Chapter 18 Parallel Processing

Taxonomy of Computer Architectures (Flynn's)

instruction stream – sequence of instructions that are executed in a processing unit

data stream – sequence of operands that are manipulated in a processor

SISD - sequential (serial) computer

MISD - pipelined vector processor?

SIMD - array processor, hardware synchronization

MIMD - multiprocessor, most general, software synchronization

- Shared-memory multiprocessor (tightly-coupled)
  - shared memory through central switching mechanism
- Distributed-memory multicomputer (loosely-coupled)
  - shared memory is formed by combining local memories.
  - message passing

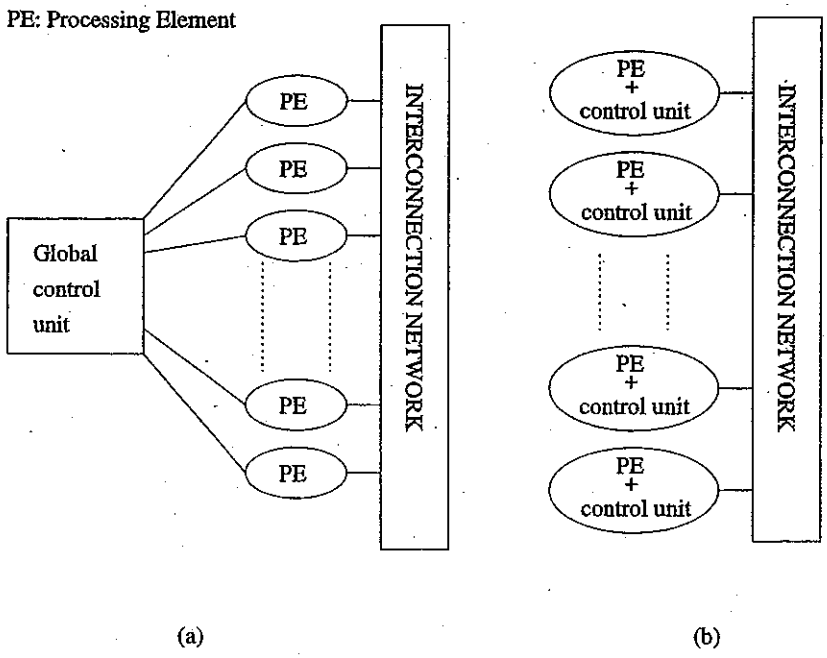
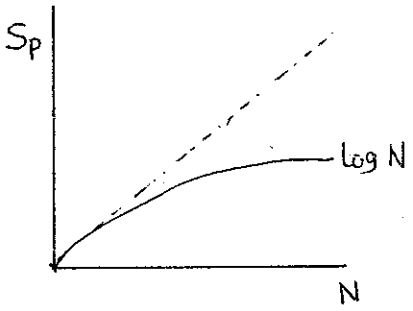


Figure A typical SIMD architecture (a) and a typical MIMD architecture (b).

Arguments against parallel computing

- **Minsky's conjecture** (1973)

  Speedup $\propto$ log N



- **Amdahl's law (1967)**

  The maximum speedup of a parallel processor is limited by the fraction of serial instructions it has to execute.

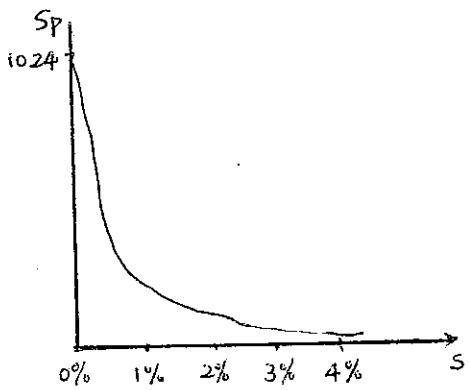
$$\textbf{Speedup} = \frac{S + P}{(S + \frac{P}{N})} = \frac{1}{(S + \frac{P}{N})}$$

Ex $S = \frac{1}{10}$ $N = 100$ $S_p \fallingdotseq 9.17$

$S = \frac{1}{100}$ $N = 100$ $S_p \fallingdotseq 50.25$

$S = \frac{1}{100}$ $N = 1000$ $S_p \fallingdotseq 91$

$S = \frac{2}{3}$ $N = 10$ $S_p \fallingdotseq 1.42$



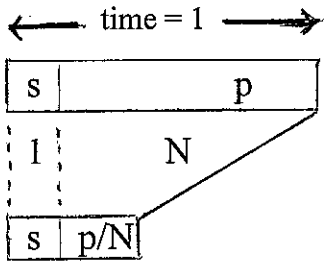
Gustafson-Barsis Law (1988)

- s and N (number of processor) are not independent of one another
- run time, not problem size, is fixed

s: vector start-up, program loading, serial bottleneck, I/O :fixed

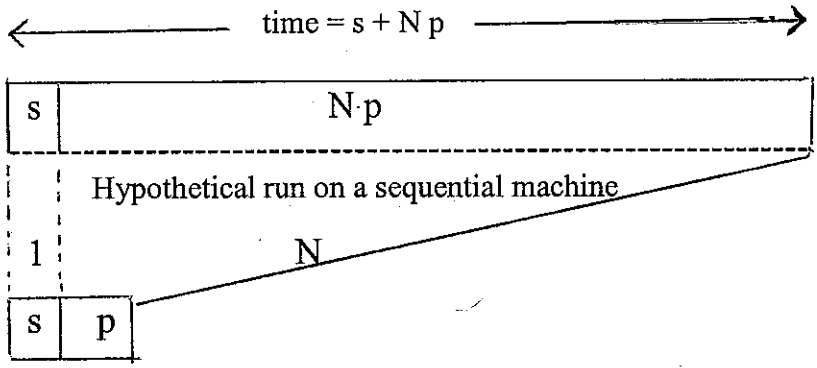
p: scale up :variable

- Fixed-size model:



$$Sp = \frac{1}{(s + p/N)}$$

- Scaled-size model:



←time=1→

$$Sp = (s + N\,p) / (s+p) = s + N \cdot p$$

Ex. $s = 2/3$, $N = 10$

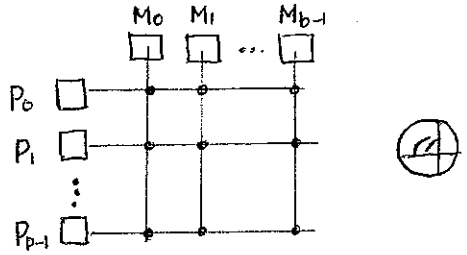
$Sp = 2/3 + 1/3 \cdot 10 = 4$

Ref. Gustafson, Reevaluating Amdahl's Law
Comm. of the ACM, 31(3), May 1988, pp. 532-533

Processor Interconnection

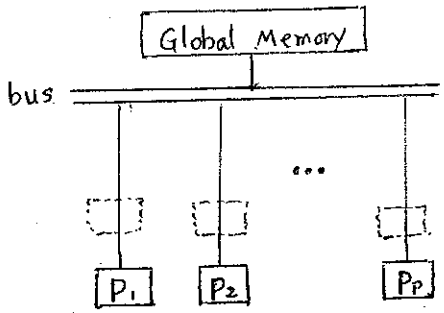
Dynamic Interconnection Network

. Crossbar Switching Network



CMU C.mmp
Cray Y/MP
Fujitsu VPP500

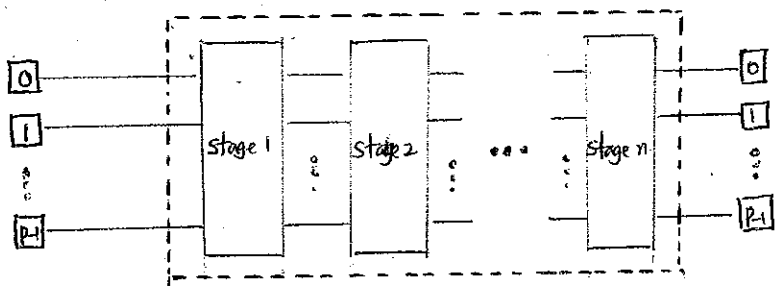
. Bus-based

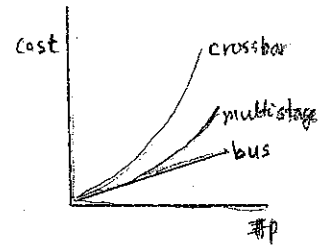


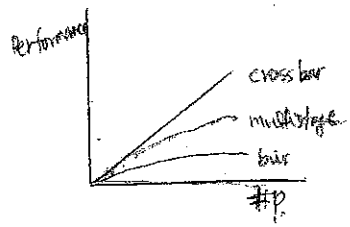
Sequent Symmetry

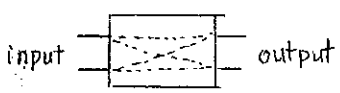
Note. Performance Vs. cost - tradeoff

. Multistage Interconnection Network

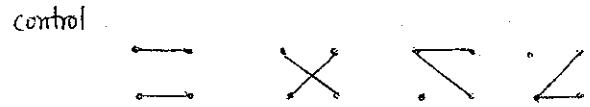








Switching box



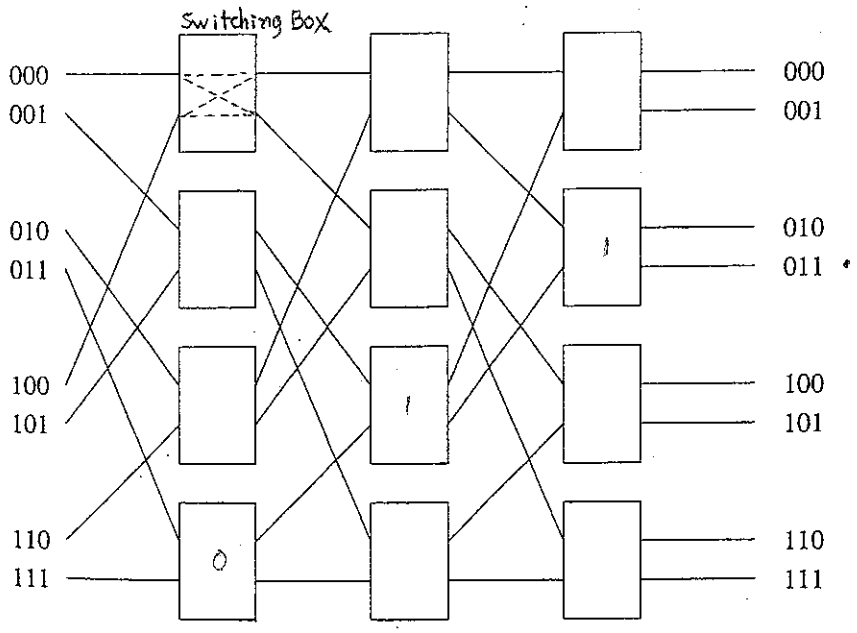


Figure A complete omega network connecting eight inputs and eight outputs .

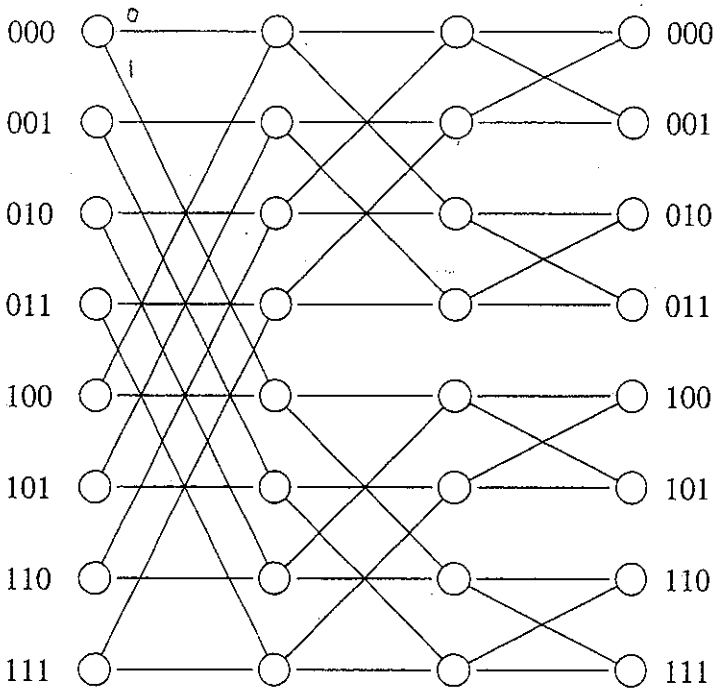
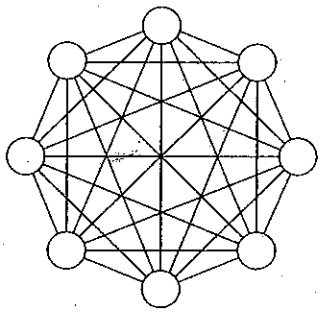


Figure An eight-processor butterfly network.

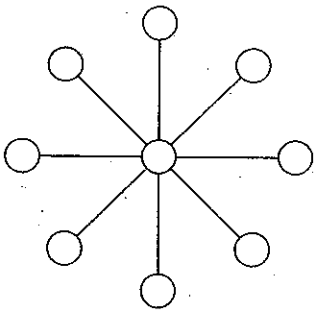
Static Interconnection Network

- **Completely-connected netwok**

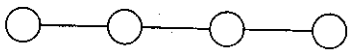


$\frac{n(n-1)}{2}$

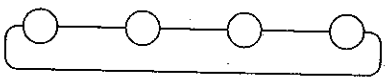
- **Star-connected**



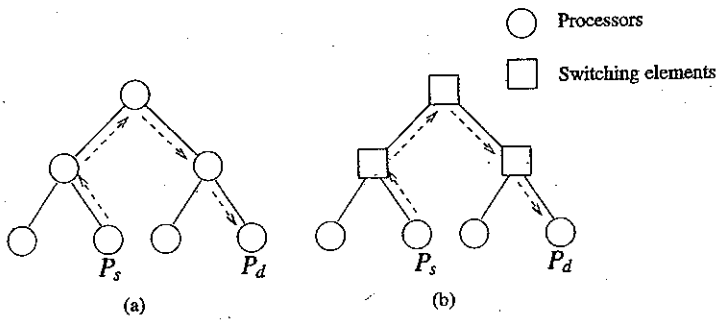
- **Linear** (1-D)



- **Ring**

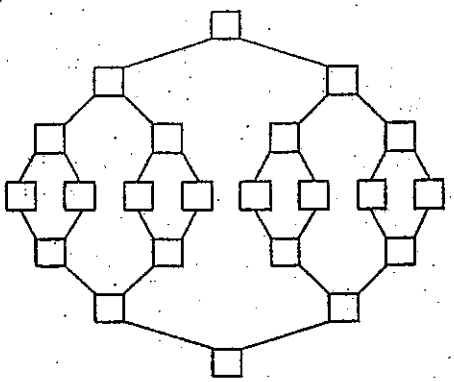


- **Tree Network**

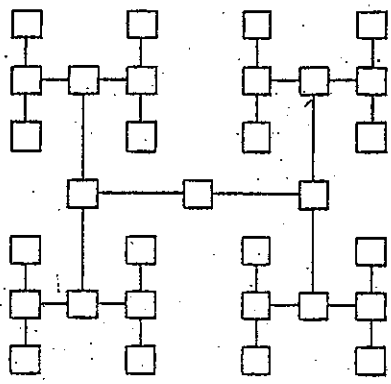


Processor communication : SEND/RECEIVE

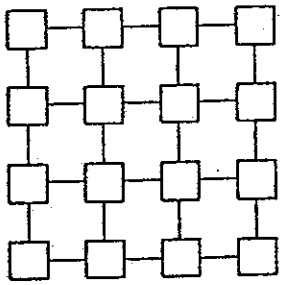
Summation in log N steps

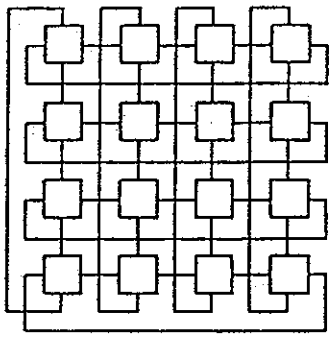


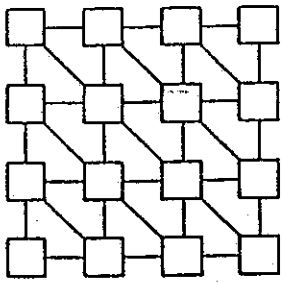
Layout

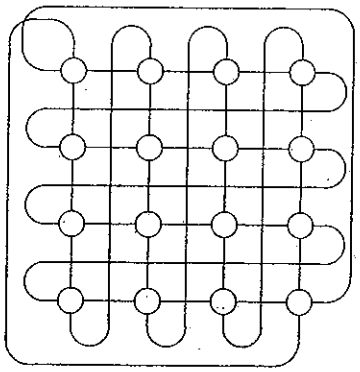


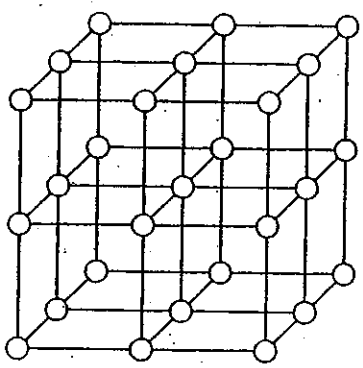
- Mesh



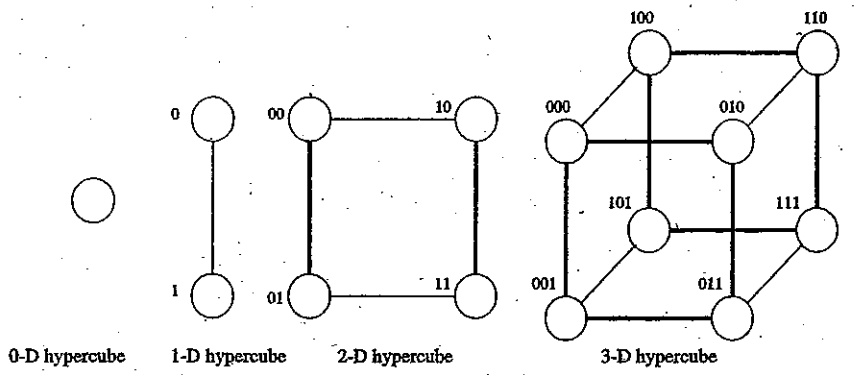




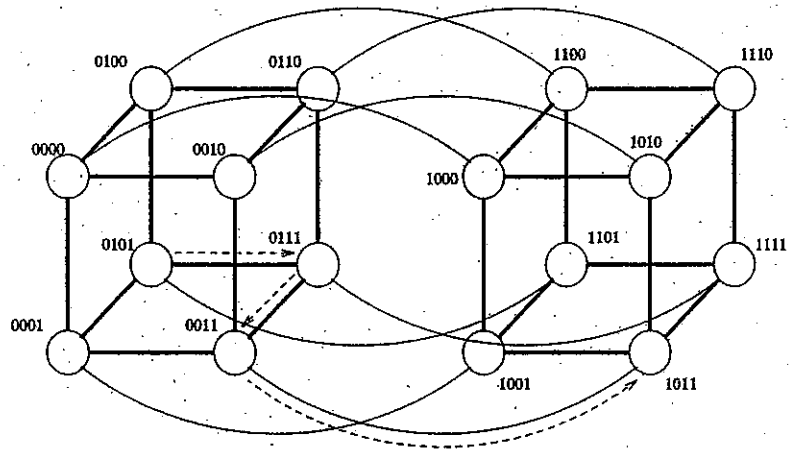




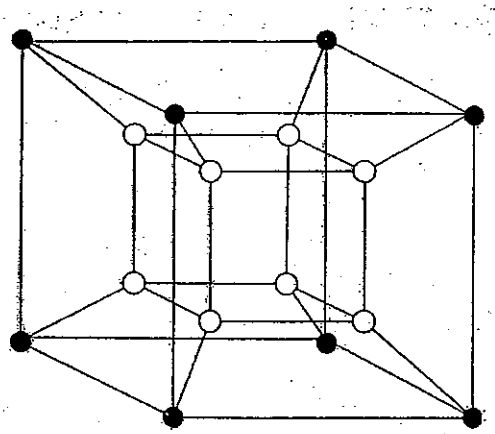
• Hypercube



0-D hypercube 1-D hypercube 2-D hypercube 3-D hypercube



4-D hypercube



Performance Criteria

(1) Latency
(2) Bandwidth
(3) Connectivity
(4) Hardware cost
(5) Reliability
(6) Functionality

Goodness Measures

- maximum degree
- diameter
- bisection-width

Characteristics of various interconnection networks

Network	#nodes	#edges	Max.node.deg	Diameter	Bisection-width
Linear	n	$n-1$	2	$n-1$	1
Ring	n	n	2	$n/2$	2
Star	n	$n-1$	$n-1$	2	-
2-D Mesh	$n \times n$				
2-D Torus	$n \times n$				
3-D Mesh	$n \times n \times n$				
3-D Torus	$n \times n \times n$				
Tree	$n = 2^k - 1$				
Double tree					
Hyper tree					
Pyramid	$n = (4k^2-1)/3$				
Butterfly	$n=(k+1)2^k$				
Shuffle-xchg	$n = 2^k$				
Hypercube	$n = 2^k$				
Completely connected	n	$\frac{n(n-1)}{2}$	$n-1$	1	$n^2/4$