Interconnections

Communication Model

Message Passing 或是 MPI

可以是 unicast，broadcast 或是 multicast

message passing 相比 shared memory

- message passing
  - 内存私有
  - 显式的 send/receive
  - message 中有数据和同步信息
  - 需要知道信息的 src 或 dest
- shared memory
  - 内存共享
  - 隐式的 load/store
  - 隐式的同步
  - 不需要知道信息的 src 和 dest

可以用 shared memory 实现 message passing：软件上将 load/store 变为 send/receive，硬件上将总线上的数据交流变成信息的发送（核与核，核与内存通过 switch 相连）

可以用 message passing 实现 shared memory：使用 queue 直接在核之间通过 load/store 和锁实现数据传输

Switching

总线的特点是共享通信媒介，不需要路由，便于监听，而交换机网络是点对点连接各节点，需要路由

有多种 switching 方法

- circuit switched
- store and forward
- cut-through
- wormhole

Topology

节点网络的拓扑也可以有很多种

- 1-D
  - 总线连接所有节点
  - 节点首尾相接成环
- 2-D
  - mesh
  - 节点组织成二维的环（每一行/一列首尾相接成环）
  - full mesh，每个节点间都直接相连
  - hyper cube，立方体
  - omega network
  - fat tree，交换机组成二叉树，叶节点是主机
  - clos network，同 fat tree，但是增加了冗余，每个非叶非根的交换机都有两个父节点

topology 有各种参数

- routing distance：两点间需要多少 hop
- diameter：最大的 routing distance
- average distance
- minimum bisection bandwidth：最小的将网络分成两个不相交的 set 的 cut 的带宽
- degree of a router

Network	# Nodes	Router Deg.	Diameter	Bisection BW	# Links
1D Torus	N	2	$\lfloor N/2 \rfloor$	2	N
2D Mesh	$\sqrt{N} \times \sqrt{N}$	4	$2\sqrt{N}-2$	$\sqrt{N}$	$2(N-\sqrt{N})$
2D Torus	$\sqrt{N} \times \sqrt{N}$	4	$2\lfloor \sqrt{N}/2 \rfloor$	$2\sqrt{N}$	$2N$
Full Mesh	N	$N-1$	1	$N^2/4$	$N(N-1)/2$
Binary Tree	N	3	$2(\lceil \log_2 N \rceil - 1)$	1	N-1
Hyper Cube	$N=2^n$	n	n	N/2	$nN/2$

Network Performance

bandwidth：给定时间内能传输的数据总量

latency：信息从发送到接受需要的时间

带宽上升可以减小延迟：减少拥塞

延迟会限制带宽，如 round trip 的形式

Routing

routing 一般是根据网络情况动态适应的

- link state
- distance vector

flow control 可以是局部或是端到端的

一个流控例子：token bucket