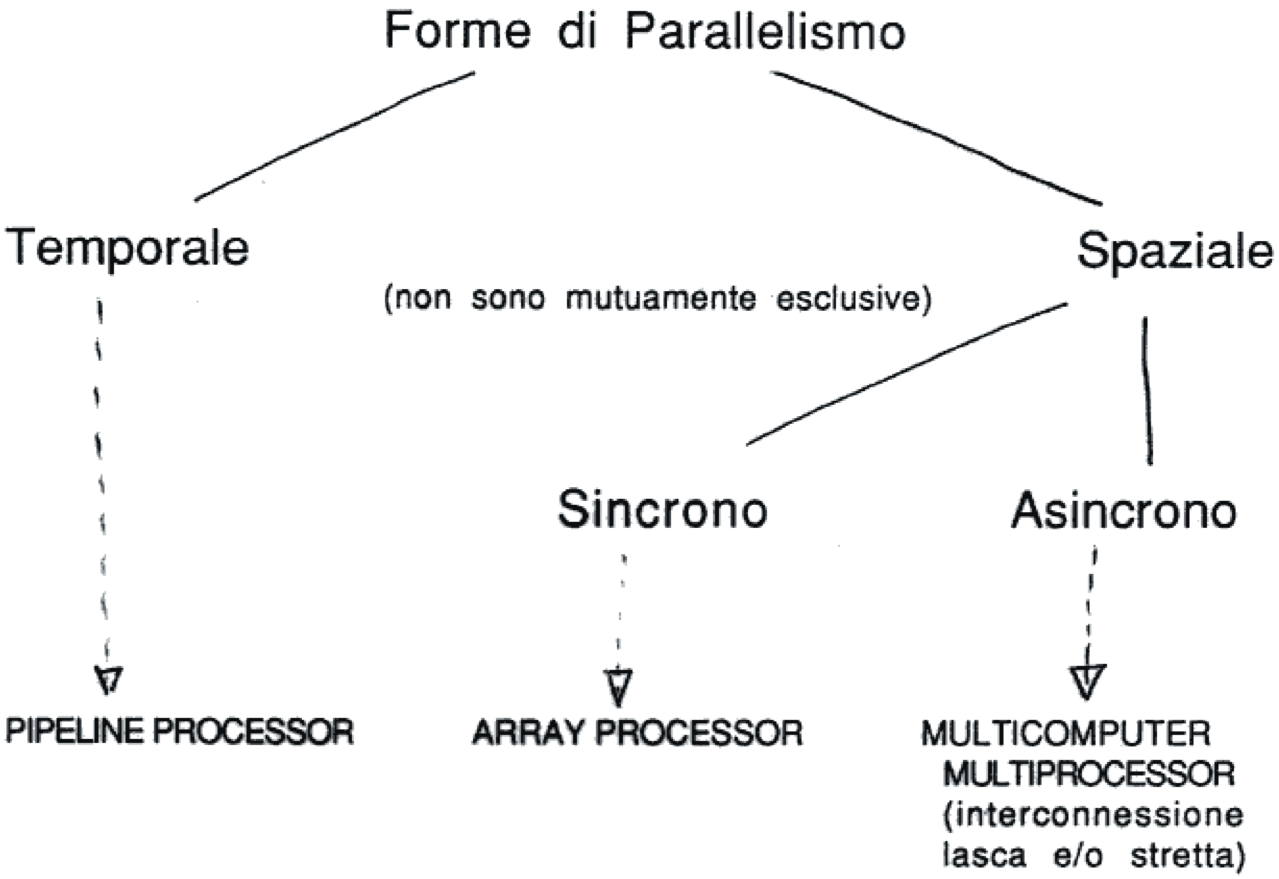
Meccanismi per Implementare il Parallelismo nei Sistemi di Elaborazione Uniprocessor

1. Molteplicità di unità funzionali
2. Parallelismo and pipelining all'interno della CPU
3. Sovrapposizione tra operazioni della CPU e dell'I/O
4. Uso di memoria organizzata gerarchicamente
5. Multiprogrammazione e time sharing

Strutturazione dei Calcolatori Paralleli



Forme di Parallelismo

Temporale

Spaziale

(non sono mutuamente esclusive)

Sincrono

Asincrono

PIPELINE PROCESSOR

ARRAY PROCESSOR

MULTICOMPUTER
MULTIPROCESSOR
(interconnessione
lasca e/o stretta)

Classificazione di Flynn

L'organizzazione dell'elaboratore è caratterizzata dalla molteplicità dell'hardware fornito per gestire il flusso delle istruzioni e dei dati

- **S**ingle **I**nstruction stream - **S**ingle **D**ata stream (SISD)

- **S**ingle **I**nstruction stream - **M**ultiple **D**ata stream (SIMD)

- **M**ultiple **I**nstruction stream - **S**ingle **D**ata stream (MISD)

- **M**ultiple **I**nstruction stream - **M**ultiple **D**ata stream (MIMD)

Esempi di SISD, SIMD, MISD, MIMD

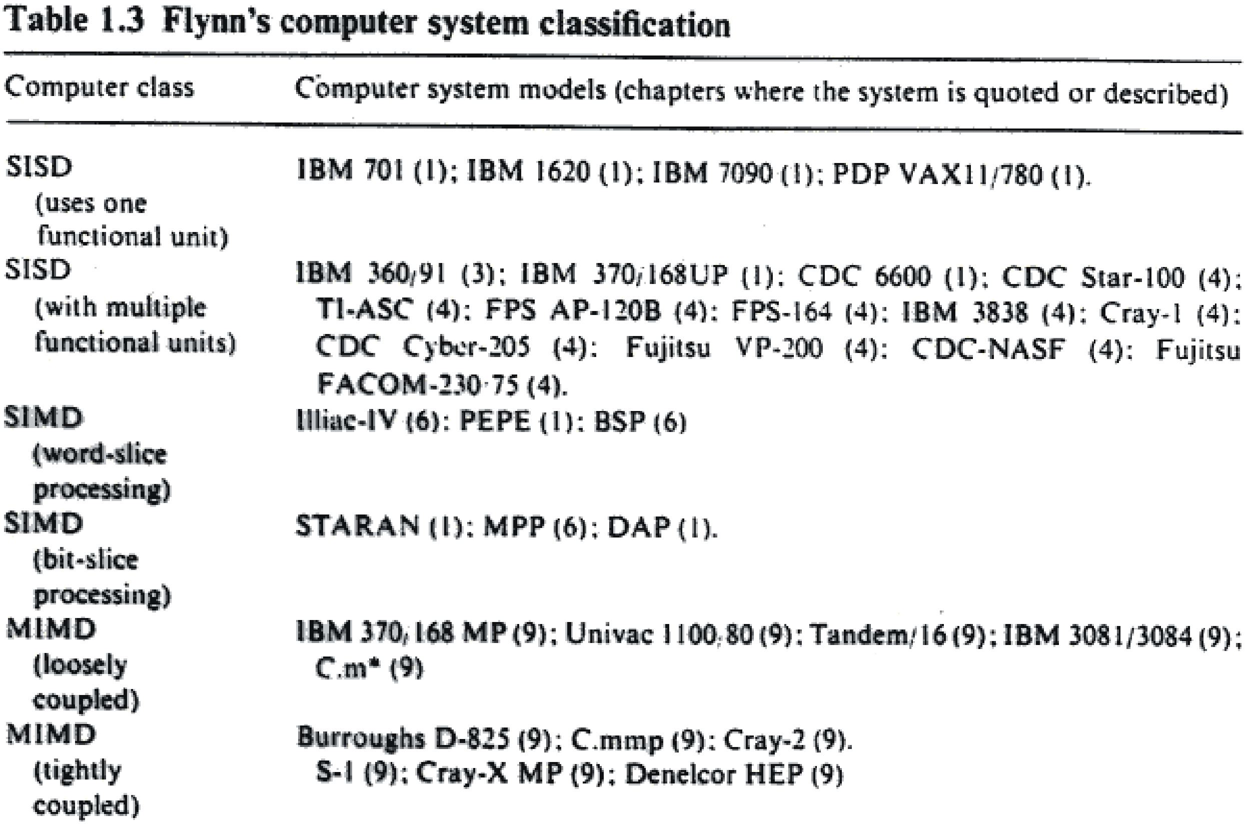


Table 1.3 Flynn's computer system classification

Computer class	Computer system models (chapters where the system is quoted or described)
SISD (uses one functional unit)	IBM 701 (1); IBM 1620 (1); IBM 7090 (1); PDP VAX11/780 (1).
SISD (with multiple functional units)	IBM 360/91 (3); IBM 370/168UP (1); CDC 6600 (1); CDC Star-100 (4); TI-ASC (4); FPS AP-120B (4); FPS-164 (4); IBM 3838 (4); Cray-1 (4); CDC Cyber-205 (4); Fujitsu VP-200 (4); CDC-NASF (4); Fujitsu FACOM-230-75 (4).
SIMD (word-slice processing)	Illiac-IV (6); PEPE (1); BSP (6)
SIMD (bit-slice processing)	STARAN (1); MPP (6); DAP (1).
MIMD (loosely coupled)	IBM 370/168 MP (9); Univac 1100/80 (9); Tandem/16 (9); IBM 3081/3084 (9); C.m* (9)
MIMD (tightly coupled)	Burroughs D-825 (9); C.mmp (9); Cray-2 (9). S-1 (9); Cray-X MP (9); Denelcor HEP (9)

Istituzione	Nome	Massimo numero di proc.	Bit per proc.	Frequenza di clock del proc.	Numero di FPU	Memoria massima per sistema (MB)	Banda passante massima per sistema (MB/s)	Anno
U. Illinois	Illiac IV	64	64	5 MHz	64	0,125	2560	1972
ICL	DAP	4 096	1	5 MHz	0	2	2560	1980
Goodyear	MPP	16 384	1	10 MHz	0	2	20 480	1982
Thinking Machines	CM-2	65 536	1	7 MHz	2048 (opzionale)	512	16 384	1987
Maspar	MP-1216	16 384	4	25 MHz	0	256 o 1024	23 000	1989

FIGURA 9.1 Caratteristiche di cinque calcolatori SIMD. Il numero di FPU indica il numero delle unità dedicate ai calcoli in virgola mobile.

Istituzione	Nome	Massimo numero di proc.	Bit per proc.	Frequenza di clock del proc.	Numero di FPU	Memoria massima per sistema (MB)	Banda passante massima per sistema (MB/s)	Anno
Sequent	Symmetry	30	32	16 MHz	30	240	53	1988
Silicon Graphics	4/360	16	32	40 MHz	16	512	320	1990
Sun	4/640	4	32	40 MHz	4	768	320	1991

FIGURA 9.4 Caratteristiche di tre calcolatori MIMD collegati tramite un singolo bus generico di sistema. Il numero di FPU indica il numero delle unità dedicate ai calcoli in virgola mobile. Per queste macchine, la banda passante per le comunicazioni corrisponde alla banda passante del bus.

Name	Maximum number of processors	Processor name	Processor clock rate	Maximum memory size/ system	Communi-cations BW/ system
Compaq ProLiant 5000	4	Pentium Pro	200 MHz	2,048 MB	540 MB/sec
Digital AlphaServer 8400	12	Alpha 21164	440 MHz	28,672 MB	2150 MB/sec
HP 9000 K460	4	PA-8000	180 MHz	4,096 MB	960 MB/sec
IBM RS/6000 R40	8	PowerPC 604	112 MHz	2,048 MB	1800 MB/sec
SGI Power Challenge	36	MIPS R10000	195 MHz	16,384 MB	1200 MB/sec
Sun Enterprise 6000	30	UltraSPARC 1	167 MHz	30,720 MB	2600 MB/sec

FIGURE 9.3 Characteristics of multiprocessor computers connected by a single backplane bus that are for sale in 1997. The communication style for these machines is shared memory with uniform memory access times. These machines are generally designed to be used with multiple generations of microprocessors both to allow customers to upgrade their existing machines and to allow companies to amortize their research and development investment. For example, the SGI Power Challenge was first delivered in 1993 with the MIPS R4400 and then again in 1995 with the R8000. Note that the bus and memory system did not change over this time. (See *www.mkp.com/cod2e.htm* for pointers to these and more recent bus-connected multiprocessors.)

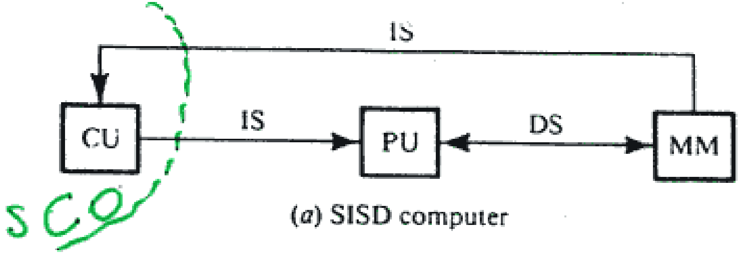
Istituzione	Nome	Numero di proc.	Bit per proc.	Frequenza di clock del proc.	Numero di FPU	Dimensione di memoria per sistema (MB)	Banda passante per la comunicazione (MB/s)		Anno
							Picco	Bisezione	
Intel	iPSC/2	128	16	16 MHz	128	512 MB	896	345	1988
nCube	nCube/ten	1024	32	10 MHz	1024	512 MB	10 240	640	1987
Intel	Delta	540	32	40 MHz	540	17 280 MB	21 600	640	1991
Thinking Machines	CM-5	1024	32	33 MHz	4096	32 768 MB	5120	5120	1991

FIGURA 9.13 Caratteristiche di quattro calcolatori MIMD collegati tramite una rete di interconnessione. Il numero di FPU indica il numero delle unità dedicate ai calcoli in virgola mobile. Tutte queste macchine hanno una memoria fisica distribuita e spazi di indirizzamento multipli e privati.

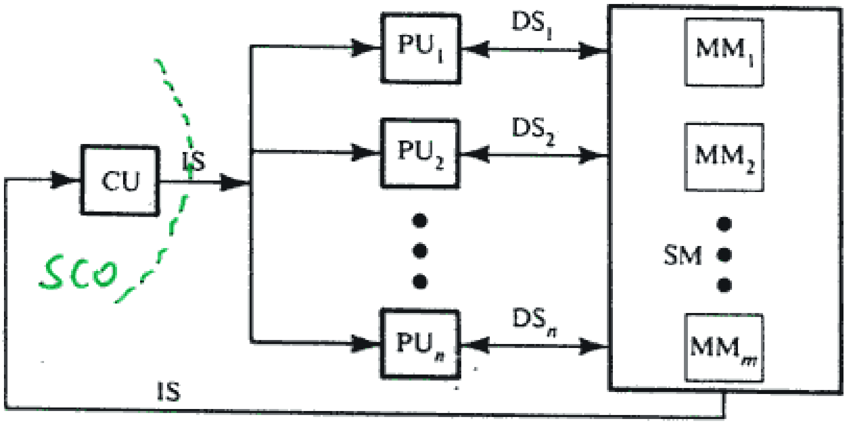
Name	Maximum number of processors	Processor name	Processor clock rate	Maximum memory size/ system	Communi-cations BW/link	Node	Topology
Cray Research T3E	2048	Alpha 21164	450 MHz	524,288 MB	1200 MB/sec	4-way SMP	3-D torus
HP/Convex Exemplar X-class	64	PA-8000	180 MHz	65,536 MB	980 MB/sec	2-way SMP	8-way crossbar + ring
Sequent NUMA-Q	32	Pentium Pro	200 MHz	131,072 MB	1024 MB/sec	4-way SMP	Ring
SGI Origin2000	128	MIPS R10000	195 MHz	131,072 MB	800 MB/sec	2-way SMP	6-cube
Sun Enterprise 10000	64	UltraSPARC 1	250 MHz	65,536 MB	1600 MB/sec	4-way SMP	16-way crossbar

FIGURE 9.9 Characteristics of multiprocessor computers connected by a network that are for sale in 1997. All these machines have a shared address space with nonuniform memory access time except for the Sun Enterprise 10000, which offers a shared address with uniform memory access time. And all these machines except the Cray Research T3E are cache coherent, with the HP, Sequent, and SGI using directories. The Sun machine uses buses for addresses and a switch for data, so it supports coherency with conventional snooping on the address buses. Communication bandwidth is peak per link, counting all bytes sent including network headers. The bisection bandwidth typically scales with the number of processors. (See *www.mkp.com/cod2e.htm* for pointers to these and more recent network-connected multiprocessors.)

Organizzazione base SISD, SIMD, MISD, MIMD

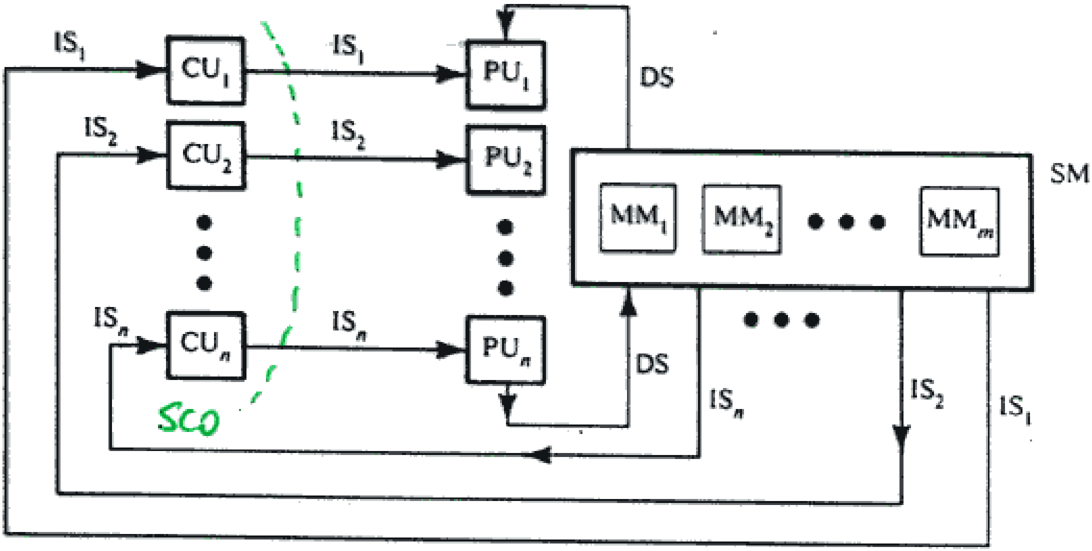


(*a*) SISD computer

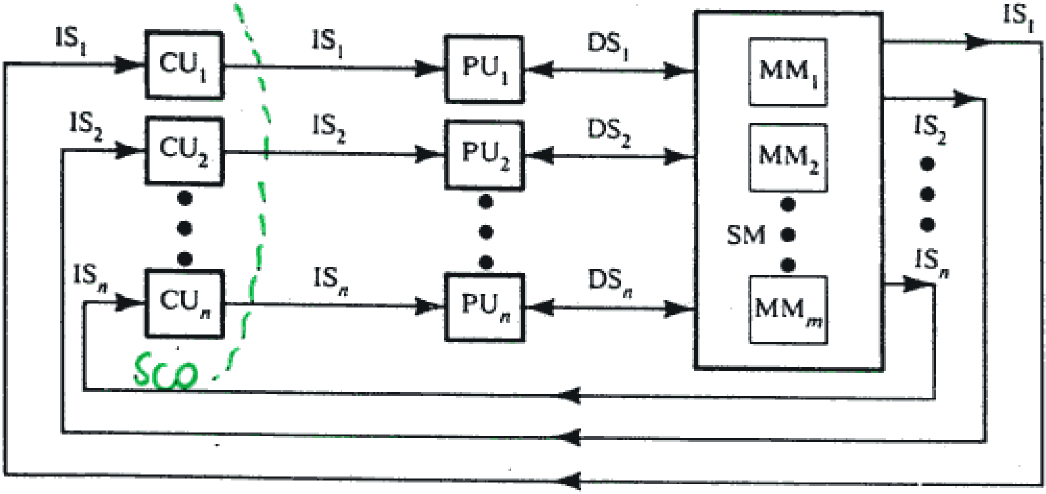


CU: control unit
PU: processor unit
MM: memory module
SM: shared memory
IS: instruction stream
DS: data stream

(*b*) SIMD computer



(*c*) MISD computer



(*d*) MIMD computer

Figure 1.16 Flynn's classification of various computer organizations.

Struttura Funzionale degli Array Processor

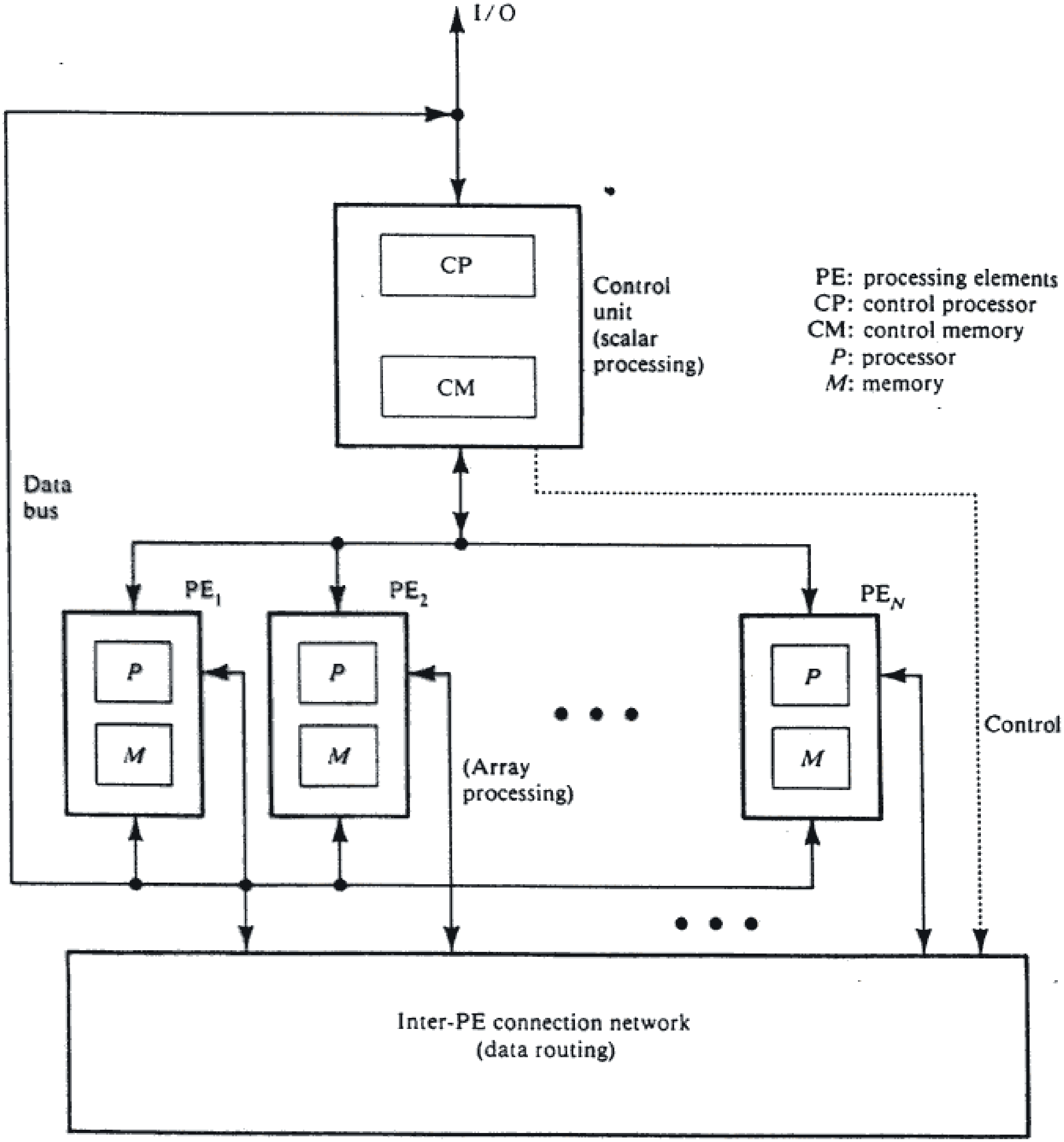


Figure 1.12 Functional structure of an SIMD array processor with concurrent scalar processing in the control unit.

Struttura Funzionale dei Multicomputer

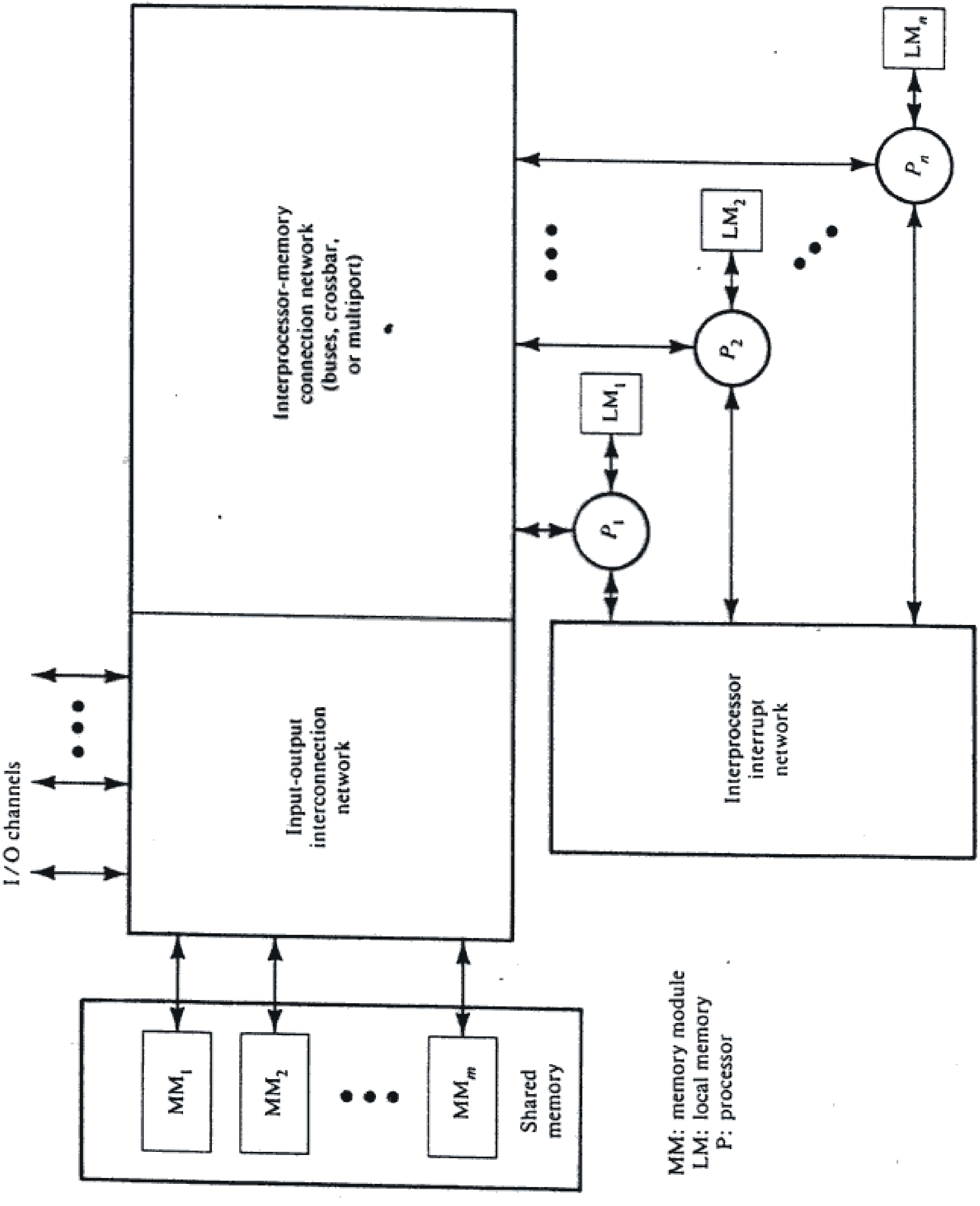

I/O channels
Input-output interconnection network
Interprocessor-memory connection network (buses, crossbar, or multiport)
Interprocessor interrupt network
P_1
P_2
P_n
LM_1
LM_2
LM_n
MM_1
MM_2
MM_m
Shared memory
MM: memory module
LM: local memory
P: processor

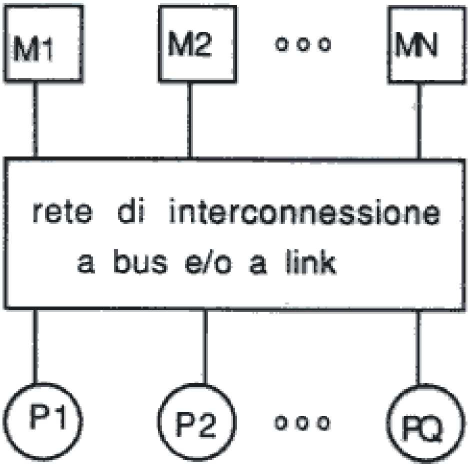
Figure 1.13 Functional design of an MIMD multiprocessor system.

Macchine MIMD

Multiprocessor

UMA Uniform Memory Access

Tightly Coupled Processors

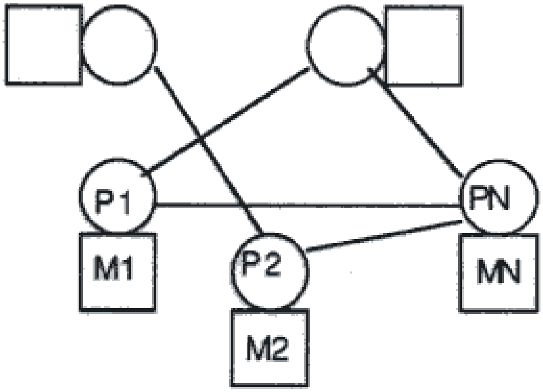


multiprocessor

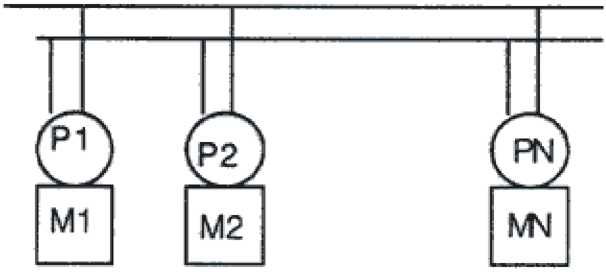
Multicomputer

NORMA No Remote Memory Access

Loose Coupled Processors



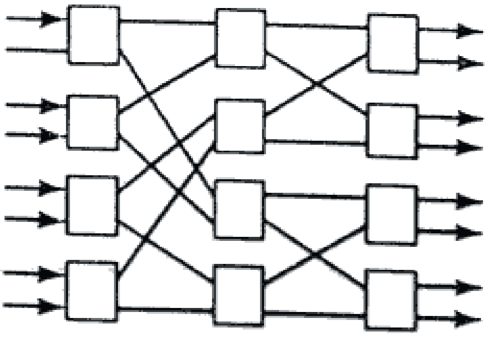
multicomputer a link



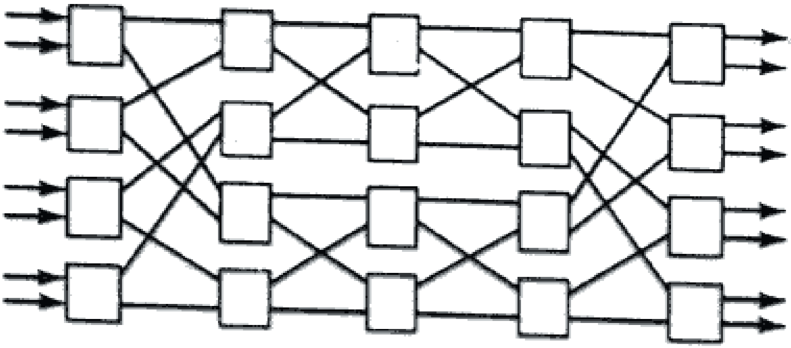
multi~~pro~~computer a bus

Esempi di

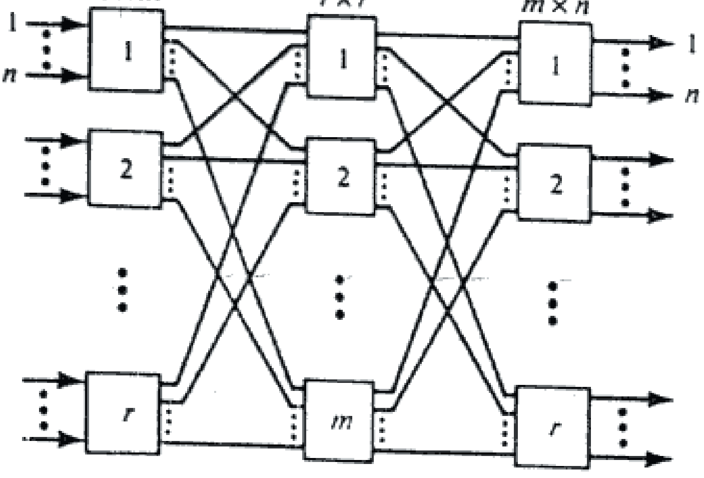
Strutture di Interconnessione
INDIRETTA
per Macchine MIMD



(*a*) 8 × 8 baseline network

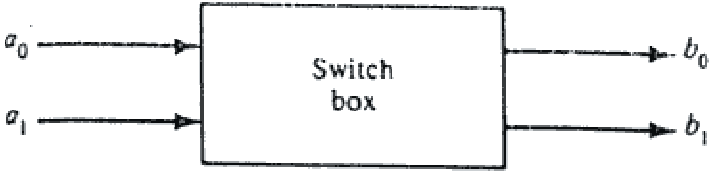


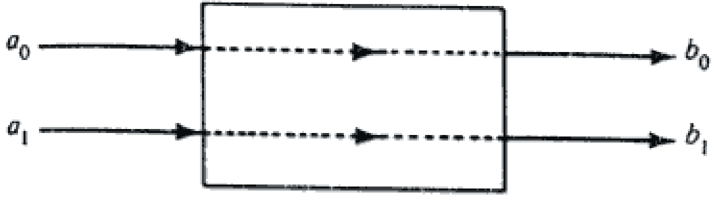
(*b*) 8 × 8 Benes network



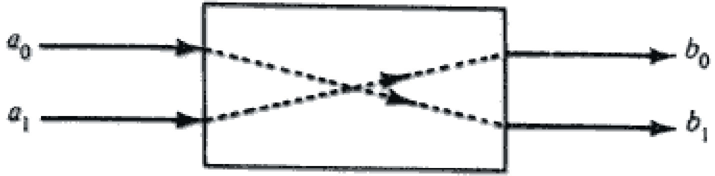
(*c*) Clos network

Figure 5.7 Several multistage interconnection networks.

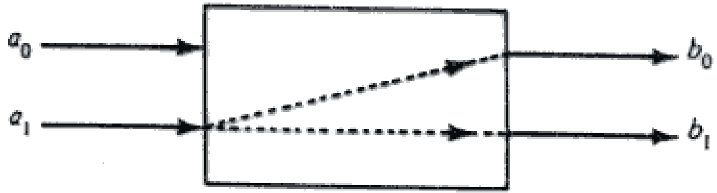




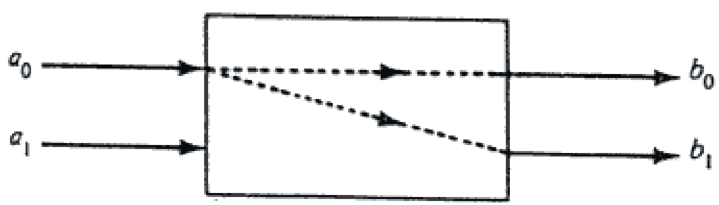
Straight



Exchange



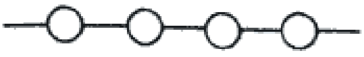
Upper broadcast



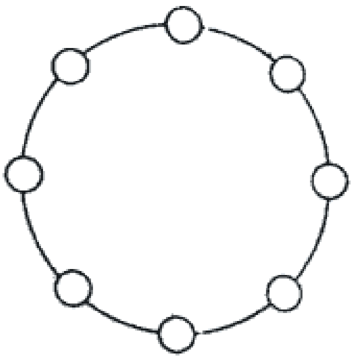
Lower broadcast

Esempi di

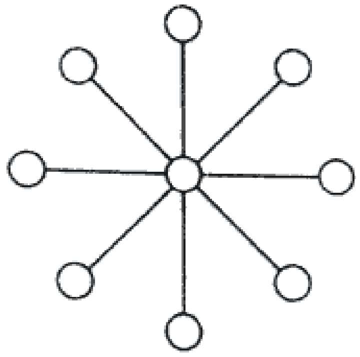
Strutture di Interconnessione DIRETTA per Macchine MIMD



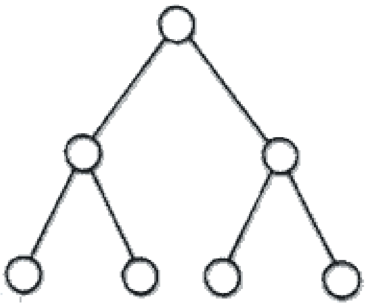
(*a*) Linear array



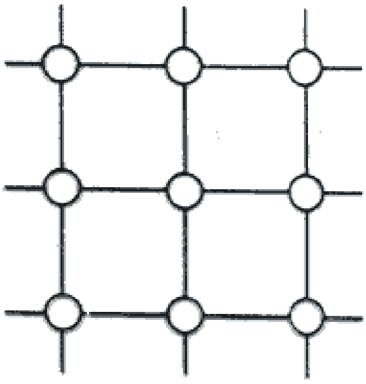
(*b*) Ring



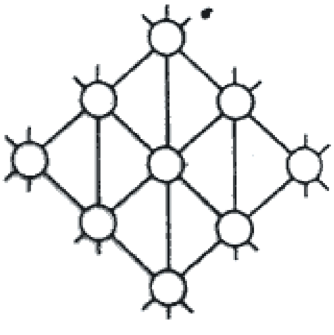
(*c*) Star



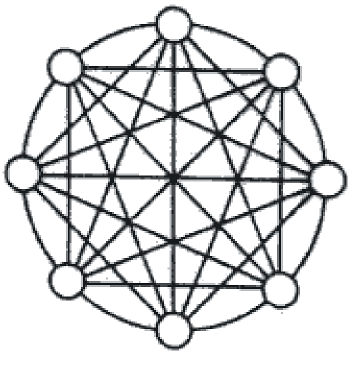
(*d*) Tree



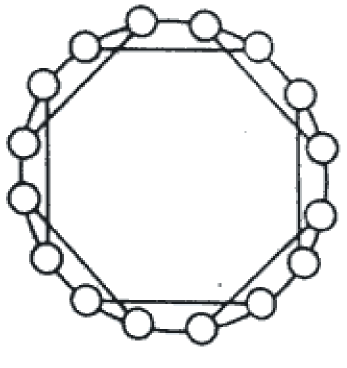
(*e*) Near-neighbor mesh



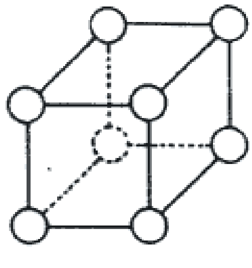
(*f*) Systolic array



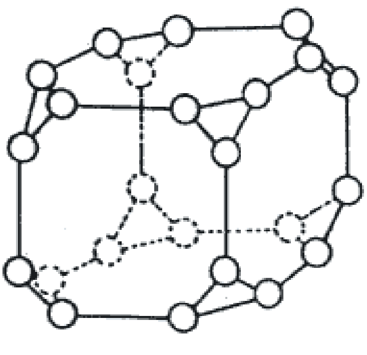
(*g*) Completely connected



(*h*) Chordal ring



(*i*) 3 cube



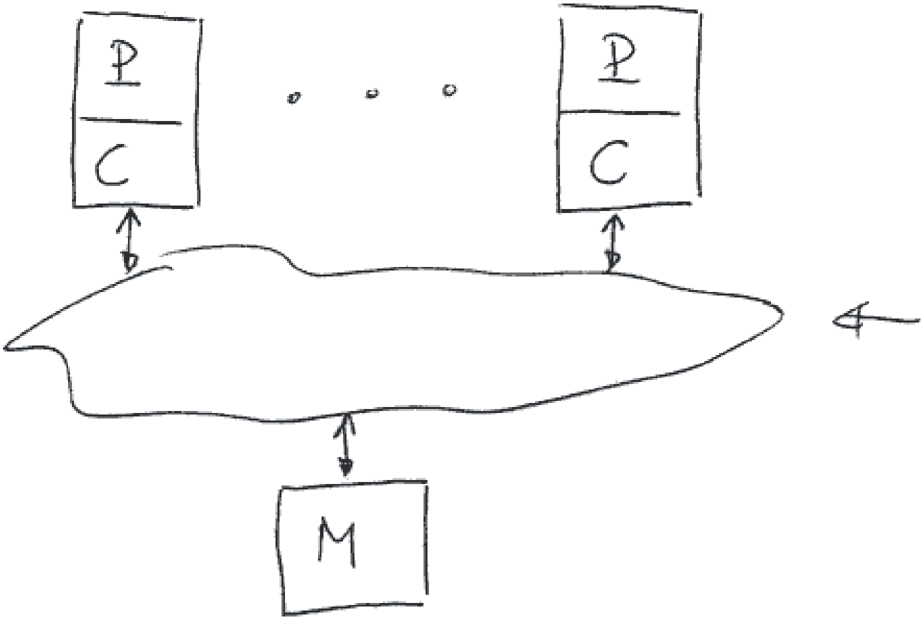
(*j*) 3-cube-connected cycle

Istituzione	Nome	Numero di nodi	Topologia di base	Bit per canale	Frequenza di clock della rete	BP*/canale (MB/s)	BP/Sistema (MB/s)	Bisezione (MB/s)	Anno
U. Illinois	Illiac IV	64	Griglia 2D	64	5 MHz	40	2560	320	1972
ICL	DAP	4096	Griglia 2D	1	5 MHz	0,6	2560	40	1980
Goodyear	MPP	16 384	Griglia 2D	1	10 MHz	1,2	20 480	160	1982
Thinking Machines	CM-2	da 1024 a 4096	12-cubo	1	7 MHz	1	65 536	1024	1987
nCube	nCube/ten	da 1 a 1024	10-cubo	1	10 MHz	1,2	10 240	640	1987
Intel	iPSC/2	da 16 a 128	7-cubo	1	16 MHz	2	896	345	1988
Maspar	MP-1216	da 32 a 512	Griglia 2D + Omega multistadio	1	25 MHz	3	23 000	1300	1989
Intel	Delta	540	Griglia 2D	16	40 MHz	40	21 600	640	1991
Thinking Machines	CM-5	da 32 a 1024	Albero grasso multistadio	4	40 MHz	20	20 480	5120	1991

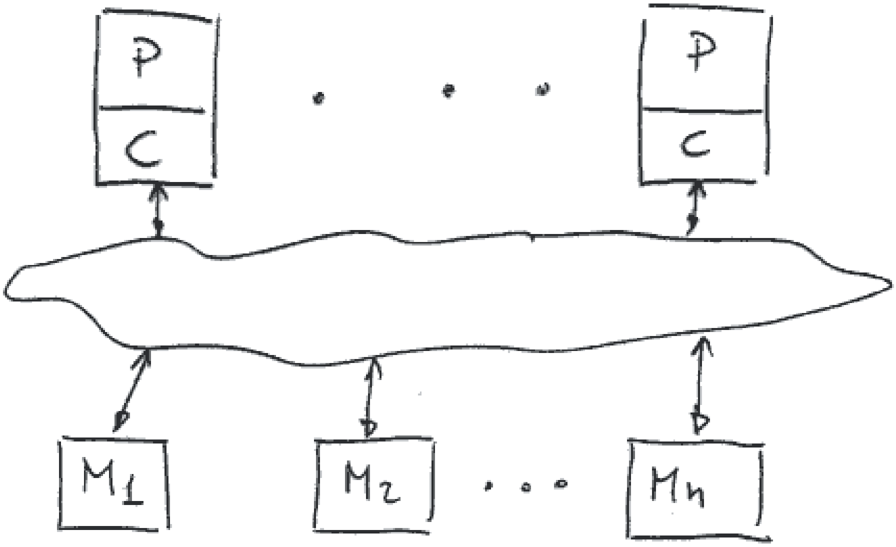
* La sigla BP sta per «banda passante», corrisponde al termine inglese «bandwidth» che spesso viene abbreviato in BW. [N.d.T.]

FIGURA 9.17 Caratteristiche delle reti di interconnessione adottate in alcuni dei processori paralleli menzionati in questo capitolo. La macchina Maspar raggruppa 32 processori da 4 bit ciascuno nei chip che si trovano in ciascun nodo, la CM-2 raggruppa invece 16 processori da 1 bit in ogni chip. La griglia 2D della macchina Intel Delta è di 16 righe per 35 colonne.

MULTIPROCESSOR



"SINGOLO BUS"



"MULTIPLE BUS
o
STRUT. INTERC.
INDIRETTA"

MULTIPROCESSOR A BUS SINGOLO

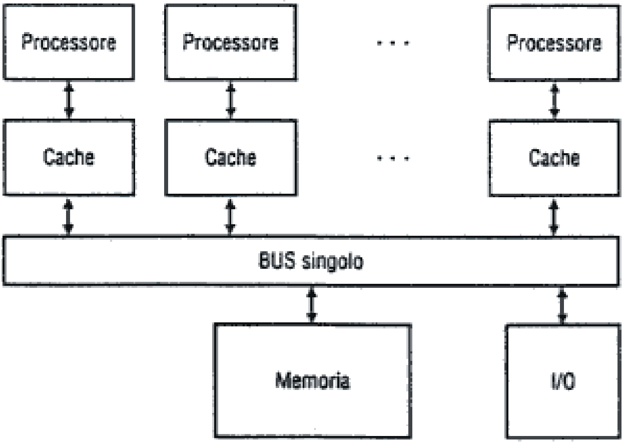


FIGURA 9.5 Un multiprocessore con bus singolo. La dimensione tipica prevede tra i 2 e i 32 processori.

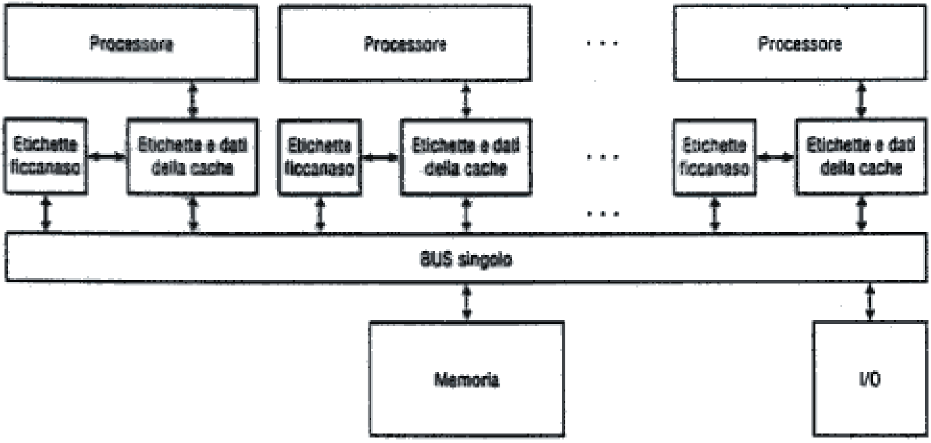


FIGURA 9.6 Multiprocessore a bus singolo che utilizza un protocollo ficcanaso per la gestione della coerenza delle cache. L'insieme delle etichette aggiuntive, che nella figura appaiono colorate, viene utilizzato per gestire le richieste di indagine. Queste etichette sono duplicate per ridurre le richieste di indagine sulle cache.

Name	Maximum number of processors	Processor name	Processor clock rate	Maximum memory size/ system	Communi-cations BW/ system
Compaq ProLiant 5000	4	Pentium Pro	200 MHz	2,048 MB	540 MB/sec
Digital AlphaServer 8400	12	Alpha 21164	440 MHz	28,672 MB	2150 MB/sec
HP 9000 K460	4	PA-8000	180 MHz	4,096 MB	960 MB/sec
IBM RS/6000 R40	8	PowerPC 604	112 MHz	2,048 MB	1800 MB/sec
SGI Power Challenge	36	MIPS R10000	195 MHz	16,384 MB	1200 MB/sec
Sun Enterprise 6000	30	UltraSPARC 1	167 MHz	30,720 MB	2600 MB/sec

FIGURE 9.3 Characteristics of multiprocessor computers connected by a single backplane bus that are for sale in 1997. The communication style for these machines is shared memory with uniform memory access times. These machines are generally designed to be used with multiple generations of microprocessors both to allow customers to upgrade their existing machines and to allow companies to amortize their research and development investment. For example, the SGI Power Challenge was first delivered in 1993 with the MIPS R4400 and then again in 1995 with the R8000. Note that the bus and memory system did not change over this time. (See *www.mkp.com/cod2e.htm* for pointers to these and more recent bus-connected multiprocessors.)

Istituzione	Nome	Massimo numero di proc.	Bit per proc.	Frequenza di clock del proc.	Numero di FPU	Memoria massima per sistema (MB)	Banda passante massima per sistema (MB/s)	Anno
Sequent	Symmetry	30	32	16 MHz	30	240	53	1988
Silicon Graphics	4/360	16	32	40 MHz	16	512	320	1990
Sun	4/640	4	32	40 MHz	4	768	320	1991

FIGURA 9.4 Caratteristiche di tre calcolatori MIMD collegati tramite un singolo bus generico di sistema. Il numero di FPU indica il numero delle unità dedicate ai calcoli in virgola mobile. Per queste macchine, la banda passante per le comunicazioni corrisponde alla banda passante del bus.

MULTIPROCESSORE SEQUENT SYMMETRY

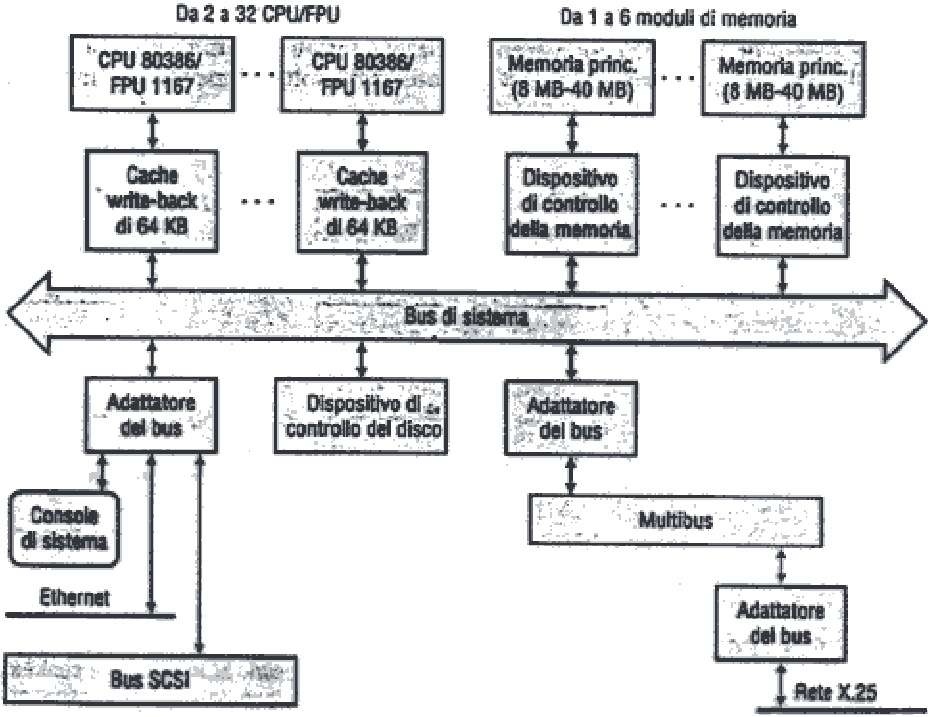

Da 2 a 32 CPU/FPU
Da 1 a 6 moduli di memoria
CPU 80386/ FPU 1167
CPU 80386/ FPU 1167
Memoria princ. (8 MB-40 MB)
Memoria princ. (8 MB-40 MB)
Cache write-back di 64 KB
Cache write-back di 64 KB
Dispositivo di controllo della memoria
Dispositivo di controllo della memoria
Bus di sistema
Adattatore del bus
Dispositivo di controllo del disco
Adattatore del bus
Console di sistema
Multibus
Ethernet
Adattatore del bus
Bus SCSI
Rete X.25

FIGURA 9.11 **Il multiprocessore Sequent Symmetry contiene fino a 30 microprocessori collegati al bus di sistema, ognuno dotato di una cache set-associativa a due vie di 64 KB che adotta una politica *write-back*. A questo bus largo 64 bit sono collegati fino a sei dispositivi di controllo della memoria, più alcune interfacce per le operazioni di I/O. In aggiunta a un dispositivo speciale per il controllo del disco, ci sono anche un'interfaccia per la console del sistema, una per l'allacciamento a una rete Ethernet e a un bus SCSI (si veda il capitolo 8), come pure un'ulteriore interfaccia per il Multibus. I dispositivi di I/O possono essere collegati sia al bus SCSI che al Multibus, secondo il volere dell'utente. (Sebbene tutte le interfacce siano definite «adattatori del bus», ognuna corrisponde a un diverso progetto.) Per trovare più dettagli sul comportamento della cache in questa macchina si può consultare il seguente articolo: T. Lovett e S. Thakkar, «The Symmetry multiprocessor system», in *Proc. 1988 International Conference on Parallel Processing*, pagg. 303-310.**

MULTIPROCESSOR VERSUS MULTICOMPUTER

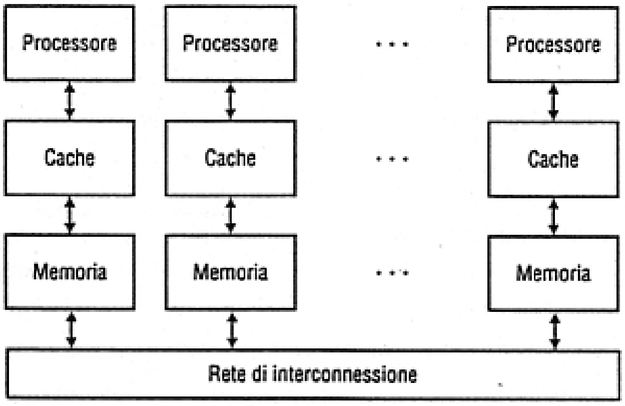


FIGURA 9.12 Organizzazione di un multiprocessore dotato di una rete di interconnessione. La dimensione tipica è tra i 32 e i 1024 processori. Si noti che, in contrasto con quanto riportato nella figura 9.5, la parte di interconnessione del multiprocessore non si trova più tra la memoria e i processori. Sono stati costruiti anche dei MIMD in cui la rete si trova tra i processori e la memoria; i multiprocessori Cray XMP e YMP sono forse gli esempi più conosciuti, ma attualmente questa organizzazione gode di scarsa considerazione.

Istituzione	Nome	Numero di proc.	Bit per proc.	Frequenza di clock del proc.	Numero di FPU	Dimensione di memoria per sistema (MB)	Banda passante per la comunicazione (MB/s)		Anno
							Picco	Bisezione	
Intel	iPSC/2	128	16	16 MHz	128	512 MB	896	345	1988
nCube	nCube/ten	1024	32	10 MHz	1024	512 MB	10 240	640	1987
Intel	Delta	540	32	40 MHz	540	17 280 MB	21 600	640	1991
Thinking Machines	CM-5	1024	32	33 MHz	4096	32 768 MB	5120	5120	1991

FIGURA 9.13 Caratteristiche di quattro calcolatori MIMD collegati tramite una rete di interconnessione. Il numero di FPU indica il numero delle unità dedicate ai calcoli in virgola mobile. Tutte queste macchine hanno una memoria fisica distribuita e spazi di indirizzamento multipli e privati.

Name	Maximum number of processors	Processor name	Processor clock rate	Maximum memory size/ system	Communications BW/link	Node	Topology
Cray Research T3E	2048	Alpha 21164	450 MHz	524,288 MB	1200 MB/sec	4-way SMP	3-D torus
HP/Convex Exemplar X-class	64	PA-8000	180 MHz	65,536 MB	980 MB/sec	2-way SMP	8-way crossbar + ring
Sequent NUMA-Q	32	Pentium Pro	200 MHz	131,072 MB	1024 MB/sec	4-way SMP	Ring
SGI Origin2000	128	MIPS R10000	195 MHz	131,072 MB	800 MB/sec	2-way SMP	6-cube
Sun Enterprise 10000	64	UltraSPARC 1	250 MHz	65,536 MB	1600 MB/sec	4-way SMP	16-way crossbar

FIGURE 9.9 Characteristics of multiprocessor computers connected by a network that are for sale in 1997. All these machines have a shared address space with nonuniform memory access time except for the Sun Enterprise 10000, which offers a shared address with uniform memory access time. And all these machines except the Cray Research T3E are cache coherent, with the HP, Sequent, and SGI using directories. The Sun machine uses buses for addresses and a switch for data, so it supports coherency with conventional snooping on the address buses. Communication bandwidth is peak per link, counting all bytes sent including network headers. The bisection bandwidth typically scales with the number of processors. (See *www.mkp.com/cod2e.htm* for pointers to these and more recent network-connected multiprocessors.)

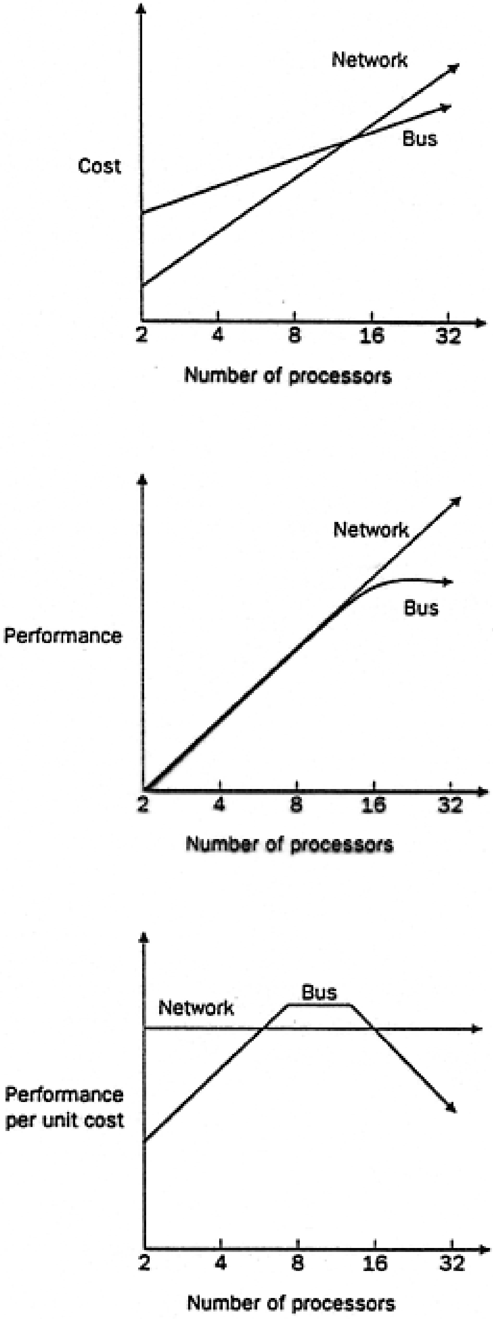
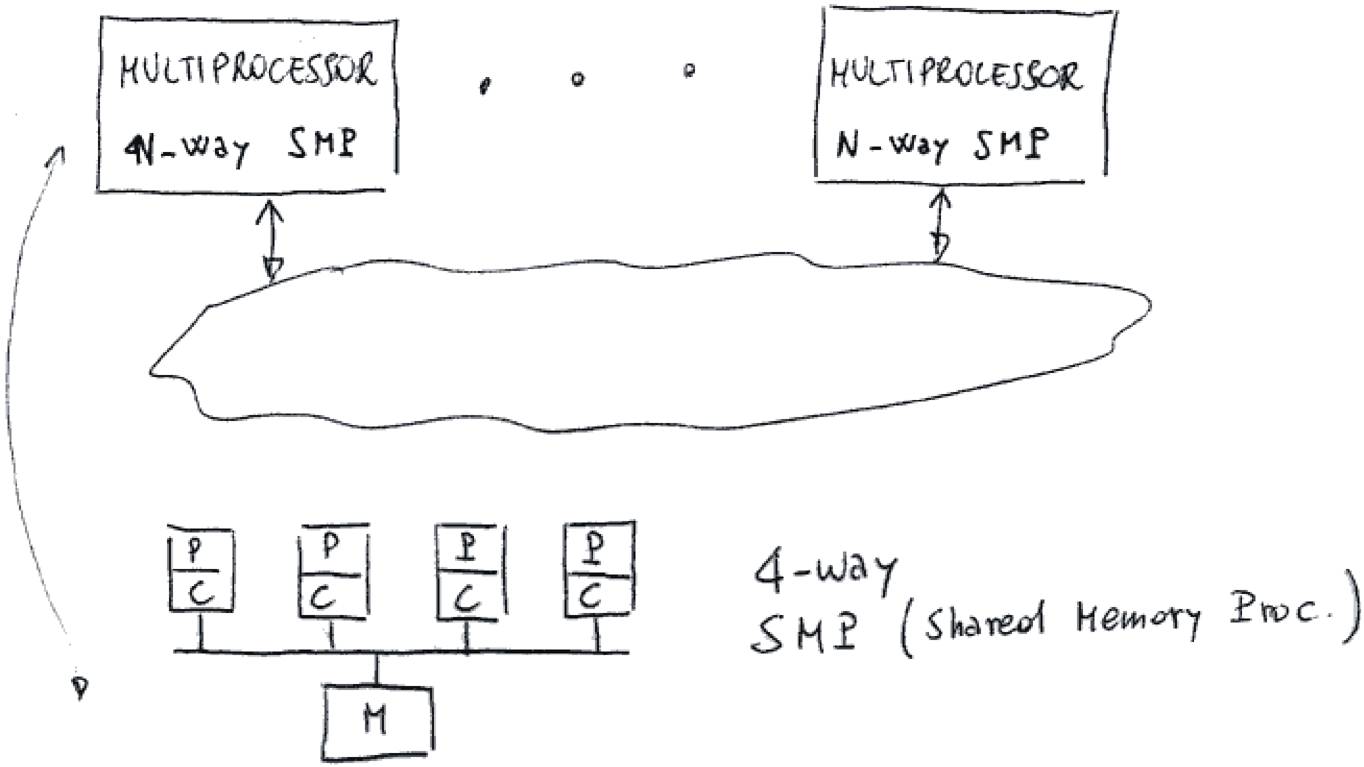


FIGURE 9.11 Cost, performance, and cost/performance of bus-connected and network-connected shared address multiprocessors. The combination of cost and performance suggests a "sweet spot" in 1997 for bus-connected multiprocessors of 8 to 16 processors, shown as a plateau in the cost/performance graph. Network-connected multiprocessors have better cost/performance to the left of the sweet spot because they are less expensive, and better cost/performance to the right of the sweet spot because they have higher performance. A bus designer effectively chooses the sweet spot by the width and the speed of the bus, which determines both the left edge of the plateau (cost) and right edge (scalability). *(page 733)*

CLUSTER – NOW (Network of Workstation)



Name	Maximum number of processors	Processor name	Processor clock rate	Maximum memory size/ system	Communi-cations BW/link	Node	Maximum number of nodes
HP 9000 EPS21	64	PA-8000	180 MHz	65,536 MB	532 MB/sec	4-way SMP	16
IBM RS/6000 HACMP R40	16	PowerPC 604	112 MHz	4,096 MB	12 MB/sec	8-way SMP	2
IBM RS/6000 SP2	512	Power2 SC	135 MHz	1,048,576 MB	150 MB/sec	16-way node	32
Sun Enterprise Cluster 6000 HA	60	UltraSPARC	167 MHz	61,440 MB	100 MB/sec	30-way SMP	2
Tandem NonStop Himalaya S70000	4096	MIPS R10000	195 MHz	1,048,576 MB	40 MB/sec	16-way SMP	256

FIGURE 9.12 Characteristics of clusters commercially available in 1997. All but the IBM SP2 are marketed for high-availability applications. The SP2 is used for number-crunching scientific applications and for data mining. (See *www.mkp.com/cod2e.htm* for pointers to these and more recent clusters.)

MESSAGE PASSING TECHNIQUES

- communication switching techniques
  - circuit switching
  - store and forward
  - wormhole
  - virtual cut through

- routing policy:
  - static or deterministic
  - dynamic or adaptive

- link conflict resolution strategy
  - hold
  - drop

INTERCONNECTION NETWORKS FOR MPP

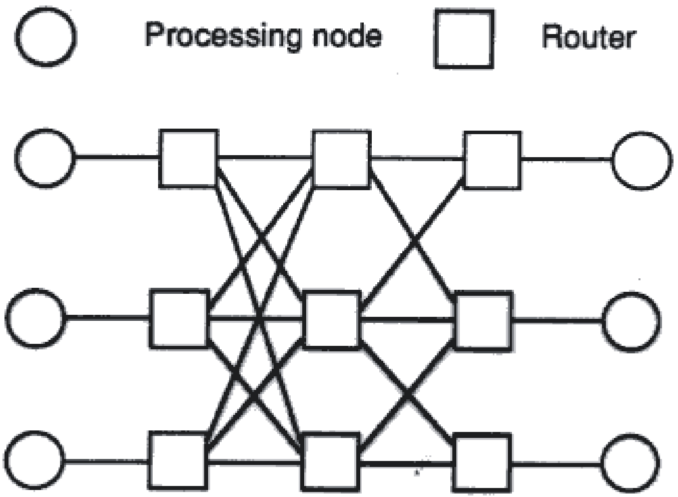
interconnection network

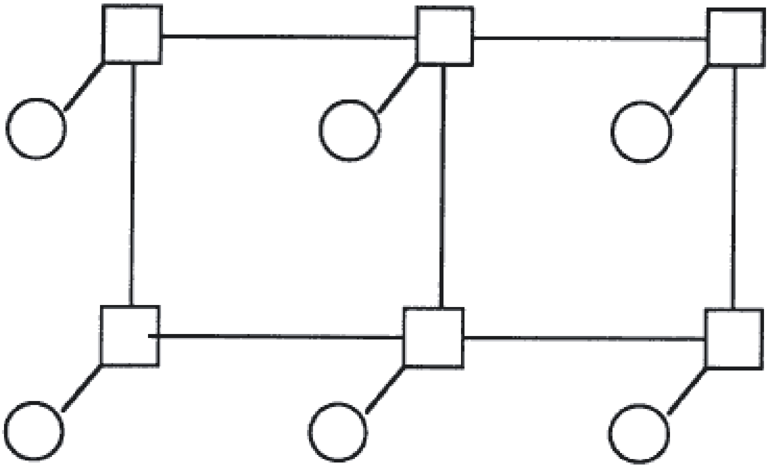
direct

computing node physically
adiacent
to communication node (router)

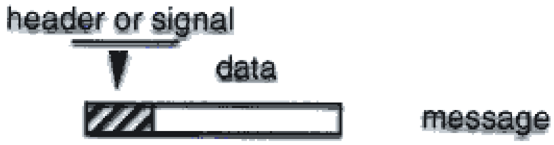
indirect

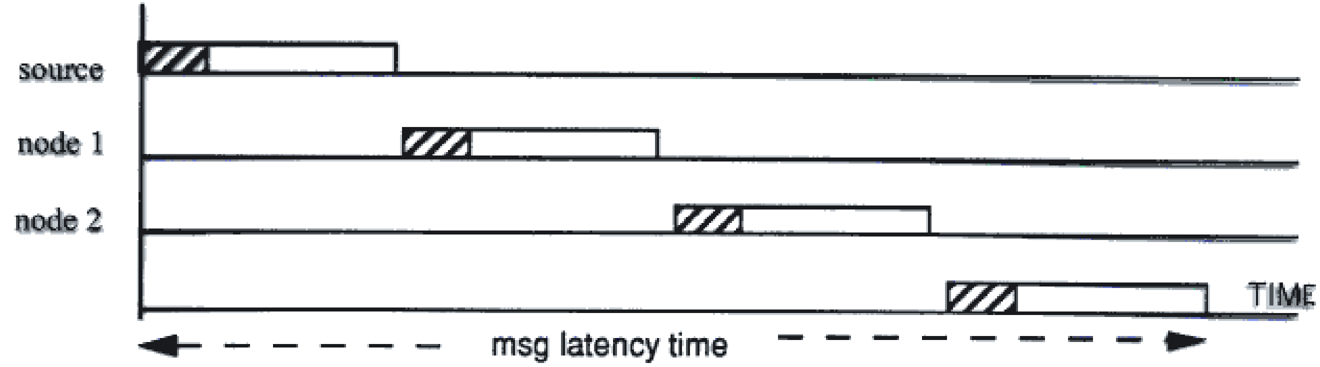
computing node physically
not adiacent
to communication node (router)



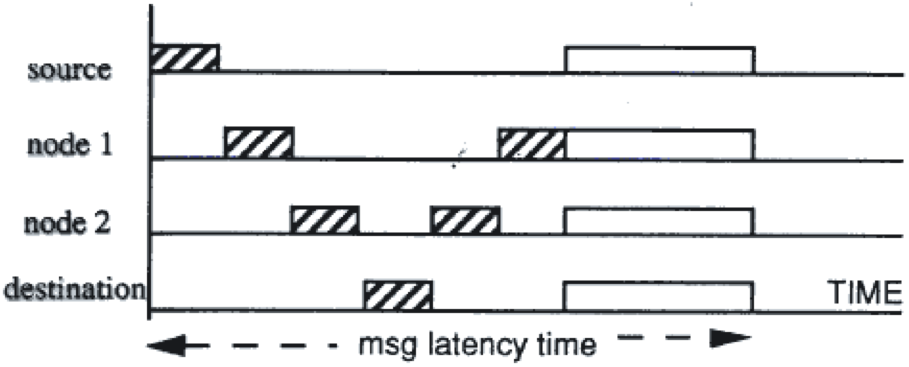


COMMUNICATION SWITCHING TECHNIQUES

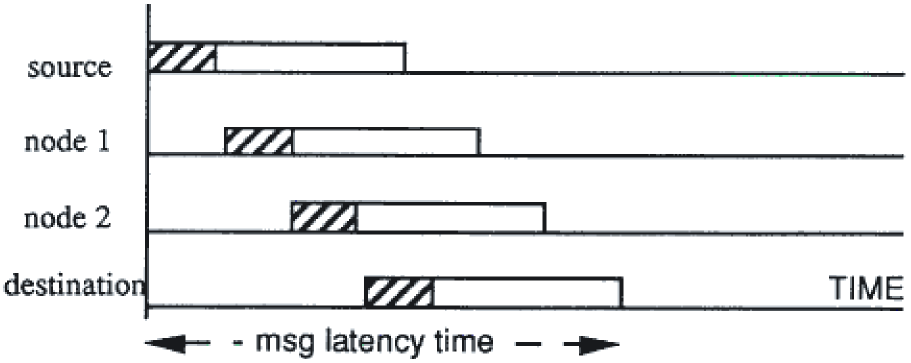




MESSAGE SWITCHING (Store and Forward)

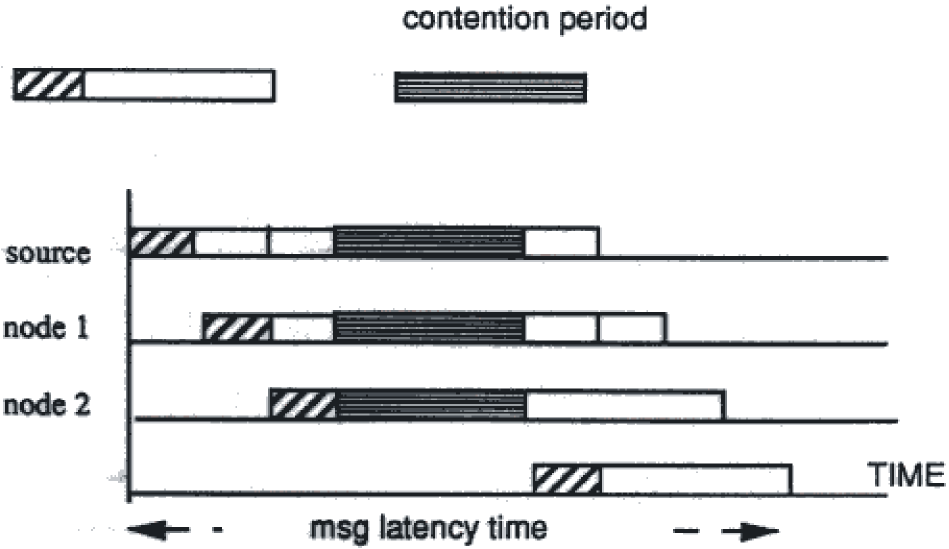


CIRCUIT SWITCHING

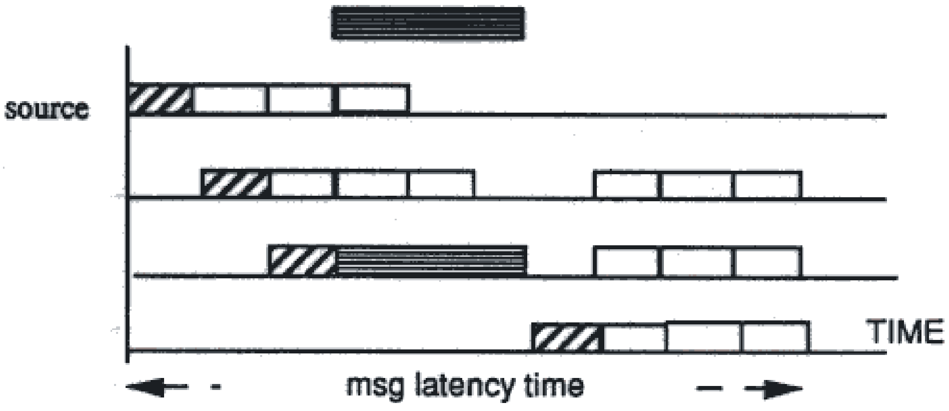


WORMHOLE & CUT THROUGH

CASE WITH CONTETION



WORMHOLE



CUT-THROUGH

The *store-and-forward* (SF) switching techniques collect the entire message (or packet) at each intermediate node of the path before requesting the next link.

The *virtual-cut-through* (VCT) technique proposed by Kermani and Kleinrock [4]. In contrast to the previous strategy, the message is divided in small parts (*flits*). Once obtained the first link, before the message is entirely received at the adjacent node, the second link is required and, if obtained, the flits are sent out to the following node. This strategy allows to reduce buffering only to the cases in which a link is not available. Nevertheless, VCT was adopted only by prototype machines.

The second generation multicomputers (such as Intel iPSC/2 and iPSC/860) preferred *circuit-switching* (CS) techniques that avoid large node buffers. In this case the message header has to establish the entire path from the source to the destination node before data can be transmitted. Since each communication requires and holds all links of its path, link contention tends to increase, thus significantly affecting performance in case of high traffic.

Last generation multicomputers (such as Intel Paragon, Ncube-2 and Ncube-3) adopted *wormhole* (WH) routing techniques. In a contention-free network WH has the same pipeline behaviour as VCT. In case of conflict, instead, WH does not gather all the flits in a node, but blocks them in the flit buffers of the built path.

The communication paths from the source to the destination node are completed by a step-by-step process, in which for each step the additional link is assigned by the local routing controller only after having verified that it is free. In case of conflict, there are several possibilities depending on the adopted *routing policy* and *link-conflict resolution strategy*.

The former may be *deterministic* or *adaptive*. Early, all the commercial multicomputers used deterministic policies in which the route between sender and receiver nodes is fixed. They are easier to implement but do not have the same ability to respond to dynamic network conditions (congestion and faults) as the adaptive policies.

THE SIMULATOR

The simulation model is discrete event driven (INTNETSIM). It has been implemented in Simula language on Unix based platforms.

INTNETSIM can model any *k*-ary *n*-cube topology, and in particular commercial interconnection networks such as 2D and 3-D mesh, torus, and hypercube.

INTNETSIM may model many features of the interconnection network, such as topology, dimension, link bandwidth, one/two directional channels, router at different levels of detail (no delay, constant delay, queuing server), node buffer of chosen dimension starting from null.

Several output parameters can be chosen to evaluate performance of message passing algorithms. Among them, mean message latency time, mean service time of router, mean length path, probability of link conflict.

The output analysis adopts the independent replication methods where the confidence intervals at 95% are based on the Jackknife estimator. The number of replications depends on network traffic.

PERFORMANCE RESULTS

For the architecture, we focus on **hypercube** that has been adopted in several multicomputers such as Cosmic Cube, Connection Machine, nCUBE, iPSC/2, and iPSC/860.

The network dimension is set to 6, the channel type is at single link.

The communication router is modelled as queuing server.

The message generation rate is chosen as a Poisson distribution.

The destination node and the message length are given by two uniform distributions.

The performance comparisons are based on *mean message latency* that includes path-set-up time, transfer message time, and link release time.

d) the analysed message passing techniques are:

Switching technique	Routing policy	Link-conflict resolution strategy
SF	deterministic	wait
SF	adaptive	wait
VCT	deterministic	wait
VCT	adaptive	wait
CS	deterministic	wait
CS	deterministic	drop all
CS	adaptive	drop one
CS	adaptive	drop all
W H	deterministic	wait
W H	adaptive	wait

1-st EXPERIMENT

deterministic routing policy
limited versus infinite input buffers

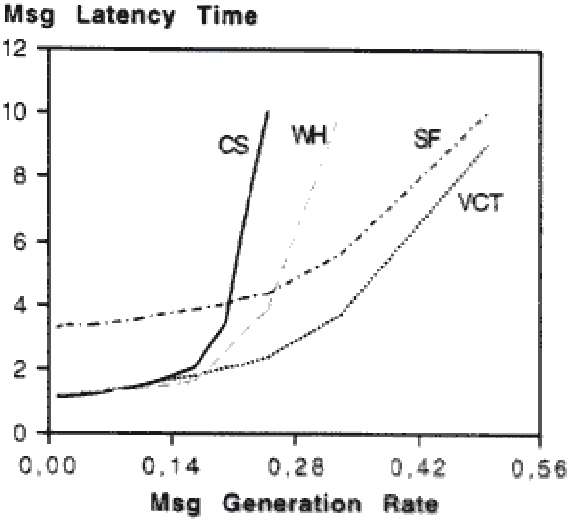


Figure 1. *infinite buffers.*

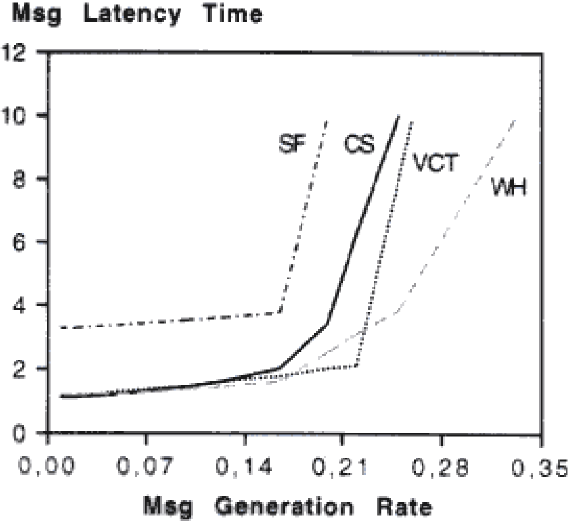


Figure 2: *finite buffers.*

2-nd EXPERIMENT

fixed communication switching technique (SF and CS)

deterministic versus adaptive routing

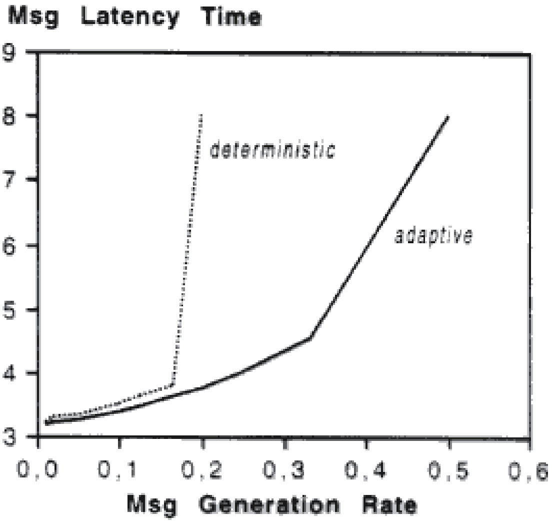


Figure 3. SF techniques.

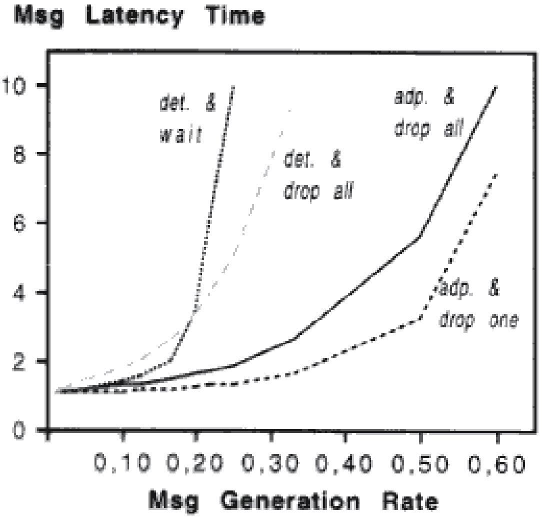


Figure 4. CS techniques.

3-rd EXPERIMENT

adaptivity versus message switching techniques

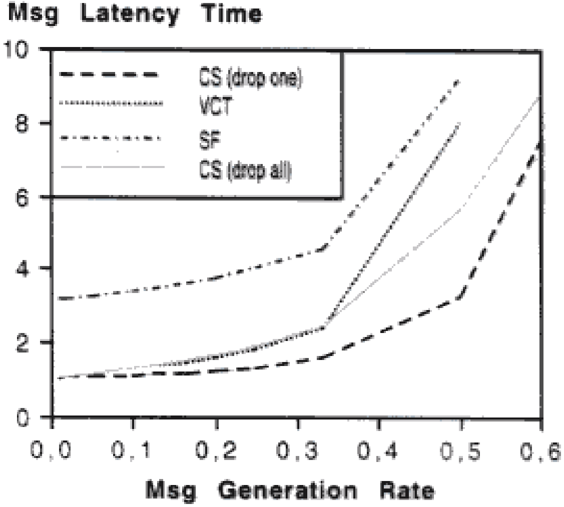


Figure 5. Adaptive routing policies.

4-th EXPERIMENT

fixed message transmission rate

variation of the message lengh

deterministic routing

finite buffer

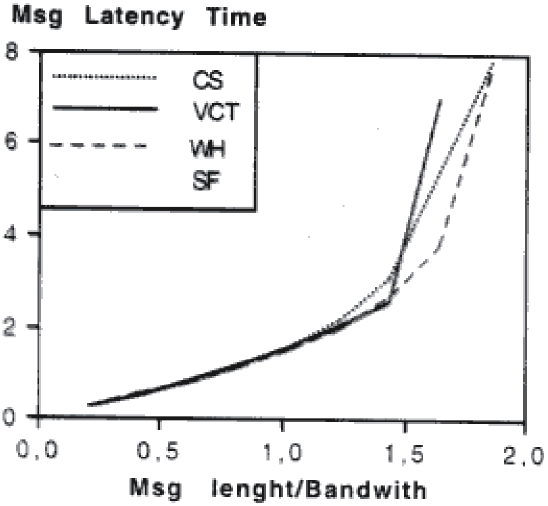


Figure 6. Four switching techniques in case of deterministic routing (Case II: finite buffers).