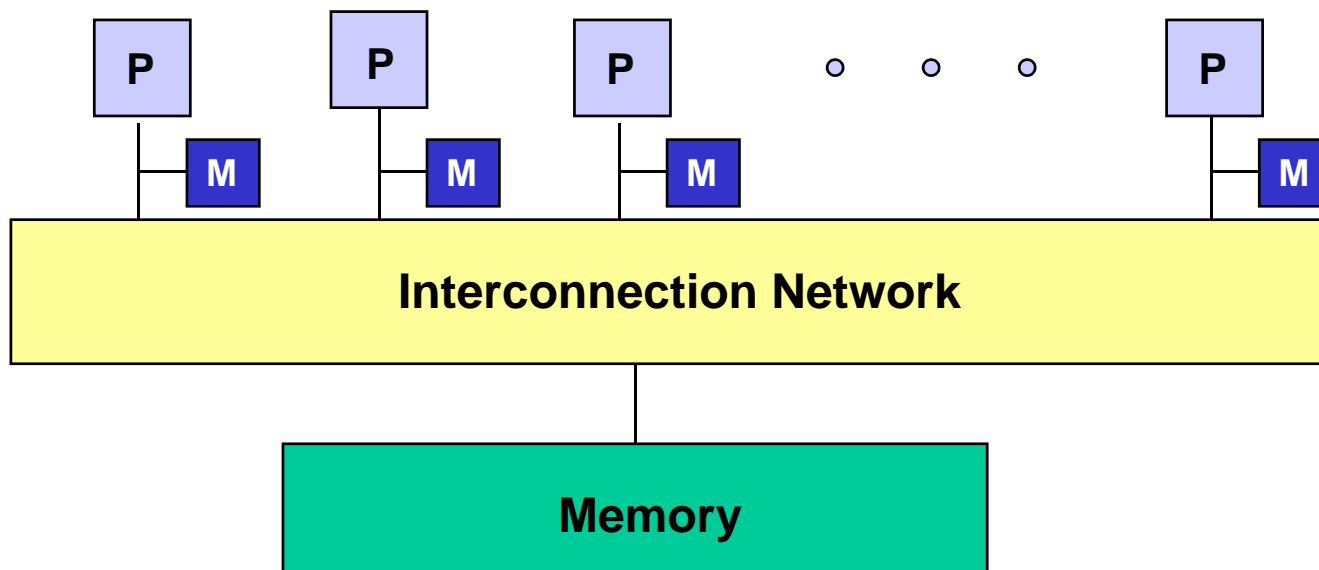
Up to 52.4
teraflops of
peak
computing
power and
65.5 TB of
memory

U.S. list pricing
starts at about
\$2.5 million.

并行结构

- 并行计算机是由一组处理单元组成的，这组处理单元通过相互之间的通信与协作，以更快的速度共同完成一项大规模的计算任务。因此，并行计算机的两个最主要的组成部分是计算节点和节点间的通信与协作机制
- 并行计算机体系结构的发展也主要体现在计算节点性能的提高以及节点间通信技术的改进两方面
- 并行计算机与传统计算机的区别在于其通信架构：
 - 互联网络的特性
 - 处理器的通信、同步方法等

一个通用的并行结构

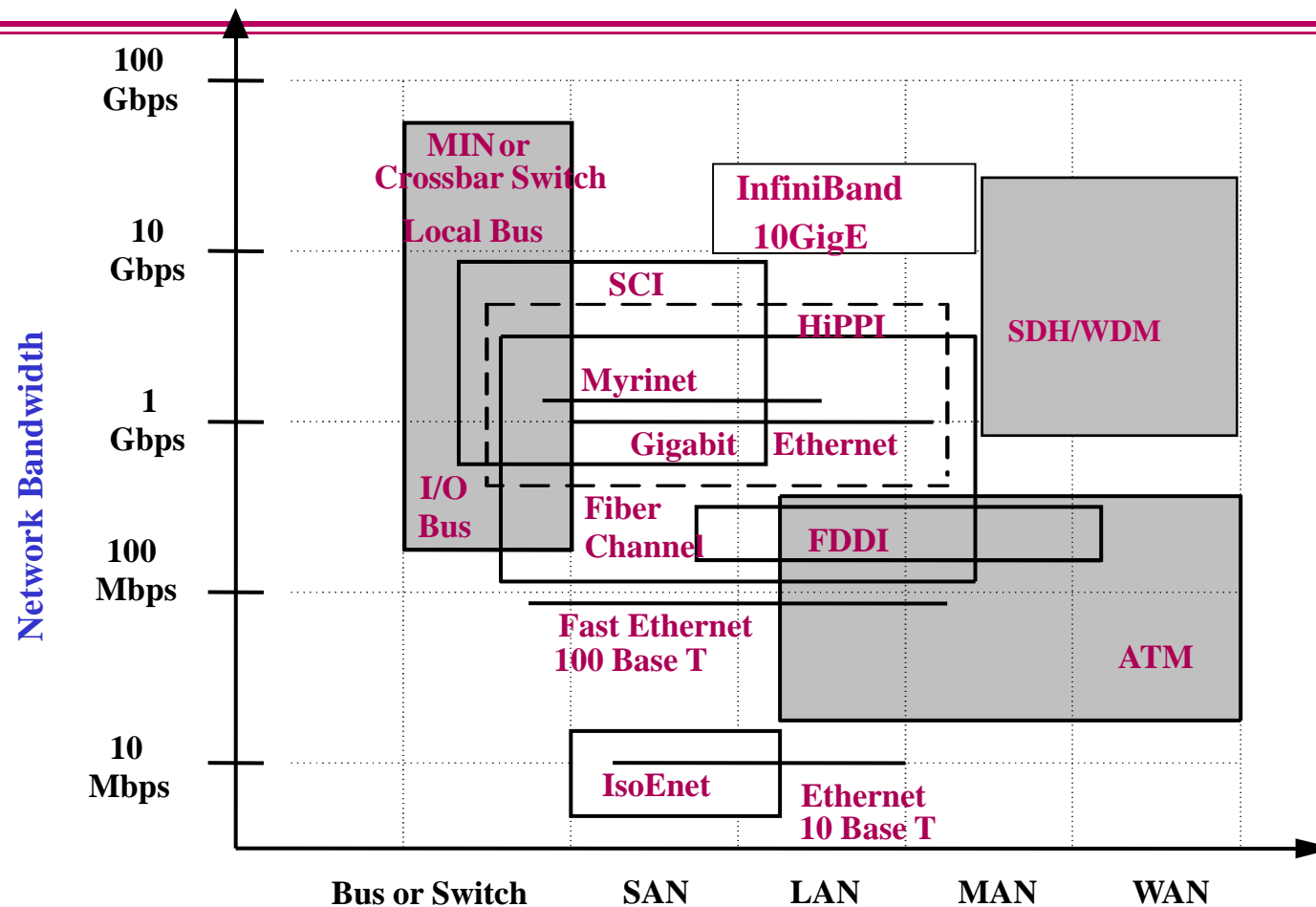


网络拓扑：处理器、内存和I/O的连接方式

主要内容

- 系统互联
- 静态互联网络
- 动态互连网络
- 标准互联网络

不同带宽与距离的互连技术



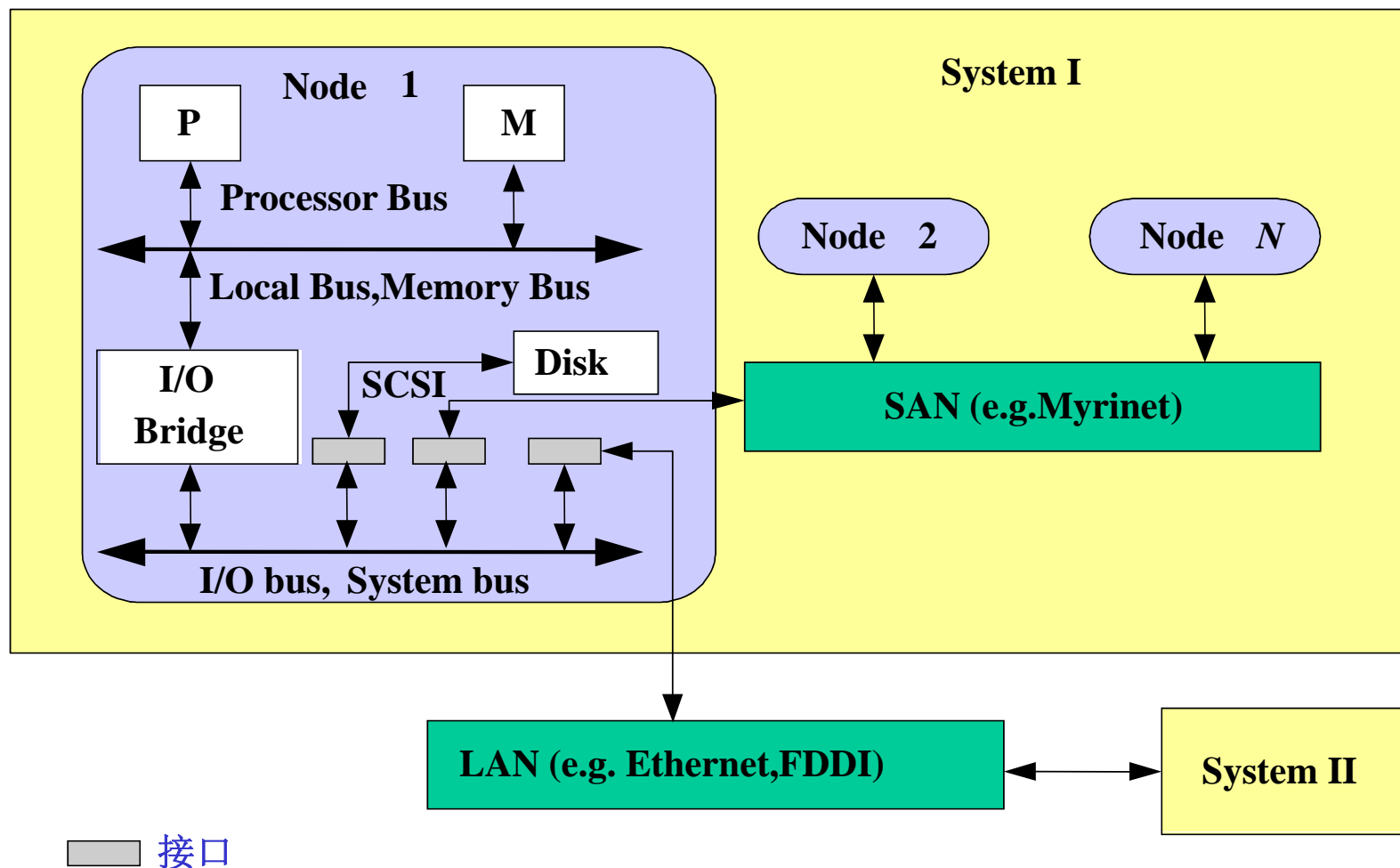
SAN: System Area Network (系统域)

LAN: Local Area Network (局域网)

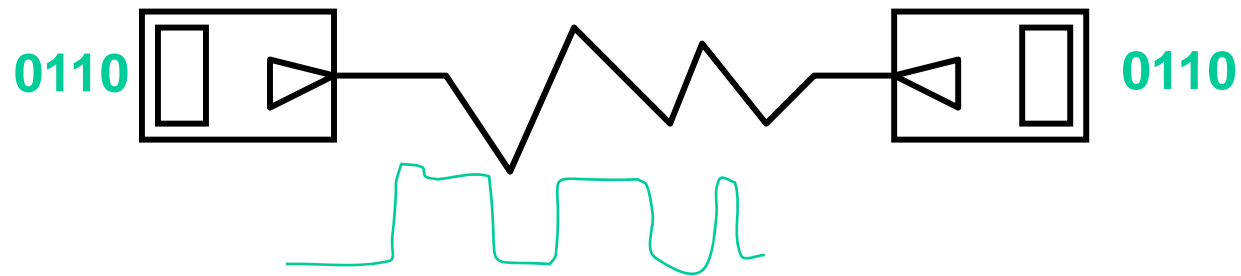
MAN: Metropolitan Area Network (城域网)

WAN: Wide Area Network (广域网)

本地总线, IO总线, SAN 和 LAN

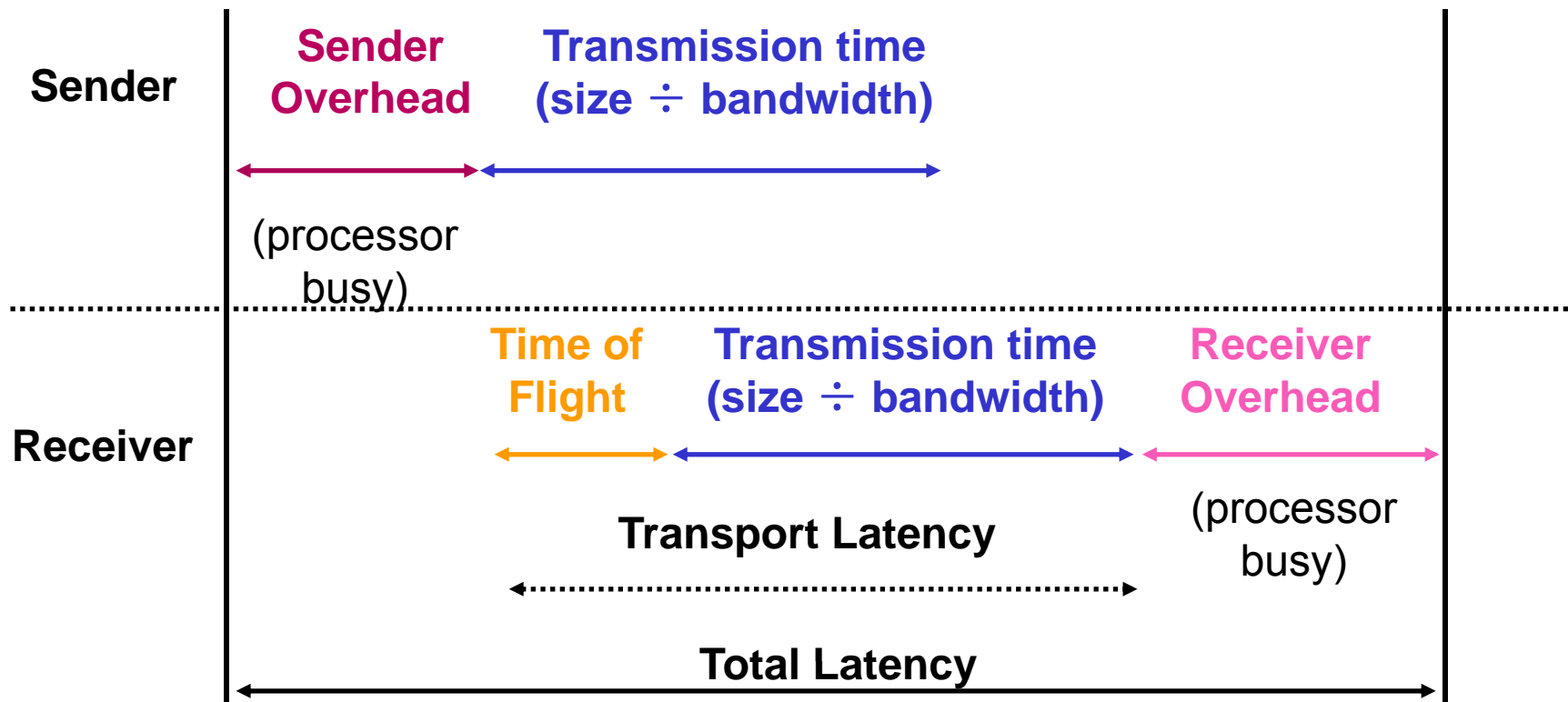


复习：基本网络部件



- 链路（link）：传输信息的物理介质
- 交换机/路由器开关（switch/router）：用于建立交换网络
- 网络接口电路（NIC）：用来连接主机和网络

网络性能指标—时延

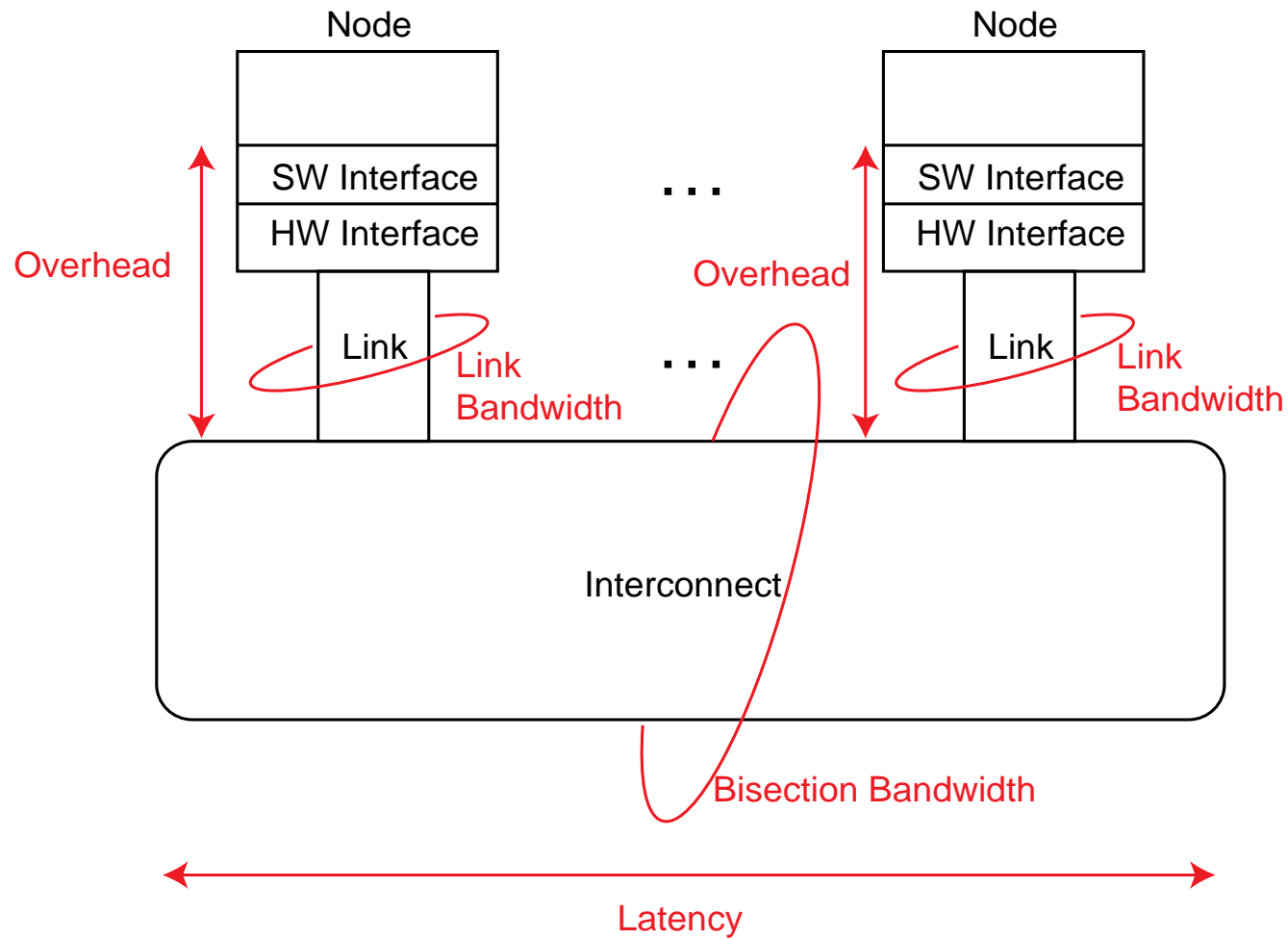


$$\text{Total Latency} = \text{Sender Overhead} + \text{Time of Flight} + \text{Message Size} \div \text{BW} + \text{Receiver Overhead}$$

时延 (Latency)

- **通信时延**：从源节点到目的节点传输一条消息所需的总时间
 - 在网络两端相应收发消息的**软件开销**
 - 由于通道占用导致的**通道时延**，即总的消息长度除以通道带宽
 - 沿选路路径作一系列选路决策期间花费在后续交换开关上的**选路时延**
 - 由于网络传输竞争导致的**竞争时延**
- **软件开销** (overhead) 主要取决于主机内核，与**竞争时延**均依赖于程序行为
- **传输时延** (transport latency)：通道时延 (transmission time) 和选路时延之和，完全由网络硬件特征决定。轻载消息传输无竞争网络系统中的网络时延通常1微秒左右，小于软件开销和竞争时延（几十或几百微秒）

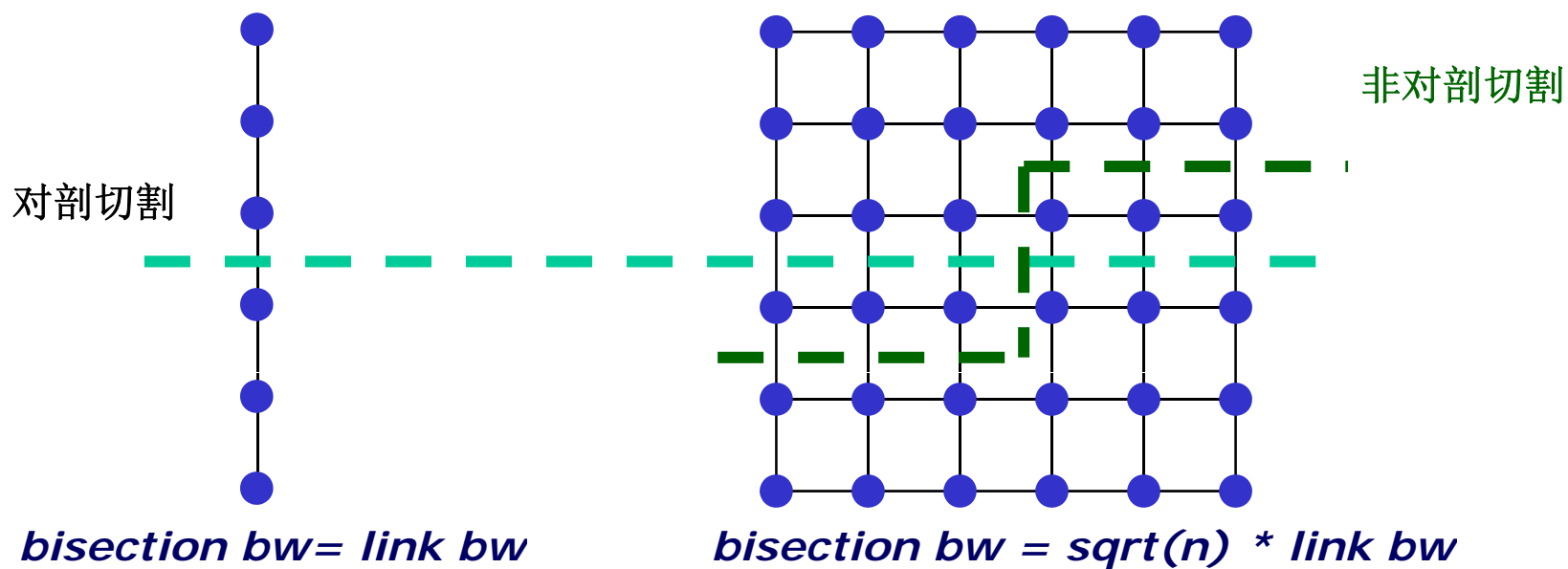
网络性能指标—带宽



带宽（Bandwidth: BW）

- 端口带宽：从任意端口到另外端口单位时间内传输消息的最大位（或字节）数
 - 如IBM HPS 每端口带宽40MB/s
- 聚集带宽：从一半节点到另一半节点，单位时间内传输消息的最大位（或字节）数
 - 如IBM HPS端口数最多为512，聚集带宽为 $512 * 40 / 2 = 10.24\text{GB/s}$
- 链路带宽（Link Bandwidth）：单位时间内链路传输消息的最大位（或字节）数
- 对剖宽度：将网络分成两个相等部分所必须移去的最少边数。
- 对剖带宽（Bisection Bandwidth）：每秒钟内，在最小的对剖平面上通过所有连线的最大信息位（或字节）数。等于对剖宽度与链路带宽之积。

对剖带宽与链路带宽



网络性能指标—网络拓扑

- **节点度** (Node Degree) : 射入或射出一个节点的边数。在单向网络中, 入射和出射边之和称为节点度
- **网络直径** (Network Diameter) : 网络中任何两个节点之间的最长距离, 即最大路径数
- 如果从任一节点观看网络都一样, 则称网络为对称的 (Symmetry)
- **边连通度** (arc connectivity): 将网络分成两个不连通的部分所必须移去的边数
- **代价** (cost) : 可以用总边数 (链路数) 来衡量

互联网络的评价标准

- 硬件复杂度（Cost）
 - 将N个处理机按一定拓扑结构连成网络所需的开关个数
- 时延（Latency）
 - 发送消息到接收消息所需的时间
- 带宽（Bandwidth）
 - 单位时间内传送的数据量
- 可扩展性（Scalability）
- 容错能力（Reliability）

静态互连网络与动态互连网络

- 静态互连网络：处理单元间有着固定连接的一类网络，在程序执行期间，这种点到点的链接保持不变；典型的静态网络：
 - 一维线性阵列（1-Linear Array）
 - 二维网孔（2-D Mesh）
 - 树（Tree）
 - 超立方（Hypercube）
 - 蝶形网络（Butterfly）
- 动态网络：用交换开关构成的，可按应用程序的要求动态地改变连接组态；典型的动态网络：
 - 总线（Bus）
 - 交叉开关（Crossbus Switcher）
 - 多级互连网络（MIN-Multistage Interconnection Network）

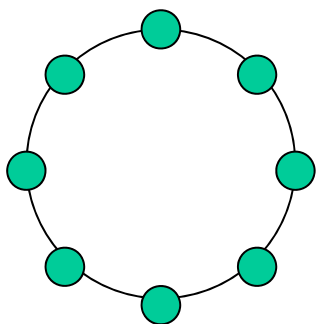
主要内容

- 系统互联
- 静态互联网络
- 动态互连网络
- 标准互联网络

一维线性阵列 (1-D Linear Array)



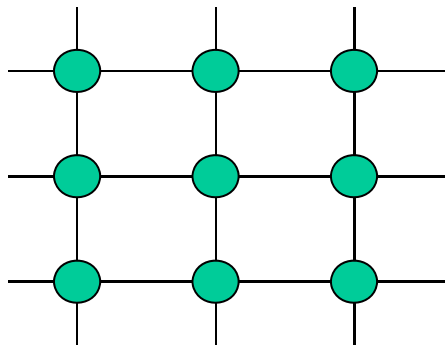
Linear Array



Ring

- 并行机中最简单最基本的互连方式
- 每个节点只与其左、右近邻相连，也叫二近邻连接，
- N 个节点用 $N-1$ 条边串接之，内节点度为2，直径为 $N-1$ ，对剖宽度为1
- 当首、尾节点相连时可构成循环移位器，在拓扑结构上等同于环，环可以是单向的或双向的
 - 节点度恒为2
 - 直径或为 $\lfloor N/2 \rfloor$ （双向环）或为 $N-1$ （单向环）
 - 对剖宽度为2
- 例子：FDDI, SCI

二维网孔 (2-D Mesh)



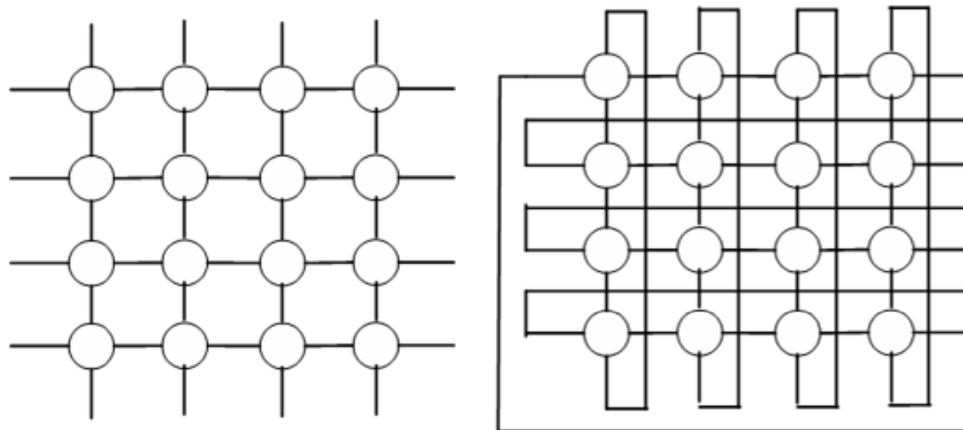
2-D Mesh

- 每个节点只与其上、下、左、右的近邻相连（边界节点除外）
 - 网络规模: N
 - 节点度: 4
 - 网络直径: $2(N^{1/2} - 1)$
 - 对剖宽度: $N^{1/2}$
- 例子: Intel paragon

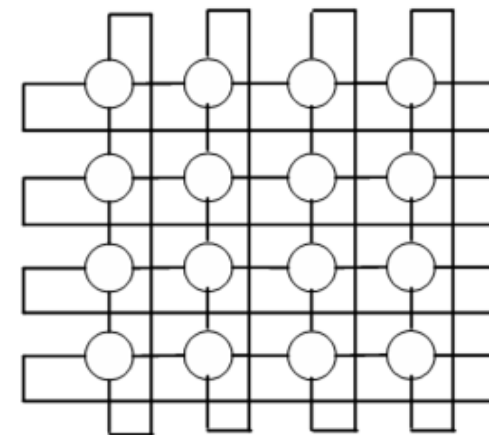
二维网孔 (2-D Mesh)

- 在垂直方向上带环绕，水平方向呈蛇状，就变成Illiac网孔了
 - 节点度恒为4，网络直径为 $N^{1/2} - 1$ ，而对剖宽度为 $2N^{1/2}$
- 垂直和水平方向均带环绕，则变成了2-D环绕 (2-D Torus)
 - 节点度恒为4，网络直径为 $2\lfloor N^{1/2}/2 \rfloor$ ，对剖宽度为 $2N^{1/2}$

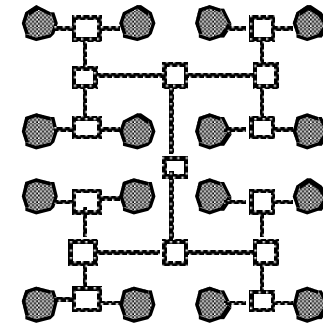
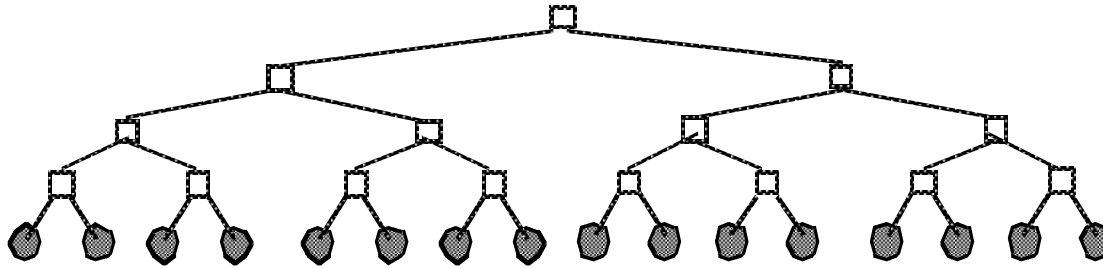
Illiac Mesh



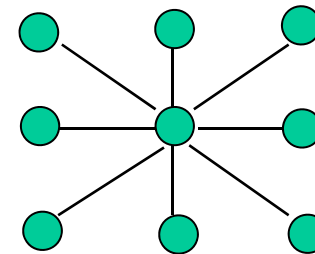
2-D Torus



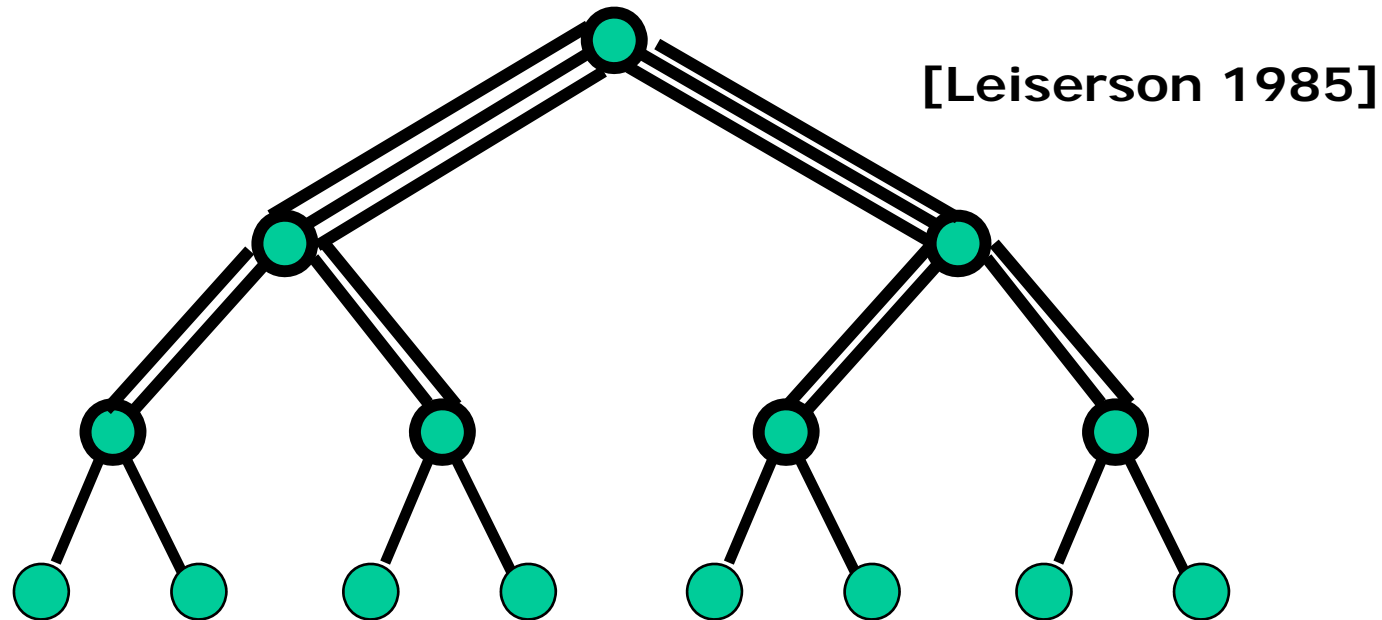
二叉树



- 除了根、叶节点，每个内节点只与其父节点和两个子节点相连。
 - 节点度为3（三近邻连接），对剖宽度为1，
 - 树的直径为 $2*k-2=2\lceil\log N\rceil-1$ （k为层数，N为树的总节点数）
- 如果尽量增大节点度为N-1，则直径缩小为2，此时就变成了星形网
 - 对剖宽度为 $\lfloor N/2 \rfloor$

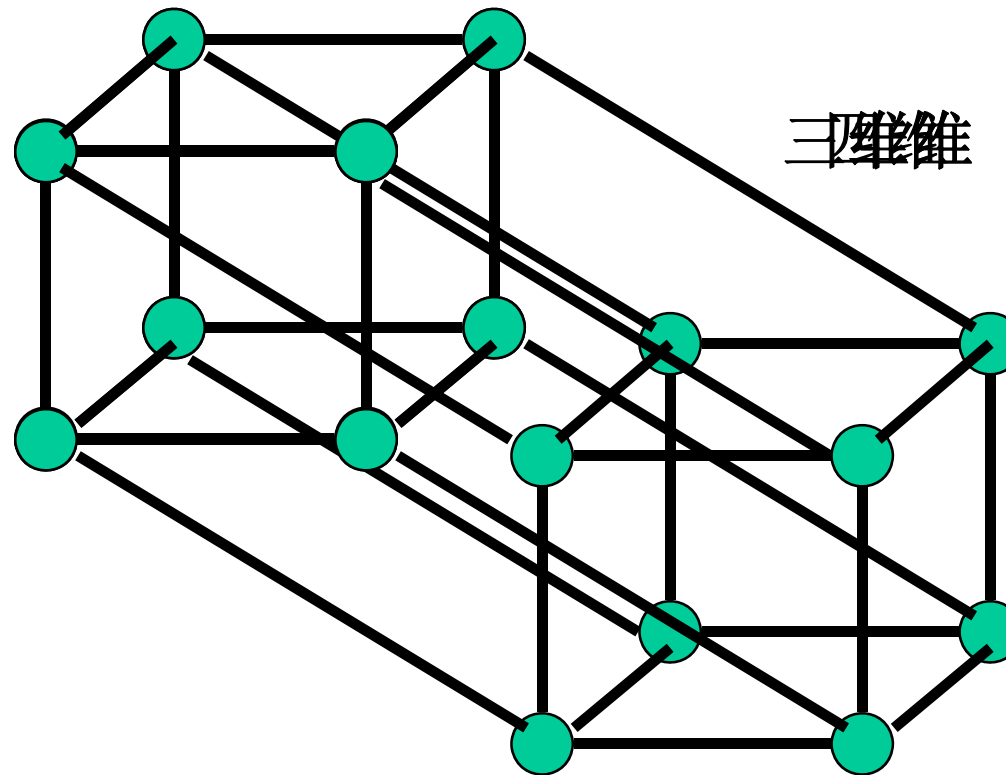


胖树 (Fat-Trees)



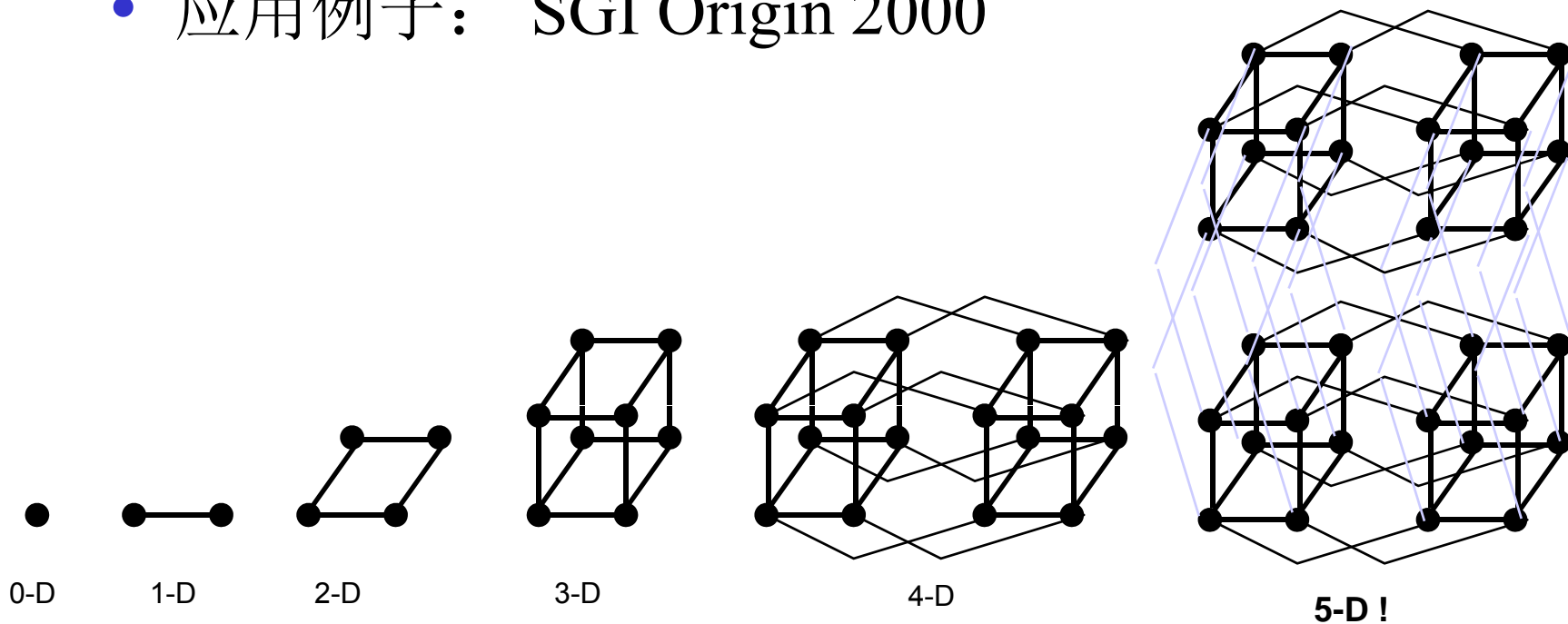
- 传统二叉树的主要问题是根易成为通信瓶颈。胖树节点间的通路自叶向根逐渐变宽
- 对剖带宽随着N的增大而增大
- 例子: Infiniband

超立方 (Hypercube)



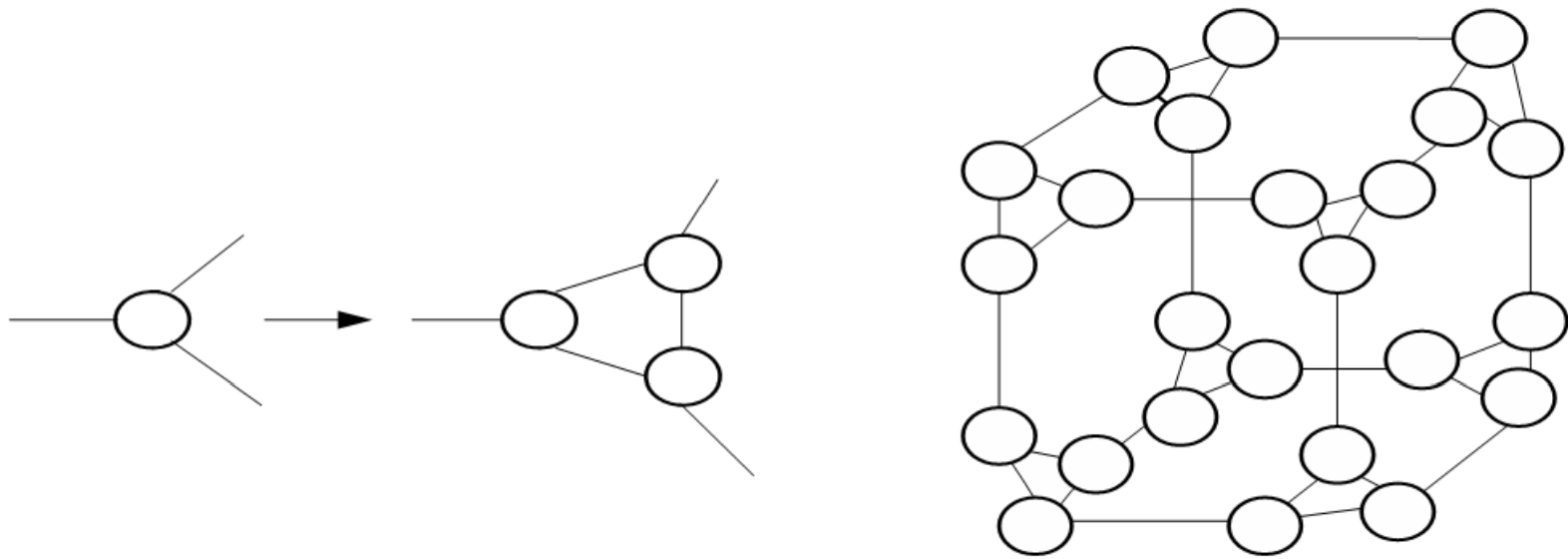
超立方

- 一个n-立方由 $N=2^n$ 个顶点组成
 - n-立方的节点度为n，网络直径也是n，而对剖宽度为 $N/2$
- 应用例子： SGI Origin 2000



3-立方环

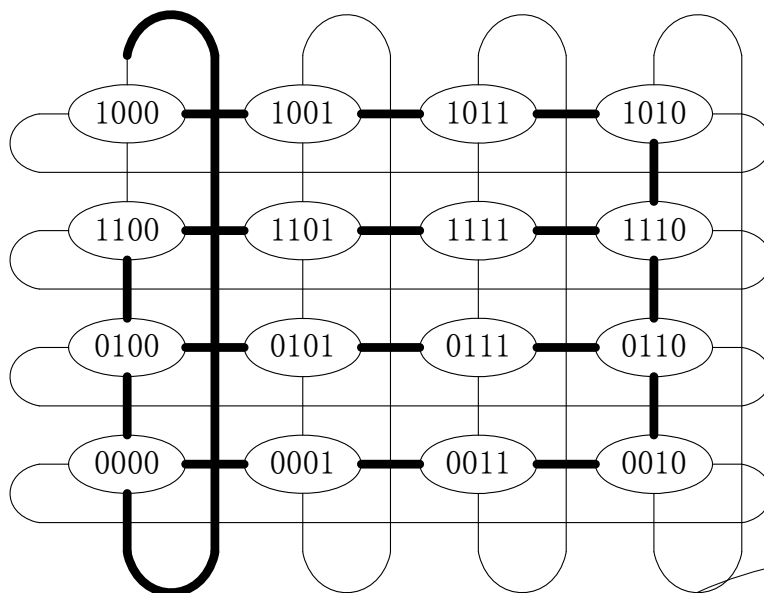
- 如果将3-立方的每个顶点代之以一个环就构成了3-立方环，此时每个顶点的度为3，而不像超立方那样节点度为 n 。



嵌入（ Embedding ）

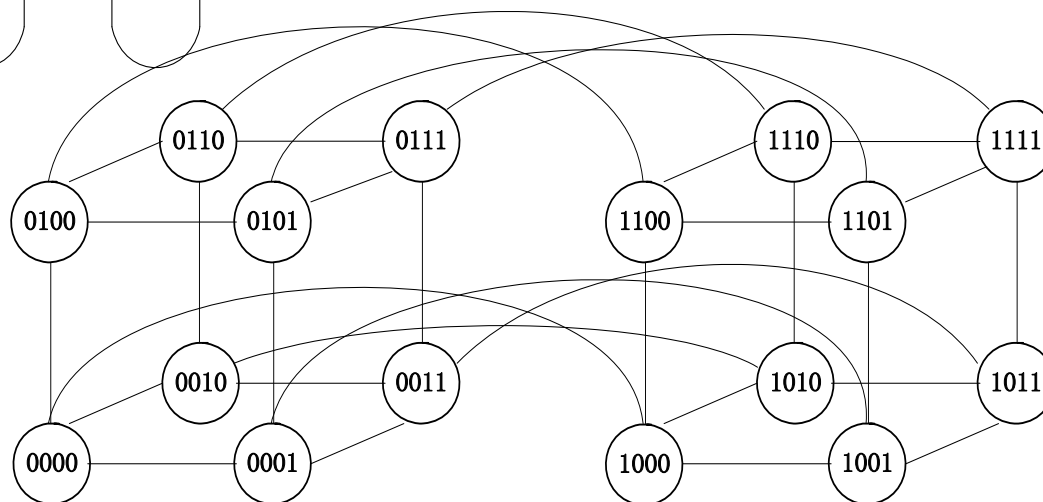
- 将网络中的各节点映射到另一个网络中去
- 用膨胀（Dilation）系数来描述嵌入的质量，它是指被嵌入网络中的一条链路在所要嵌入的网络中对应所需的最大链路数
- 如果该系数为1，则称为完美嵌入。
- 环网可完美嵌入到2-D环绕网中
- 超立方网可完美嵌入到2-D环绕网中

嵌入



Ring onto 2-D torus

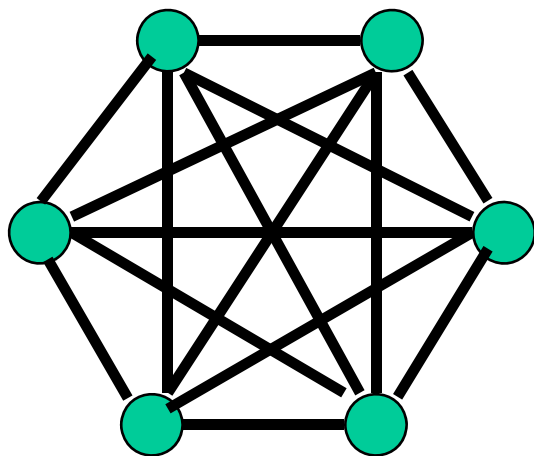
Hypercube onto 2-D torus



静态互连网络特性比较

网络名称	网络规模	节点度	网络直径	对剖宽度	对称	链路数
线性阵列	N	2	$N-1$	1	非	$N-1$
环形	N	2	$\lfloor N/2 \rfloor$	2	是	N
2-D网孔	$N = n^2$	4	$2(n-1)$	n	非	$2(N-n)$
Illiac网孔	$N = n^2$	4	$n-1$	$2n$	非	$2N$
2-D环绕	$N = n^2$	4	$2 \lfloor n/2 \rfloor$	$2n$	是	$2N$
二叉树	N	3	$2(\lceil \log_2 N \rceil - 1)$	1	非	$N-1$
星形	N	$N-1$	2	$\lfloor N/2 \rfloor$	非	$N-1$
超立方	$N=2^n$	$\log_2 N=n$	N	$N/2$	是	$nN/2$
立方环	$N=k \cdot 2^k$	3	$2k-1 + \lfloor k/2 \rfloor$	$N/(2k)$	是	$3N/2$

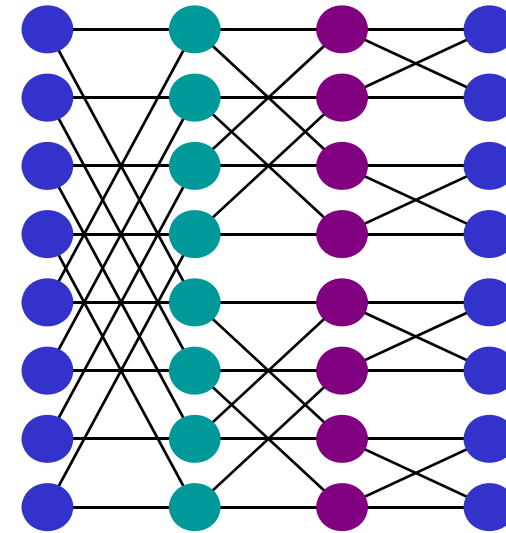
例子：完全连接



- 网络规模: N
- 连接数: $N*(N - 1) / 2$
- 节点度: $N - 1$
- 对剖宽度: $(N / 2)^2$
- 直径 = 1

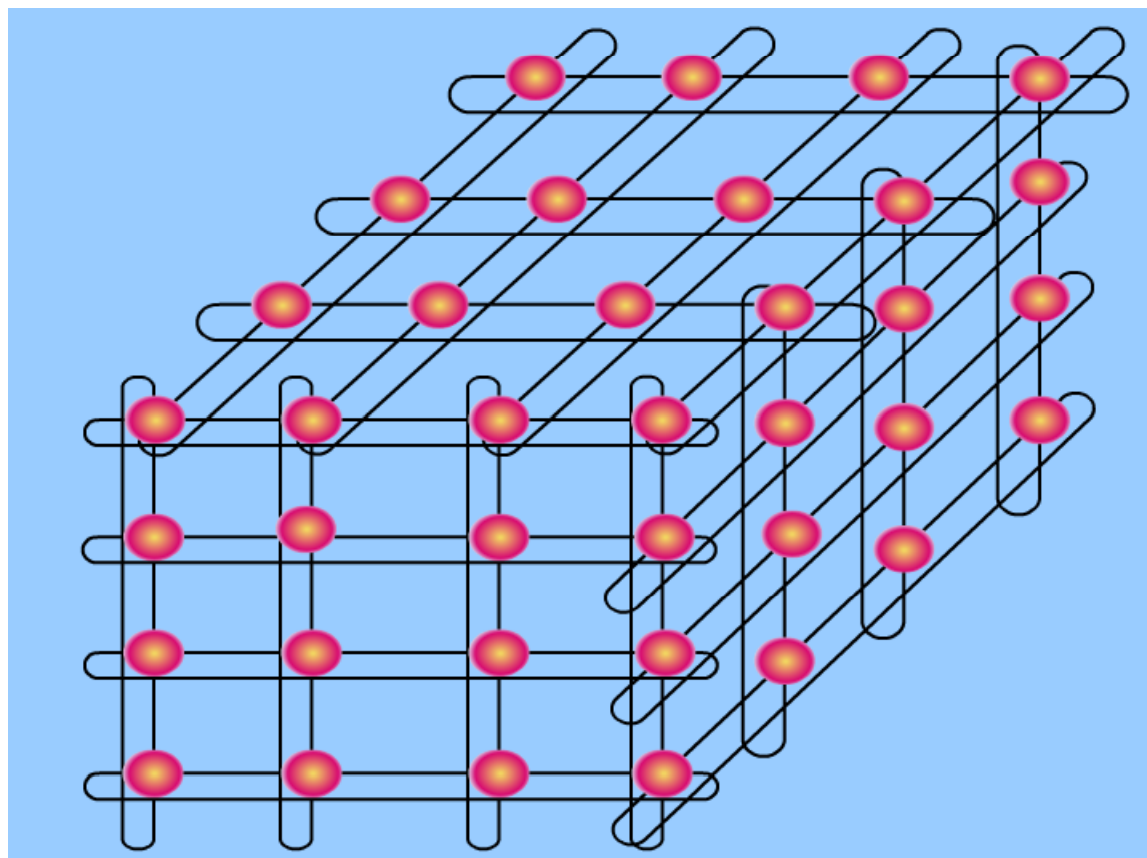
例子：蝶网 (butterfly)

- 节点: $N = (k+1)2^k$
- 直径: $2k$
- 对剖带宽: 2^k



multistage butterfly network

例子：4元3立方体的网络拓扑结构




- k元-n立方体网络中，参数k是基数或者说是沿每个方向的结点数，n是立方体的维数。这两个数与网络中节点数N的关系为 $N=k^n$

静态互联网络小结

- 大多数静态网络的节点度都小于4，这是比较理想的。若能实现所有节点的连接，节点度愈小愈好，当然要求相应的网络时延也是愈小愈好
- 节点度愈大，表示连接性愈好，但网络的连接复杂，成本高
- 对剖带宽愈大，表示网络的带宽就愈大；网络直径愈大，表示通信的时间延迟就愈大。
- 网络的总价格随节点度和链路数增大而上升。
- 固定网络规模，对剖宽度越大，或网络直径越小，则互联网络质量越高
- 对称性会影响可扩展性和路由效率
- 环、Mesh、环绕、超立方体、 k 元 n -立方体可用于MPP系统

真实机器中的拓扑结构

 older newer	Red Storm (Opteron + Cray network)	3D Mesh
	IBM Blue Gene/L	3D Torus
	SGI Altix	Fat tree
	Cray X1	4D Hypercube
	Quadrics	Fat tree
	IBM SP	Fat tree (近似)
	SGI Origin	Hypercube
	Intel Paragon	2D Mesh

主要内容

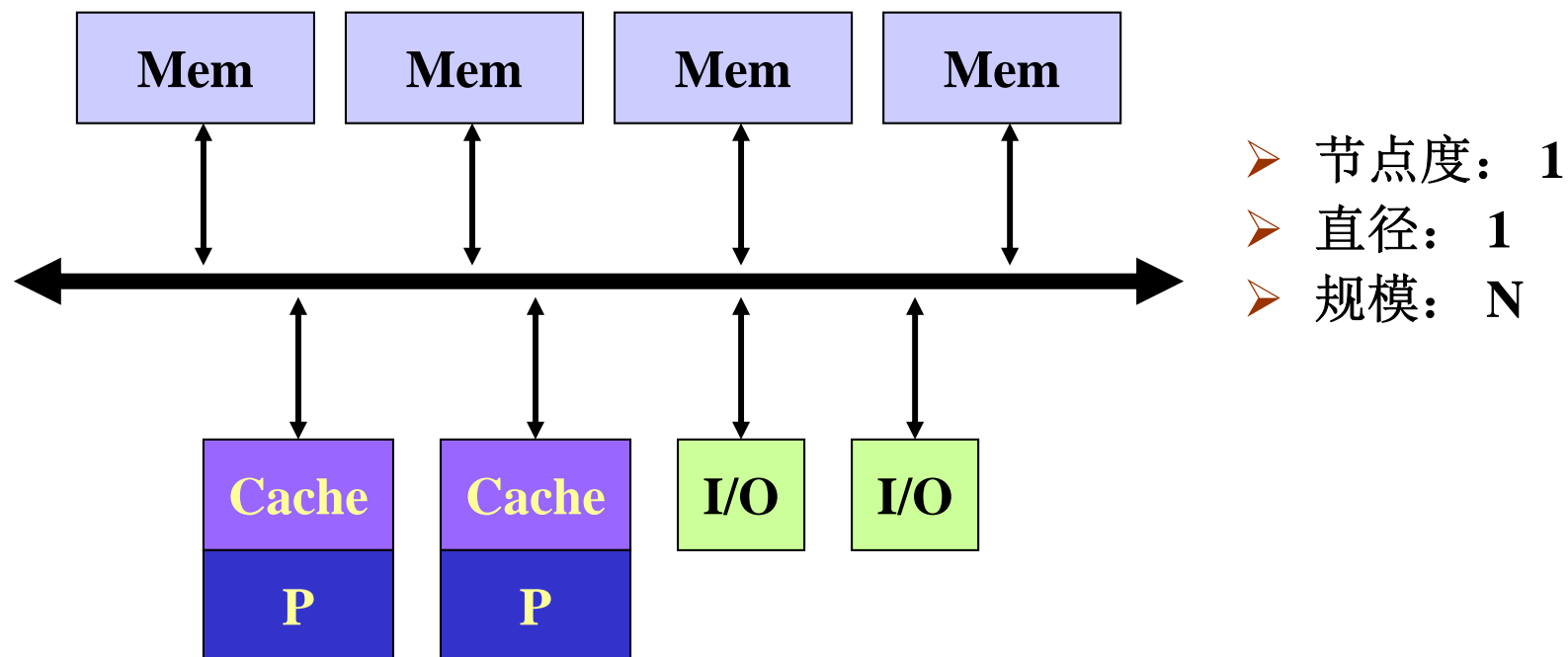
- 系统互联
- 静态互联网络
- 动态互连网络
- 标准互联网络

动态互联网络

- 通信模式是基于程序的要求
- 连接实在程序执行过程总实时建立
- 基于总线（Bus-based）或开关（Switch-based）

总线

- 总线：PCI、VME、Multics、Sbus、MicroChannel, IEEE Futurebus



多处理机总线和层次总线，常用来构筑SMP、NUMA和DSM机器

多处理机总线

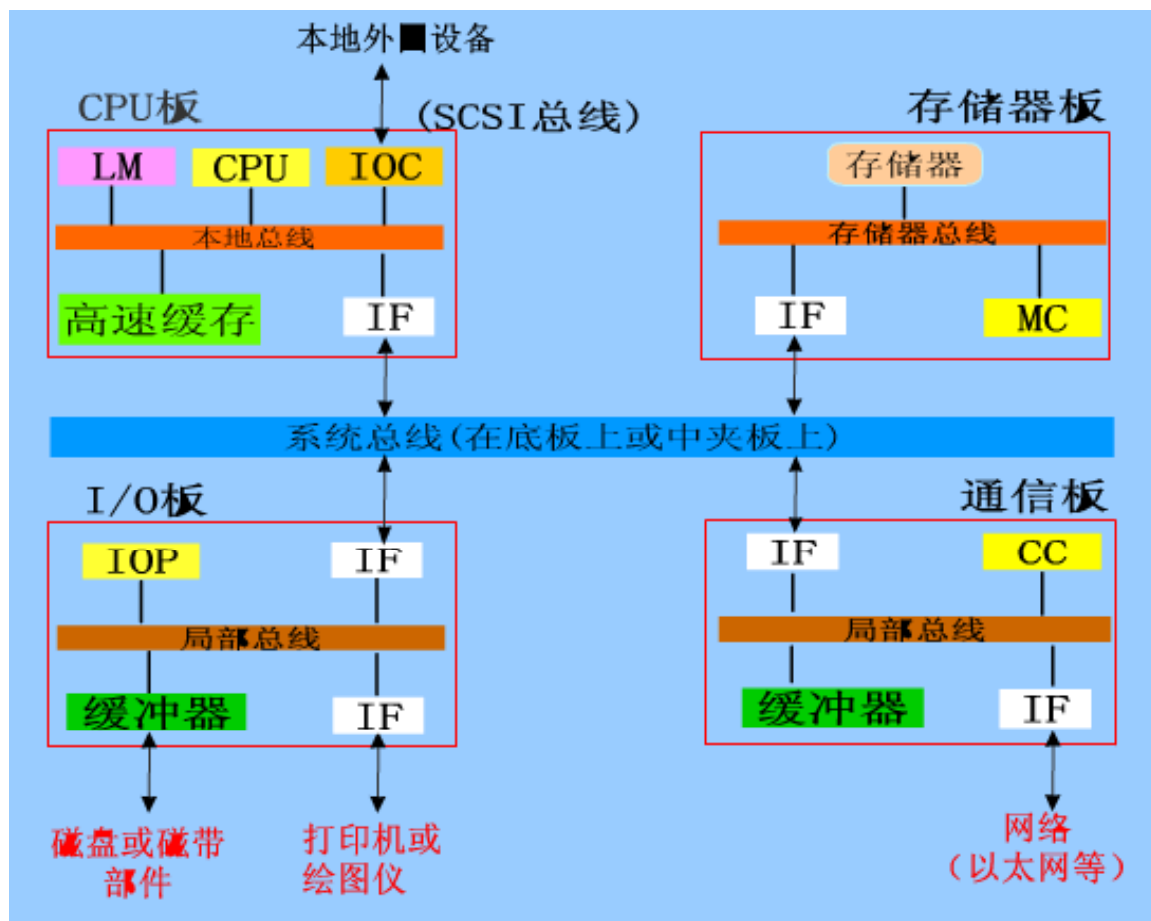
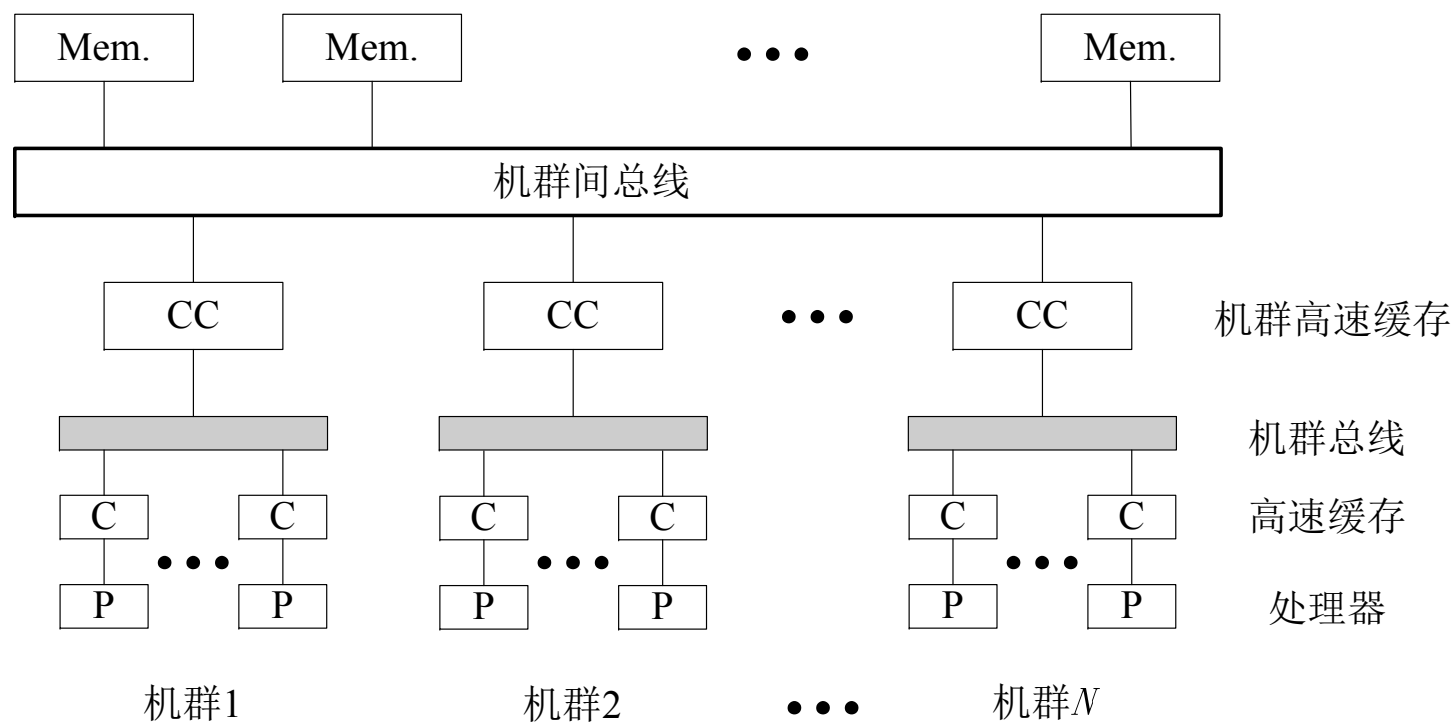


图1.9

层次总线： IEEE Futurebus

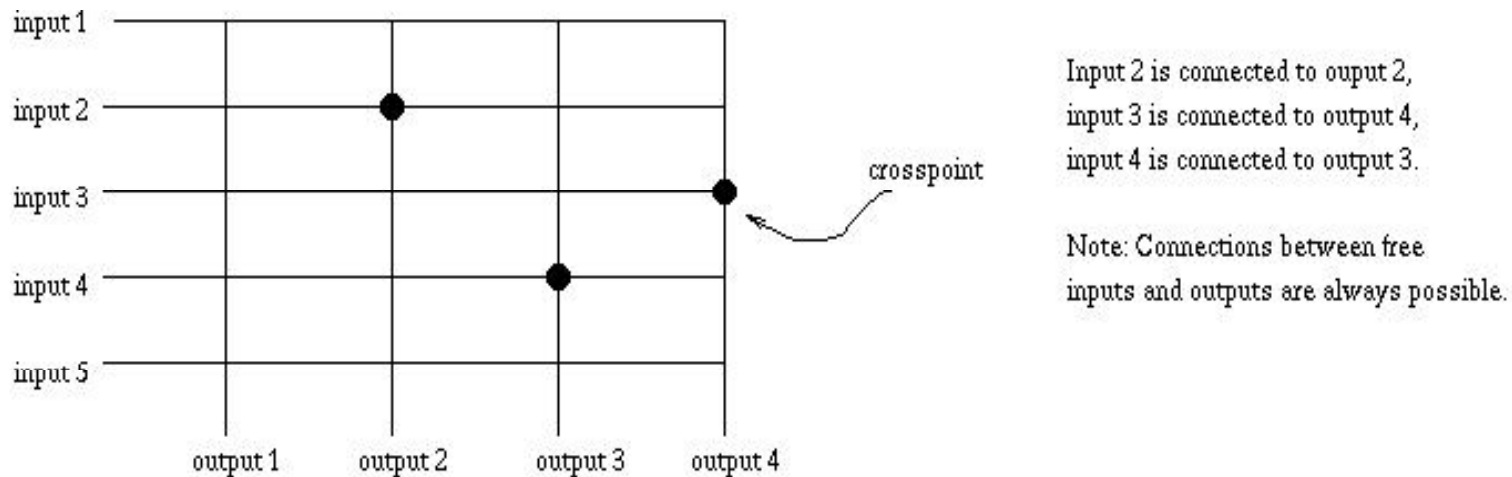


单个SMP总线在构造大规模系统时仅有有限的可扩展性，层次总线结构可在一定程度上缓和这个矛盾。

总线的特点

- 总线的优点在于成本低，不随处理器数目的增加而增加
- 总线的缺点在于扩展性不好，总线的带宽固定，随着处理器数的增加，每个处理器带宽减少
- 可利用程序中的局部性原理减少对总线带宽的需求

交叉开关

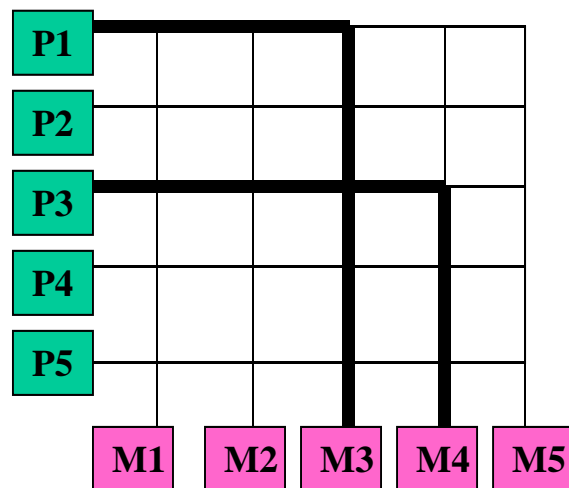


A Small Crossbar Switch

- 单级交换网络，可为每个端口提供更高的带宽。象电话交换机一样，交叉点开关可由程序控制动态设置其处于“开”或“关”状态，而能提供所有（源、目的）对之间的动态连接。
- 交叉开关一般有两种使用方式：一种是用于对称的多处理机或多计算机机群中的处理器间的通信；另一种是用于SMP服务器或向量超级计算机中处理器和存储器之间的存取。

交叉开关的应用例子

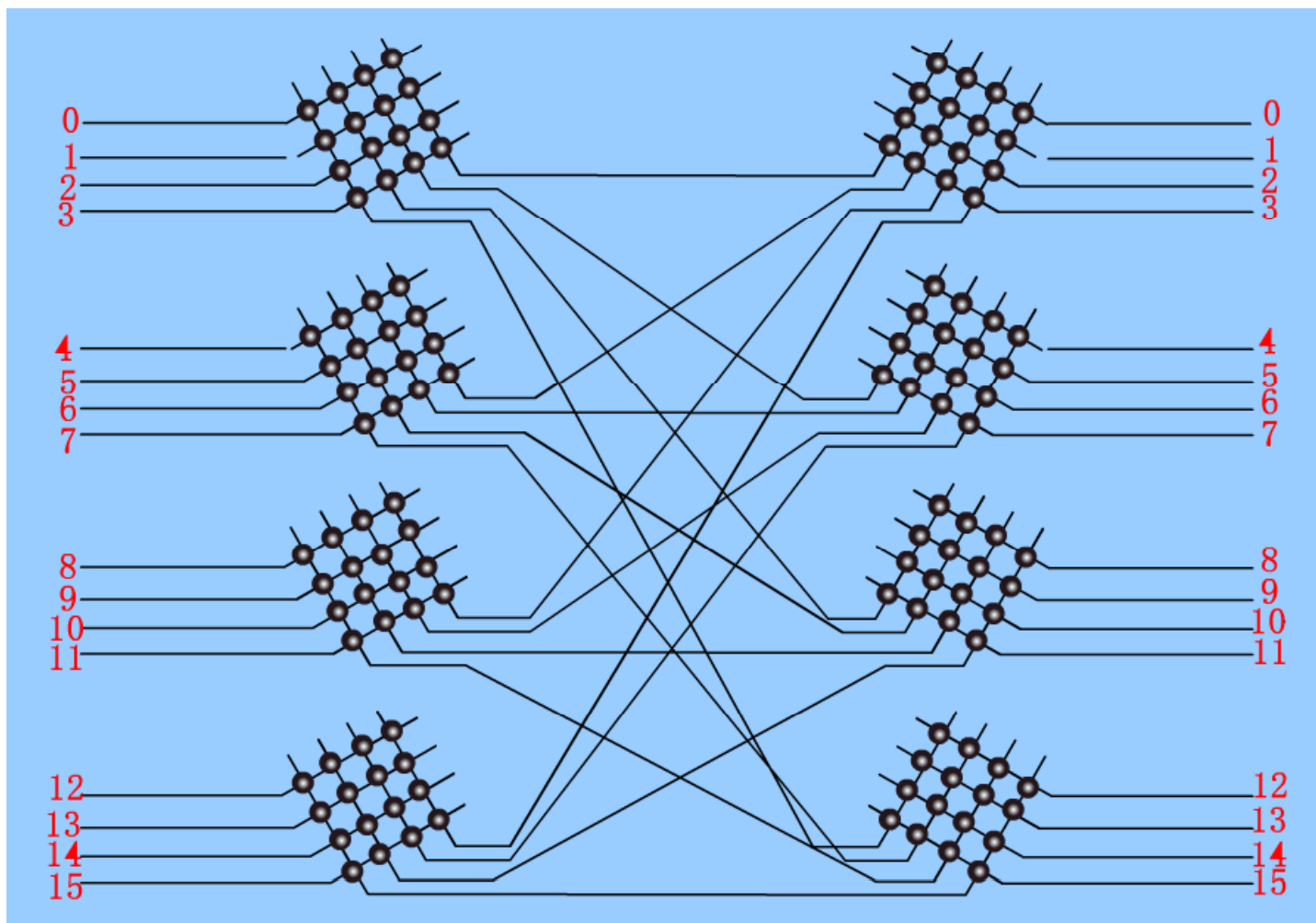
- Sun Microsystems 的 Ultra Enterprise 10000 (SMP, 1997)
 - 400 MHz UltraSparc 2 CPU's, 2 floats/cycle.
 - UMA (Uniform Memory Access)
 - 16 KB data cache (32 byte linesize), 4MB level 2 cache, 64 GB memory per processor
 - 网络: 10 GB/sec (聚合带宽), 600 ns时延



交叉开关的特点

- 交叉开关具有良好的带宽特性
- 非阻塞通信（Non-Blocking）：两个节点之间的通信，不会阻塞其他节点之间的通信。
- 代价不可缩放， $O(n^2)$ （ n 是交叉开关中的交叉点数）

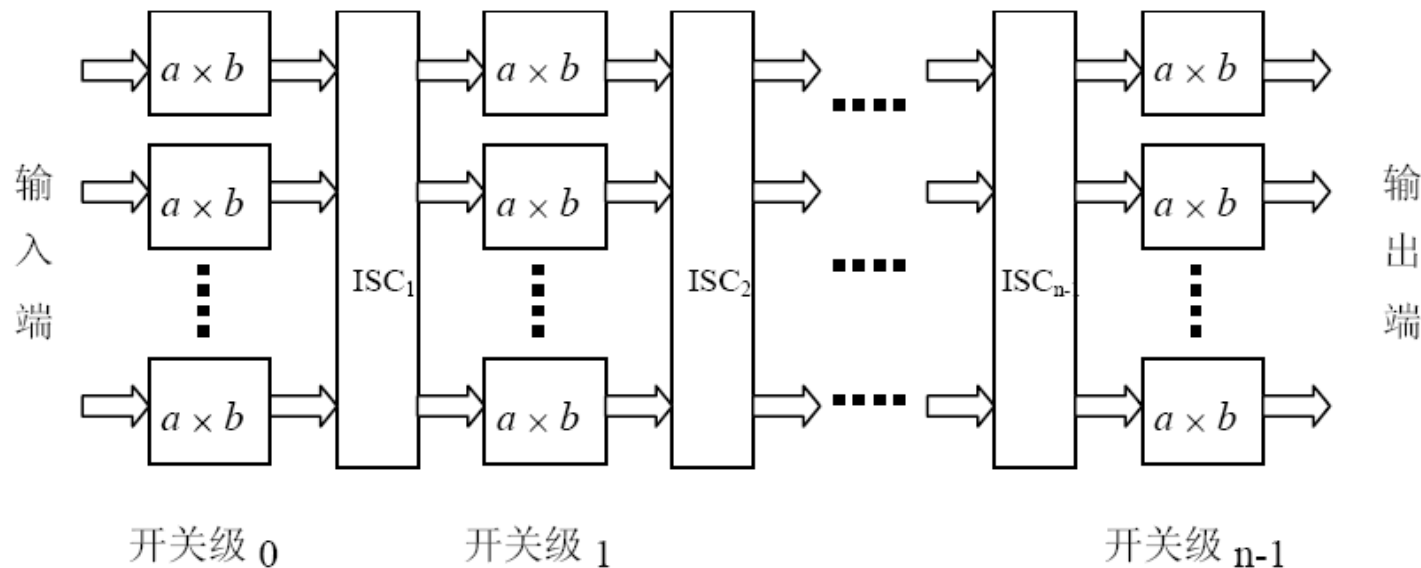
多极互联的交叉开关网络



- 用8个4×4的交叉开关构成二级16×16的交叉开关网络

多级互连网络 (1)

- 单级交叉开关级联起来形成多级互连网络MIN (Multistage Interconnection Network)

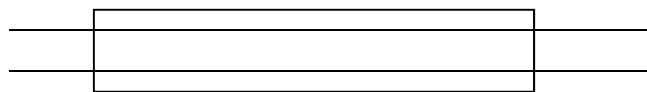


- N个输入, N个输出
- 度: 1, 直径: $\log N$, 网络规模: $N \log N$

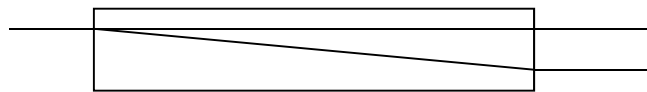
交换开关

- 交换开关模块

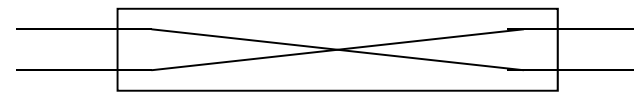
- 一个交换开关模块有 n 个输入和 n 个输出，每个输入可连接到任意输出端口，但只允许一对一或一对多的映射，不允许多对一的映射，因为这将发生输出冲突



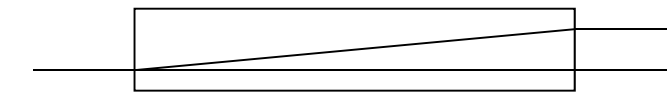
直通 (Straight)



上播 (Upper-broadcast)



交叉 (Exchange)



下播 (Lower-broadcast)

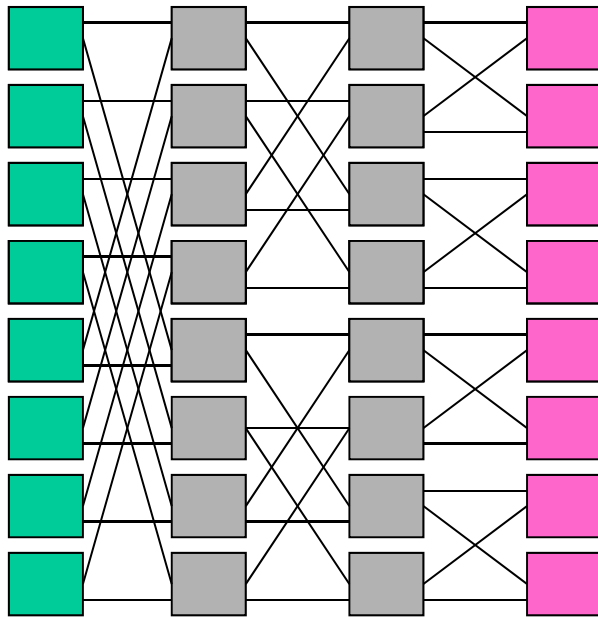
4种可能的交叉开关 (图1.12)

级间互连（ISC: Inter-Stage Connection）

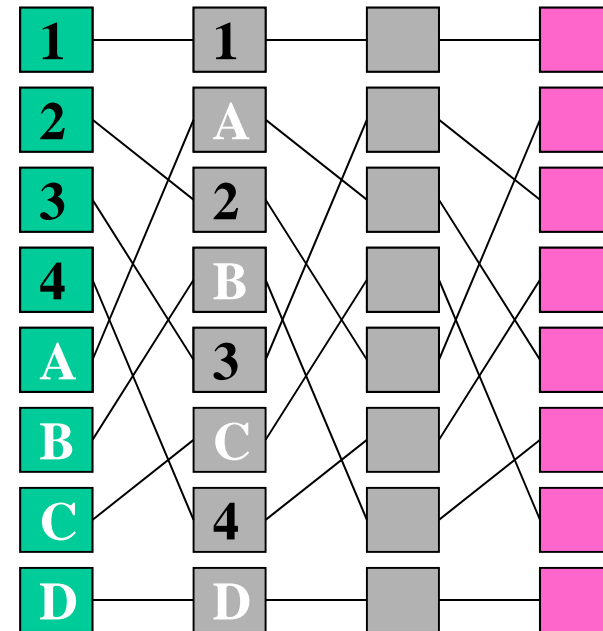
- 级间互连（ISC: Inter-Stage Connection）：
 - 每一级输入与输出之间连接，输出作为交换开关的输入连到下一级
 - 均匀洗牌（Perfect-Shuffle）、蝶式（Butterfly）、交叉开关等

MIN的例子

- 蝶式 (Butterfly)

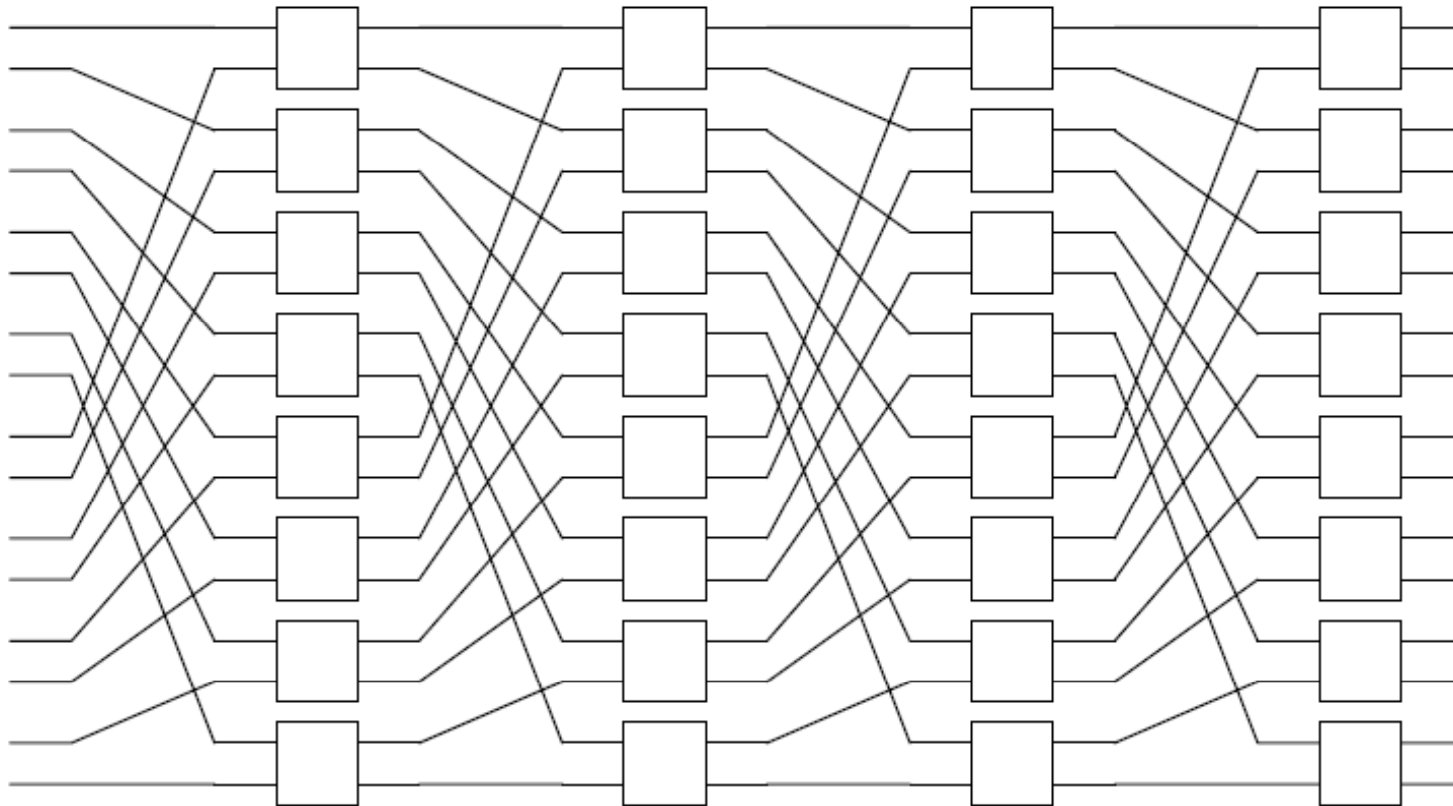


- 均匀洗牌 (Shuffle)



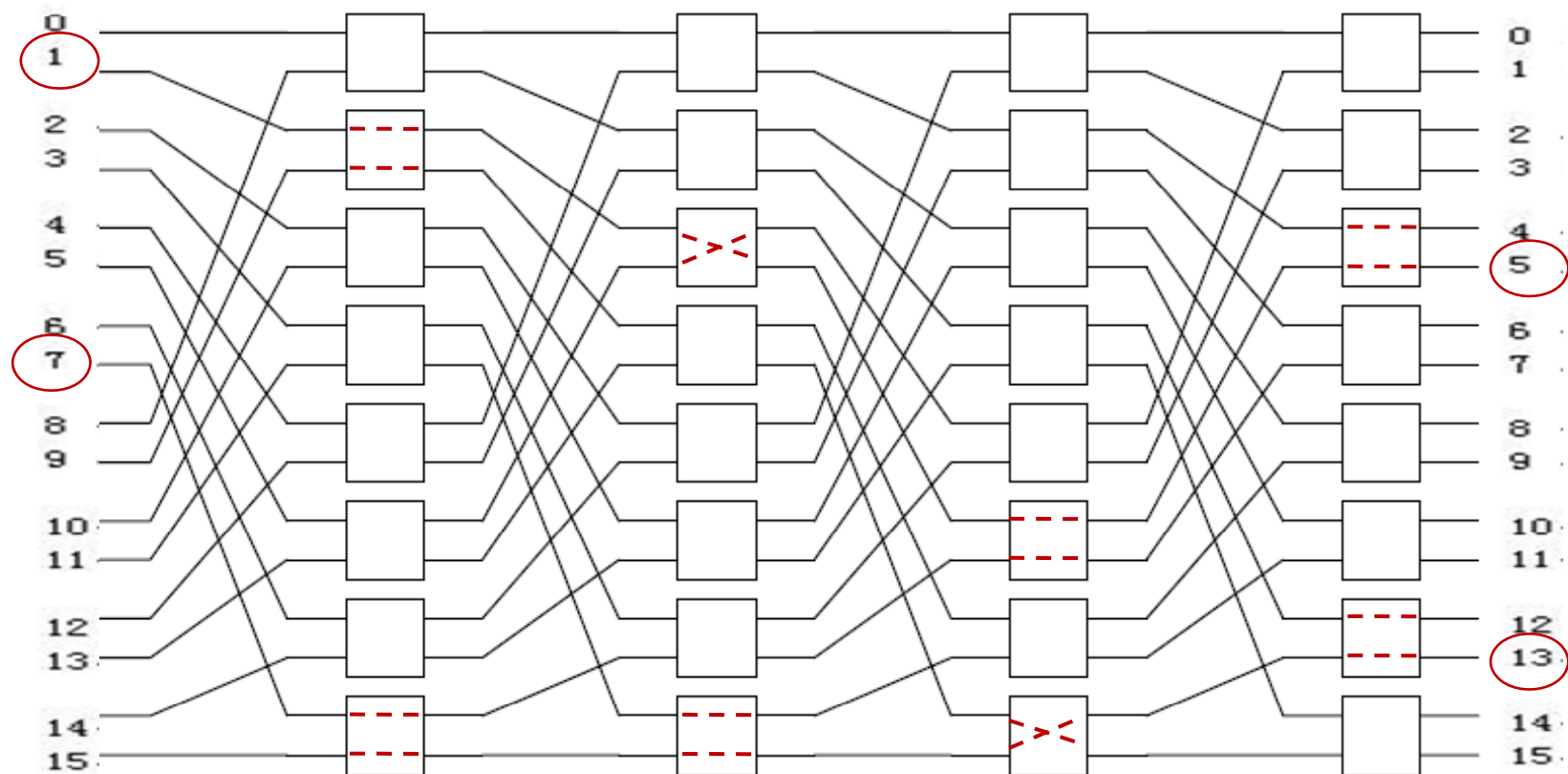
应用例子： Ω 网络

16×16 Ω 网络

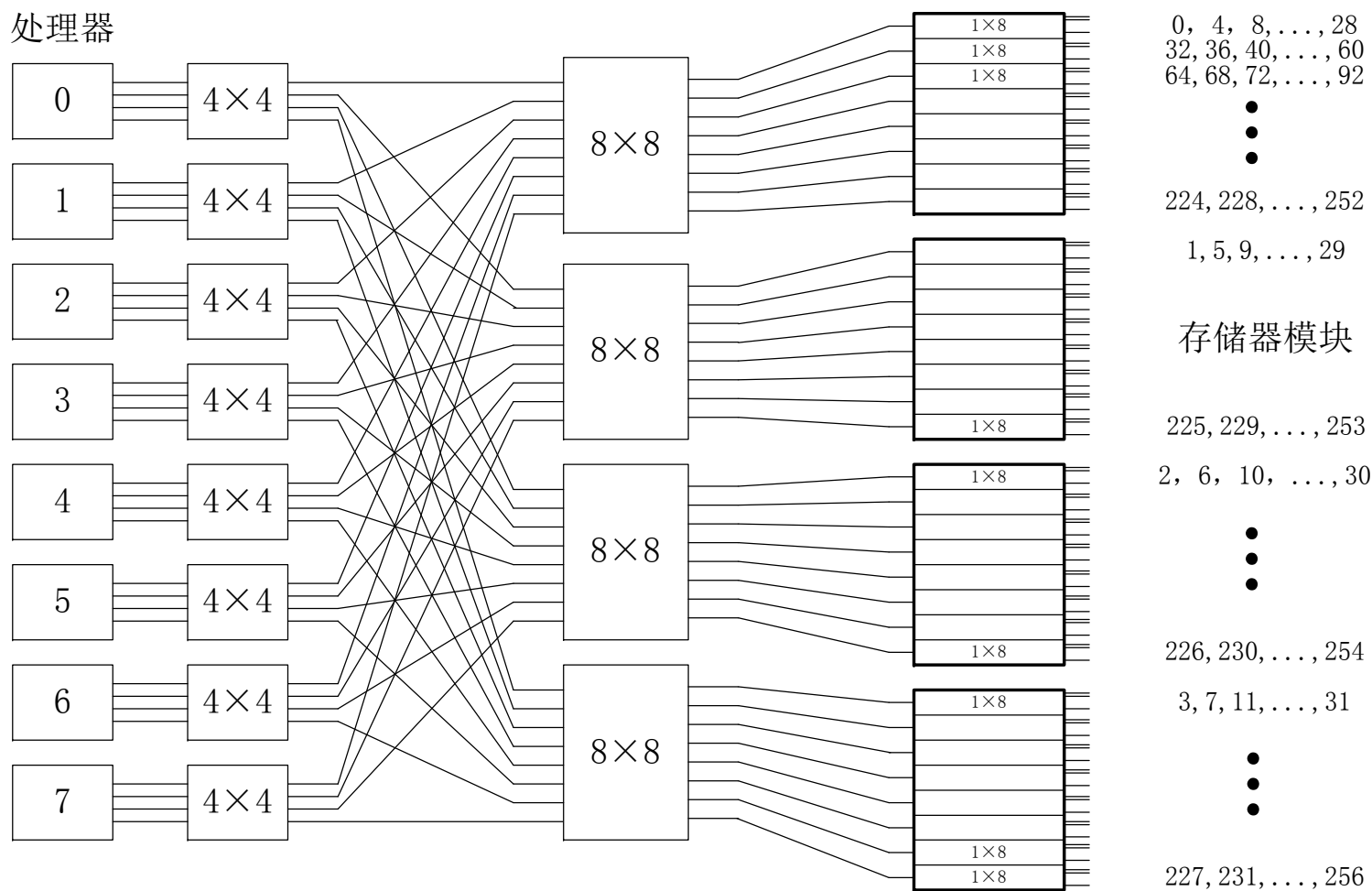


应用例子: **IBM SP2, IBM option white, IBM Blue Pacific**

Ω 网络连接示例



应用例子：Cray Y-MP多级网络



Cray Y-MP多级网络

动态互连网络比较

网络特性	每个处理器带宽	硬件复杂度	阻塞	实际机器：聚集带宽
总线	$O(wf/n) \sim O(wf)$	$O(n+w)$	是	SunFire服务器中的 Gigaplane 总线: 2.67GB/s
多级互联	$O(wf)$	$O((n \log_k n)w)$	是	IBM SP2中的512节点的HPS: 10.24GB/s
交叉开关	$O(wf)$	$O(n^2w)$	否	Digital 的千兆开关: 3.4GB/s

n : 节点规模 w : 数据宽度 f : 时钟频率

讨论

网络名称	复杂度	网络性能	扩展性	容错性能
线性阵列				
环形				
2-D网孔				
二叉树				
胖树				
星形				
超立方				
总线				
交叉开关				
MIN				

主要内容

- 系统互联
- 静态互联网络
- 动态互连网络
- 标准互联网络

标准互联网络

- 开放、高带宽、低延迟、高可靠、扩展性好的交换网络和技术
- 主要应用于集群系统
- 标准互联网络：
 - HiPPI
 - Scalable Coherent Interface (SCI)
 - Myrinet
 - Ethernet
 - QsNet
 - Infiniband

集群中的互连技术（HiPPI）

- HiPPI是Los Alamos国家实验室于1987年提出的一个标准，其目的是试图统一来自不同产商生产的所有大型机和超级计算机的接口。HiPPI作为短距离的系统到系统以及系统到外设连接的高速I/O通道
- 1993年，ANSI X3T9.3委员会认可了HiPPI标准，它覆盖了物理和数据链路层，但在这两层之上的任何规定却取决于用户
- HiPPI是单工的点到点的数据传输接口，其速率可达800Mbps到1.6Gbps。超级HiPPI（Super HiPPI）技术能提供6.4Gbps速率
- SGI公司和Los Alamos国家实验室都开发了用来构筑速率高达25.6Gbps的HiPPI交换开关的HiPPI技术
- 应用例子：SGI Power Challenge服务器、IBM 390主机、Cray Y/MP、C90和T3D/T3E等系统

使用HiPPI通道和开关的局域网

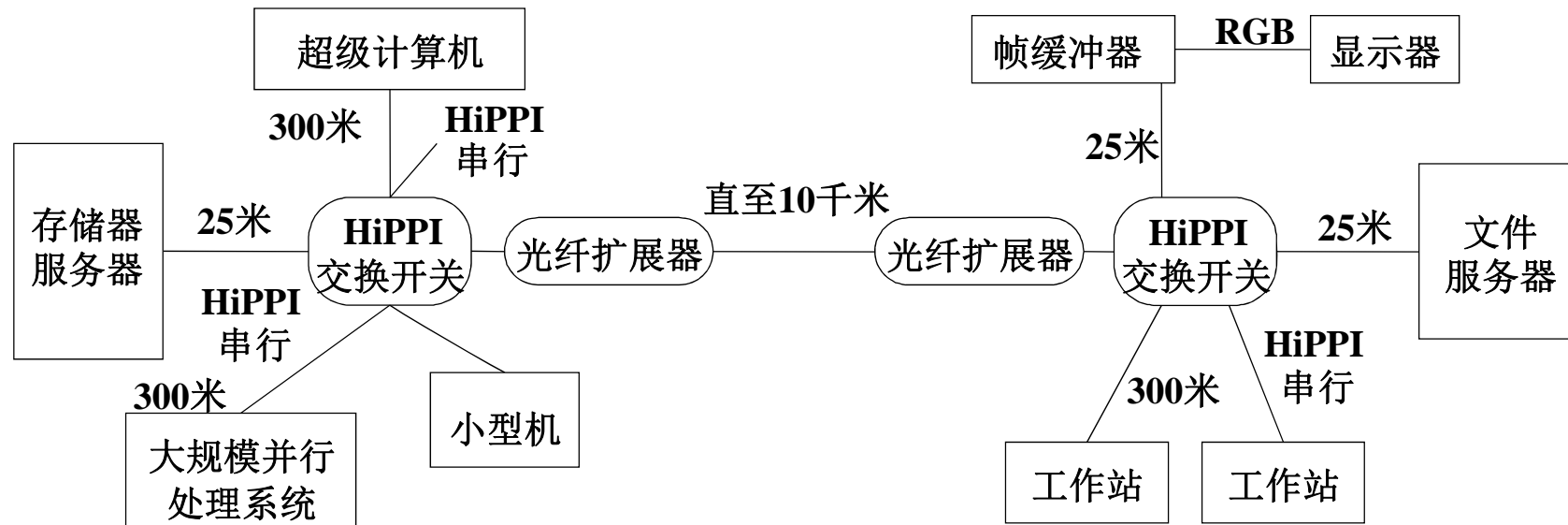
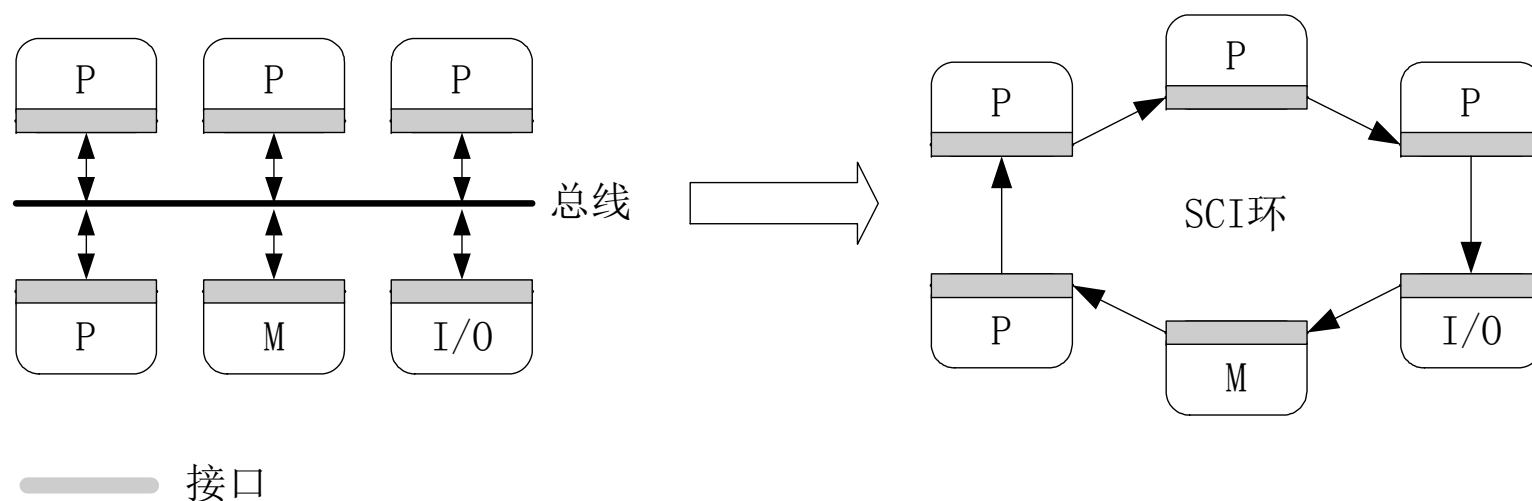


图1.16

可扩展一致性接口SCI

- SCI (Scalable Coherence Interface)
 - 既能保持总线的优点又具有传统网络空间的可扩展性
 - 通常的底板总线扩展成全双工、点到点的互连结构
 - 提供分布共享存储器一致的高速缓存映象
 - SCI被设计用来提供低时延（小于1微秒）和高带宽（高至8GB/s）的点到点互连。
 - 可连接多至64K个节点。



SCI (2)

- 关于SCI标准的 ANSI/IEEE 1596-1992定义了一个点到点的接口和一系列包协议 (packet protocols)
- Dolphin (<http://www.dolphinics.com>) SCI 产品:
 - PCI网卡: 32-bit/33Mhz 到64-bit/66Mhz
 - 最高速率: 667MB/s (或 5.33Gb/s)
 - 时延: 1.46 μ s



集群中的互连技术 (Myrinet)

- Myrinet是由Myricom(<http://www.myri.com>)公司设计的千兆位包交换网络，用于构建计算机集群
 - 基于加州理工学院开发的多计算机和VLSI技术以及在南加州大学开发的ATOMIC/LAN技术
 - 能适应任意拓扑结构，不必限定为开关网孔或任何规则的节点
- Myrinet是ANSI (American National Standards Institute) 标准
 - ANSI/VITA 26-1998

Myrinet (2)

- Myrinet在数据链路层具有可变长的包格式，对每条链路施行流控制和错误控制，并使用切通选路法以及定制的可编程的主机接口。在物理层上，Myrinet网使用全双工SAN链路，最长可达3米，峰值速率为 $(1.98+1.98)$ Gbps
- GM通信软件：实现一种高速度带宽低延迟,主机开销极少的通信方式，充分利用Myrinet在硬件架构上的特性，大大减轻了主机的通信开销，每个数据包括只耗费主机大约 $1\mu\text{s}$ 的处理时间
- 在2004年7月的TOP 500中有187台（37.4%）的系统采用Myrinet
- 在2005年7月的TOP 500中有141台（28.2%）的系统采用Myrinet
- 例子：曙光4000A



Myrinet连接的LAN/Cluster

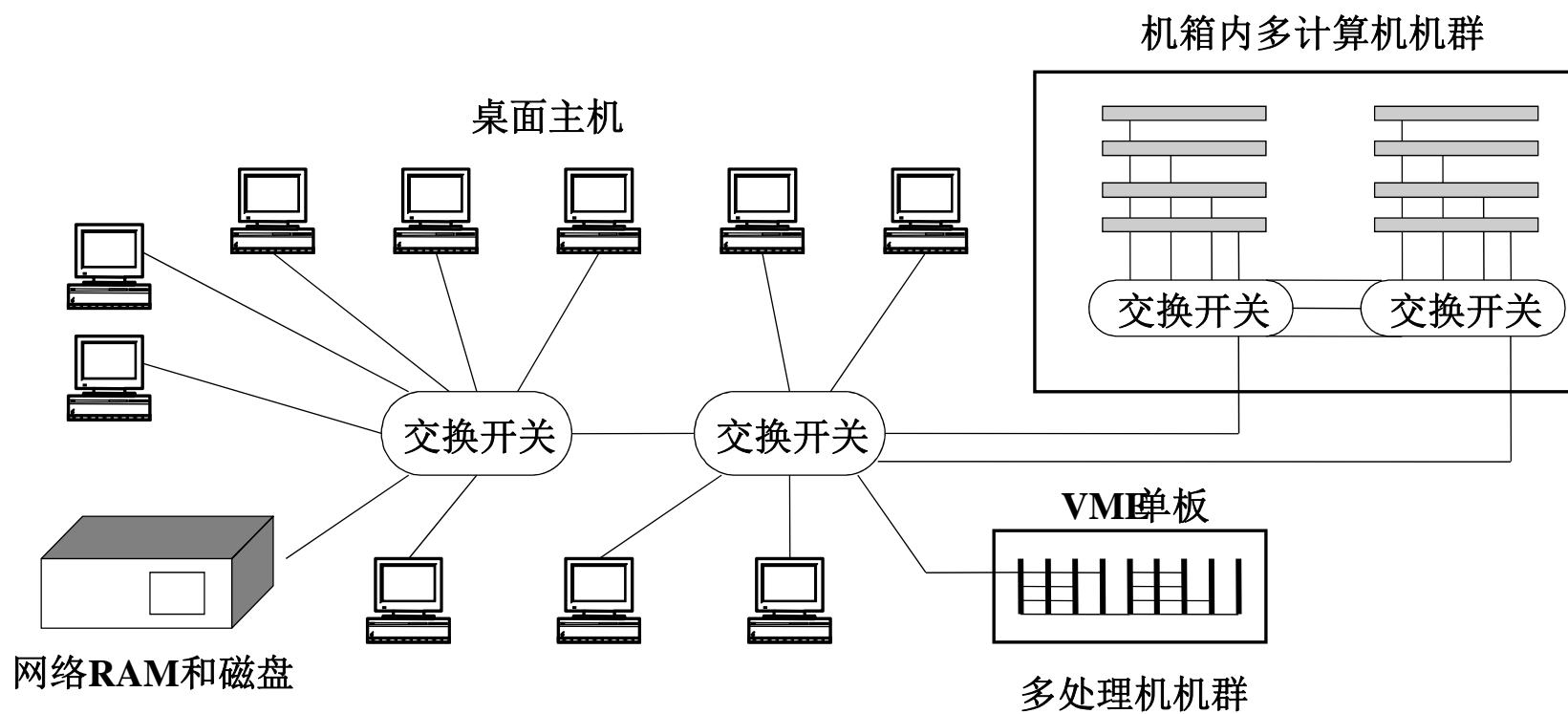


图1.15

以太网

代别 类型		以太网 10BaseT	快速以太网 100BaseT	千兆位以太网 1GB	万兆位以太网 1GB
引入年代		1982	1994	1997	2002
速度（带宽）		10Mb/s	100Mb/s	1Gb/s	10Gb/s
最大 距离	UTR (非屏蔽双扭对)	100m	100m	25—100m	25—100m
	STP (屏蔽双扭对) 同轴电缆	500m	100m	25—100m	25—100m
	多模光纤	2Km	412m (半双工) 2Km (全双工)	500m	300m
	单模光纤	25Km	20Km	3Km	10—40Km

万兆以太网

- IEEE于2002年6月批准万兆位以太网(10gige)802.3ae标准
- 万兆以太网是一种只采用全双工与光纤的技术，其技术特色首先表现在物理层面上：
 - 物理层（PHY）和OSI模型的第一层(物理层)一致，它负责建立传输介质（光纤或铜线）和MAC层的连接
 - 把PHY进一步划分为物理介质关联层（PMD）和物理代码子层（PCS）。光学转换器属于PMD层。PCS层由信息的编码方式(如64B/66B)、串行或多路复用等功能组成
- 万兆标准意味着以太网将具有更高的带宽（10G）和更远的传输距离(最长传输距离可达40公里)。
- 万兆以太网技术提供了更多的更新功能，如QoS，能更好的满足网络安全、服务质量、链路保护等多个方面需求。

Quadrics Interconnect (QsNet)

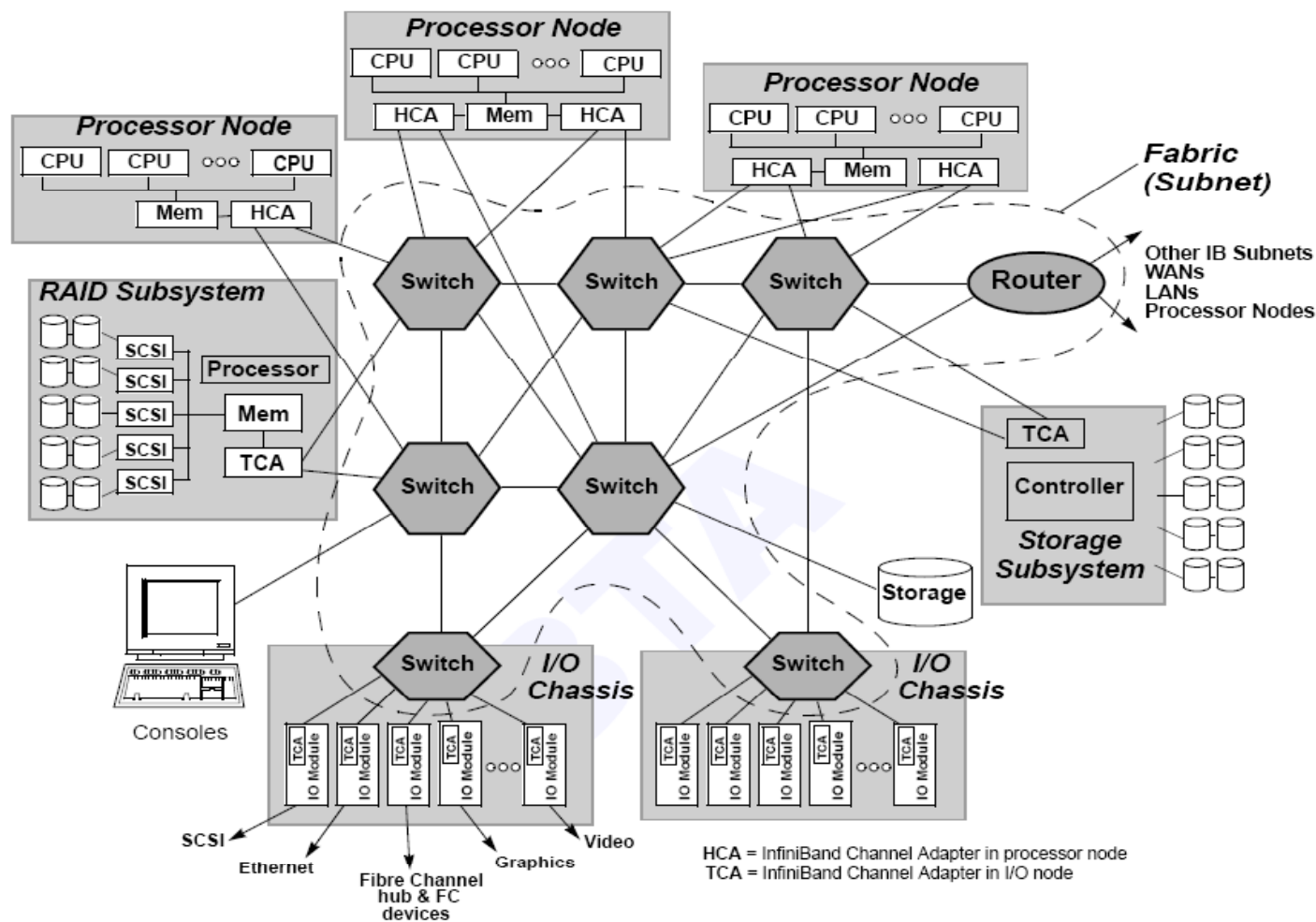
- QsNet和QsNetII 是Quadrics公司（www.quadrics.com）私有的系统互连技术，源自于为SMP系统开发的技术。
- 两个基本硬件
 - QsNet和QsNetII交换机（Elite）的基本组成部分是8端口的矩阵式交换芯片，端口速率为3.2 Gbps和10.6 Gbps
 - QsNet Elan通信接口（Network Adapter）通过将单个节点地址空间集成，提供一个全局虚拟地址空间



InfiniBand (IB)

- IB是由IB行业协会所倡导的，IBTA (Infiniband Trade Association)于1999年成立，协会主要成员有：康柏、戴尔、惠普、IBM、Intel、微软和Sun
 - <http://www.infinibandta.org>
- Infiniband是一种可满足存储区域网、高端计算集群及局域网需求的全新互连标准
- InfiniBand技术通过一种交换式通信组织（Switched Communications Fabric）提供了较局部总线技术更高的性能，它通过硬件提供了可靠的传输层级的点到点连接，并在线路上支持消息传递和内存映像技术

InfiniBand体系结构



InfiniBand技术

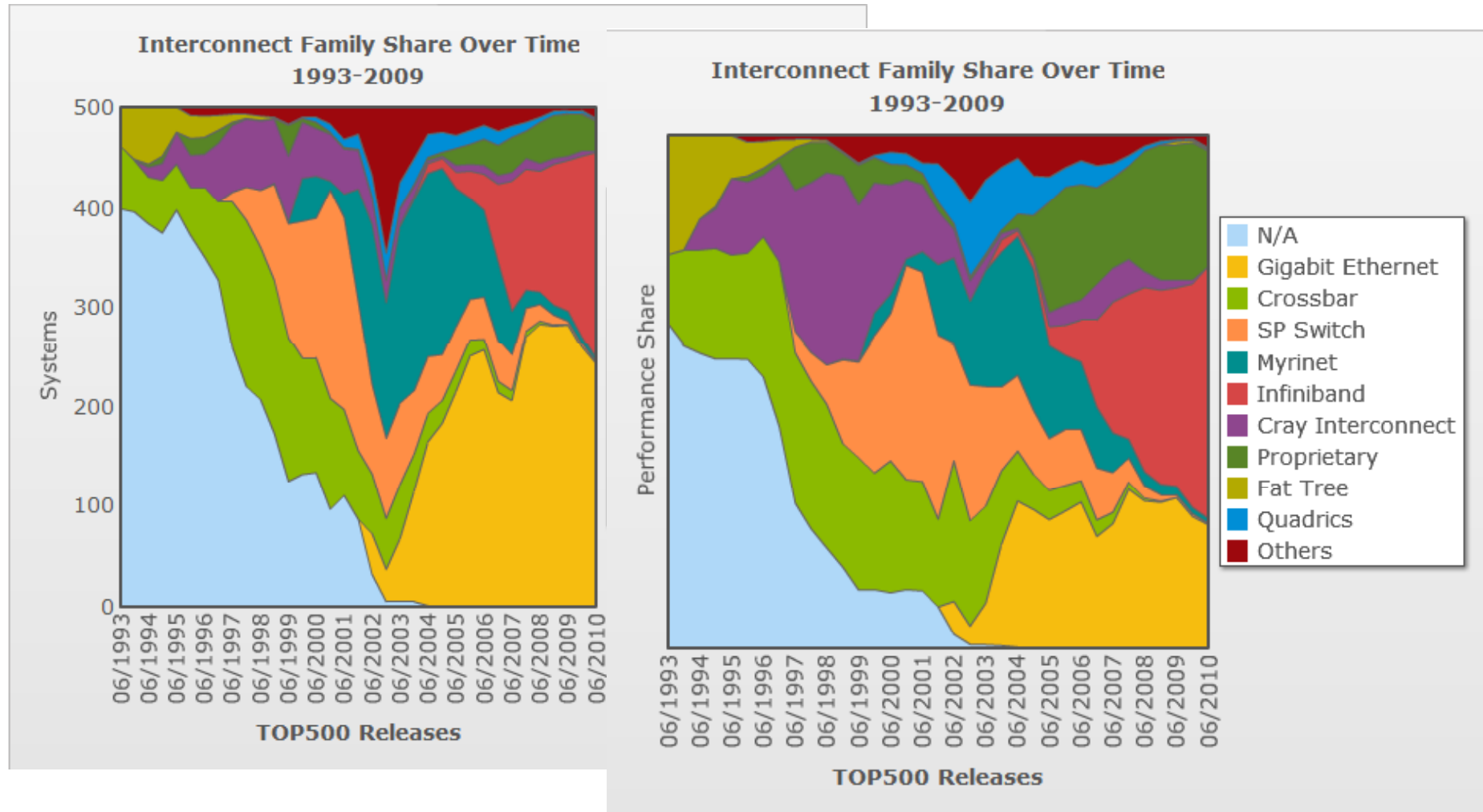
- InfiniBand技术通过连接主机通道适配器HCA(Host Channel Adapters)、目标通道适配器TCA(Target Channel Adapters)、交换机以及路由器来工作
- 拓扑
 - 不规则
 - 规则：胖树（Fat Tree）
- InfiniBand 物理层基于IEEE 802.3.z标准，与10G以太网一样
- 三种连接速率(1x, 4x, 12x)，全双工，是基本传输速率2.5Gb/s的倍数，即支持速率是2.5Gb/s、10Gb/s和30Gb/s

标准互联网络的比较

网络	接口界面	带宽 (MB/s)	MPI延时 (μ s)	备注
Gigabit Ethernet	PCI	125	80	
10G Ethernet	PCI Express	1215	2.2	Myri-10G
Myrinet 2000	PCI-X	247	2.6~3.2	Myrinet 2000
SCI	PCI Express	500~700	1.4	Dolphin D350
QsNet	PCI-X	900	1.6	QsNetII QM500
InfiniBand	PCI Express	1400	1.29	Qlogic QLE7240

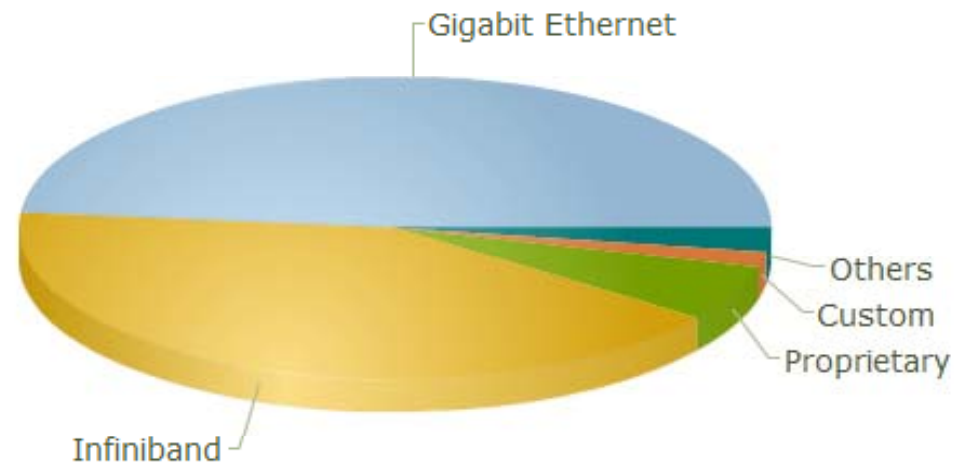
数据来源: myricom, Dolphin, Quadrics, Qlogic 公司网站

互联技术在Top500上的比较



Top 500 (Interconnect Family)

Interconnect Family / Systems
June 2010



Gigabit Ethernet: 48.44%, Infiniband: 41.4%, Proprietary: 6% , Custom: 1.6%

互联技术在Top500上的比较

Interconnect Family	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Processor Sum
Myrinet	4	0.80 %	315041	421947	45864
Quadrics	2	0.40 %	81830	98355	12848
Gigabit Ethernet	242	48.40 %	7776035	14868247	1413300
Infiniband	207	41.40 %	15895656	22572675	1798588
Mixed	1	0.20 %	66567	82944	13824
NUMalink	2	0.40 %	107961	121241	18944
SP Switch	2	0.40 %	101179	123911	16304
Proprietary	30	6.00 %	7315515	9212550	1712867
Cray Interconnect	1	0.20 %	35200	42598	8192
Fat Tree	1	0.20 %	122400	131072	1280
Custom	8	1.60 %	617300	794955	89450
Totals	500	100%	32434683.70	48470495.53	5131461

Interconnect Family (2010.6)

课程小结

- 网络环境
 - Intra-node Interconnections(Buses , Switches)
 - Inter-node Interconnections (SAN)
 - Inter-system Interconnections(LAN , MAN , WAN)
- 互联网络结构
 - Static-Connection Networks(LA,RC,MC,TC,HC,CCC)
 - Dynamic-Connection Networks (Buses, Crossbar, MIN)
 - Standard Networks (Myrinet, Ethernet, Infiniband)

推荐读物及网站

- 《并行计算》 1.2节
- 中国科大精品课程—并行计算
 - <http://www.bb.ustc.edu.cn/jpkc/guojia/bxjs/index.htm>

下一讲

- 并行计算机体系结构（2）
 - 《并行计算—结构、算法、编程》1.2，第2章