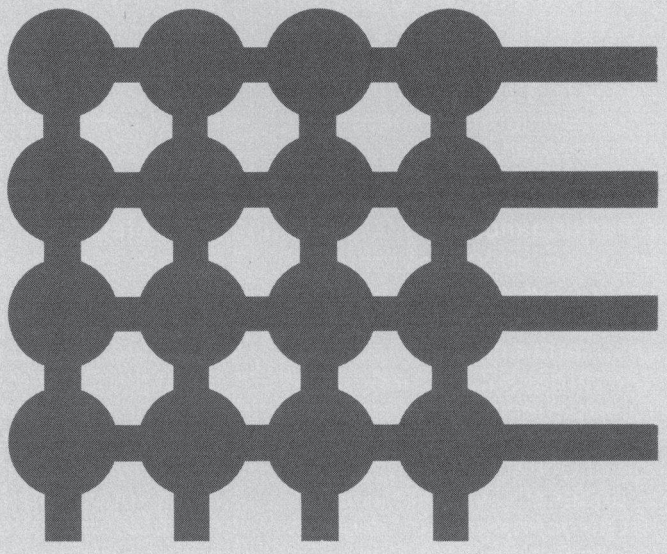
Concurrent processing depends on interconnection networks for communication among processors and memory modules. Various network topologies and switching strategies are covered here.

A Survey of Interconnection Networks



Tse-yun Feng
The Ohio State University

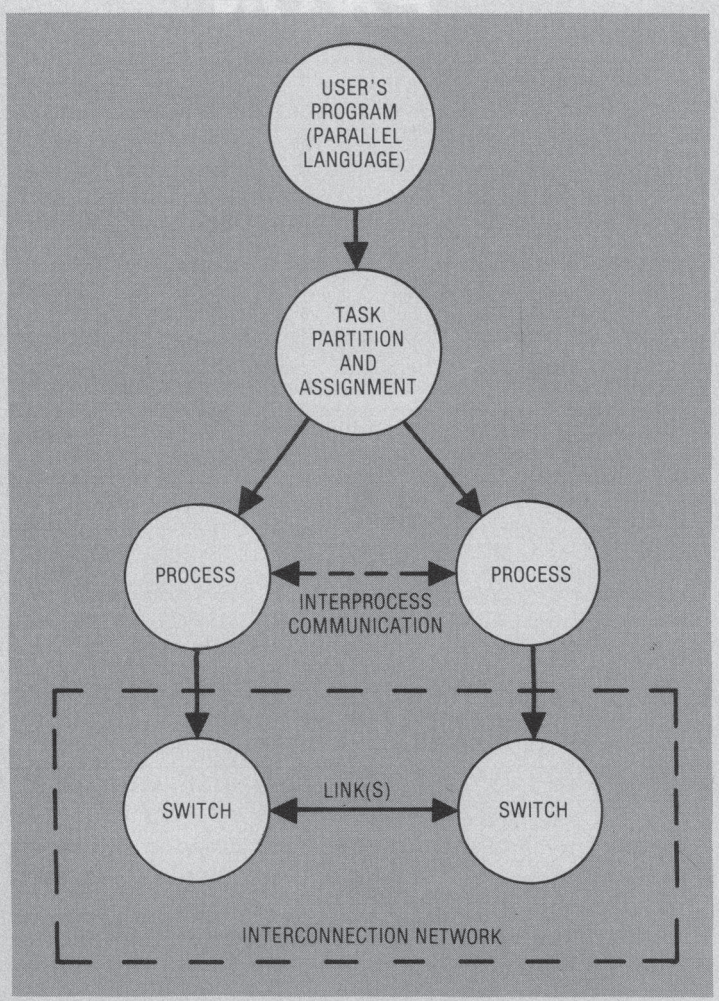


Figure 1. An overview of concurrent processing systems.

Concurrent processing of data items is considered a proper approach for significantly increasing processing speed.[1] In many real-time applications—such as image processing and weather computation, which need an instruction execution rate of more than one billion floating-point instructions per second—concurrent processing is unavoidable. And now, with the advent of LSI technology, it is economically feasible to construct a concurrent processing system by interconnecting hundreds—even thousands—of off-the-shelf processors and memory modules.

A basic concurrent processing system is shown in Figure 1. Processes, generated by compiling and partitioning a user's program, are assigned to individual processors, and an interconnection network implements interprocess communication. A general model of the hardware system is shown in Figure 2. The interconnection network facilitates communication not only among the n processors and the m memory modules but also between the processors and memory modules.

Many interconnection networks have been reviewed in other surveys.[2-9] In this article we consider interconnection networks from a practical design viewpoint. We examine design decisions that are essential in choosing a cost-effective communication network, survey the various topologies and communication protocols, and discuss connection issues related to concurrent processing.

Design decisions

In selecting the architecture of an interconnection network, four design decisions can be identified.[10] They concern operation mode, control strategy, switching method, and network topology.

0018-9162/81/1200-0012$00.75 © 1981 IEEE

Operation mode. Two types of communication can be identified: synchronous and asynchronous. Synchronous communication is needed for processing in which communication paths are established synchronously for either a data manipulating function[11] or a data/instruction broadcast. Asynchronous communication is needed for multiprocessing in which connection requests are issued dynamically. A system may also be designed to facilitate both synchronous and asynchronous processing. Therefore, typical operation modes of interconnection networks can be classified into three categories: synchronous, asnychronous, and combined.

Control strategy. A typical interconnection network consists of a number of switching elements and interconnecting links. Interconnection functions are realized by properly setting control of the switching elements. The control-setting function can be managed by a centralized controller or by the individual switching element. The latter strategy is called distributed control; the first strategy is called centralized control.

Switching methodology. The two major switching methodologies are circuit switching and packet switching. In circuit switching, a physical path is actually established between a source and a destination. In packet switching, data is put in a packet and routed through the interconnection network without establishing a physical connection path. In general, circuit switching is much more suitable for bulk data transmission, and packet switching is more efficient for short data messages. Another option, integrated switching, includes capabilities of both circuit switching and packet switching. Therefore, three switching methodologies can be identified: circuit switching, packet switching, and integrated switching.

Network topology. A network can be depicted by a graph in which nodes represent switching points and edges represent communication links. The topologies tend to be regular and can be grouped into two categories: static and dynamic. In a static topology, links between two processors are passive and dedicated buses cannot be reconfigured for direct connections to other processors. On the other hand, links in the dynamic category can be reconfigured by setting the network's active switching elements.

The cross product of the set of categories in each design decision—{operation mode} × {control strategy} × {switching methodology} × {network topology}—represents a space of interconnection networks. Obviously, the cross product contains some uninteresting cases, but a network designer can obtain a meaningful subspace by exercising a practical view of engineering technology.

Topologies

Network topology is a key factor in determining a suitable architectural structure, and many topologies have been considered for telephone switching connections.[12] Here, we review those proposed or used for connections in tightly coupled multiple-processor systems (see Figure 3).

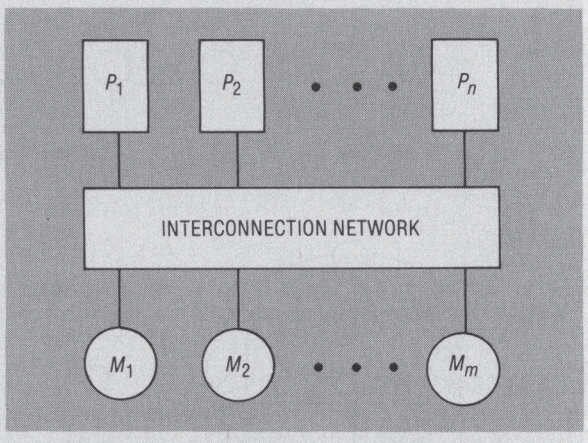


Figure 2. Hardware model of concurrent processing systems.

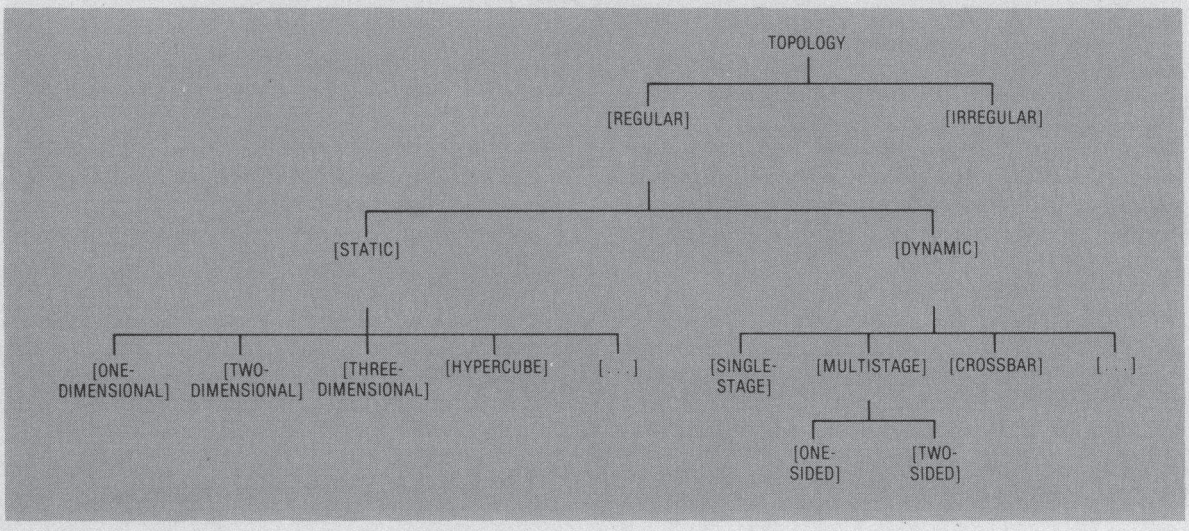


Figure 3. Topologies of interconnection networks.

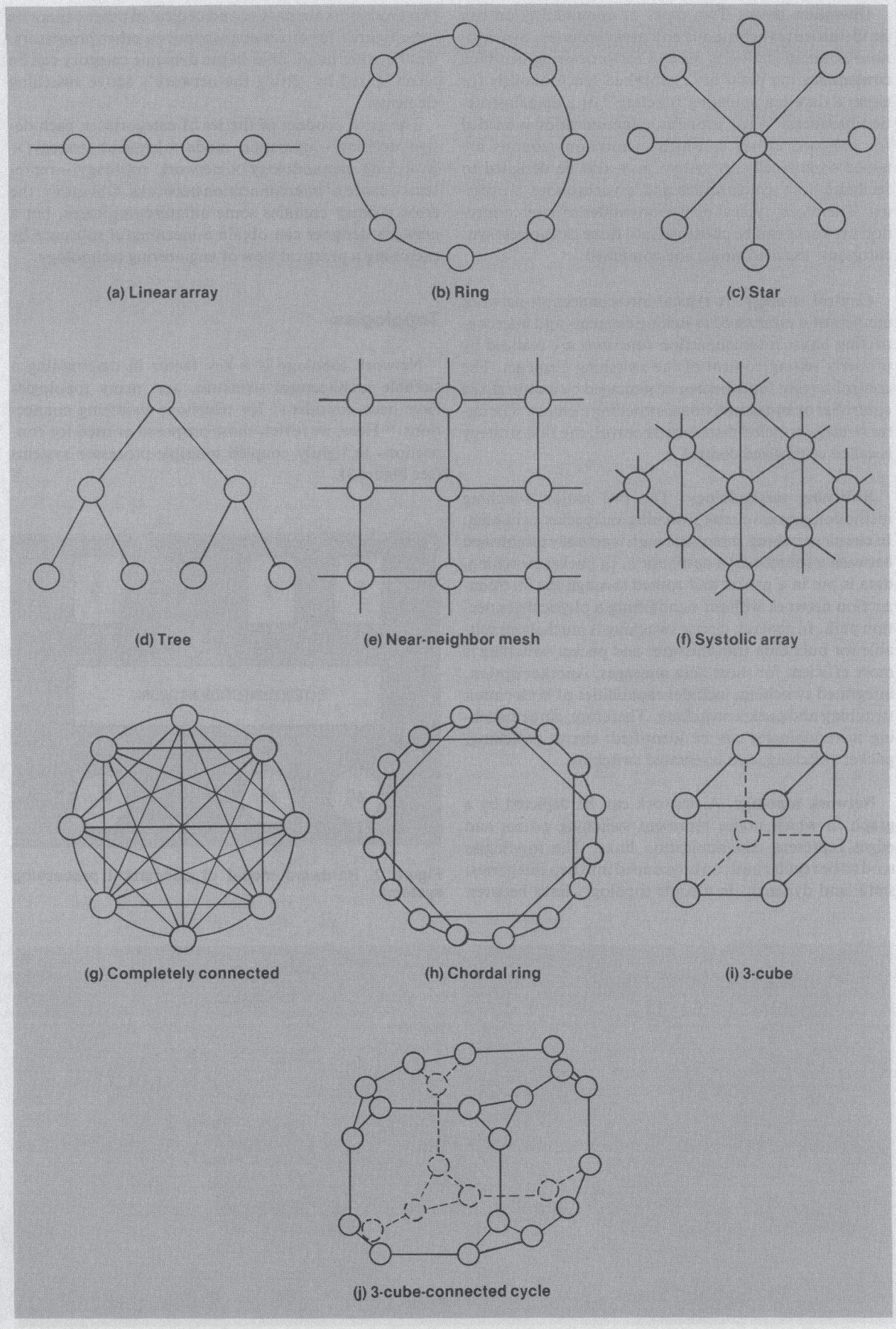


Figure 4. Examples of static network toplogies: (a) one dimensional; (b-f) two dimensional; and (g-j) three dimensional.

Static. Topologies in the static category can be classified according to dimensions required for layout —specifically, one-dimensional, two-dimensional, three-dimensional, and hypercube as shown in Figure 3. Examples of one-dimensional topologies include the linear array used for some pipeline architectures (Figure 4a).[13] Two-dimensional topologies include the ring,[14,15] star,[16] tree,[17] near-neighbor mesh,[18] and systolic array.[13] Examples are shown in Figure 4b-f. Three-dimensional topologies include the completely connected,[19] chordal ring,[20] 3-cube,[21] and 3-cube-connected-cycle[22] networks depicted in Figure 4g-j. A D-dimensional, W-wide hypercube contains W nodes in each dimension, and there is a connection to a node in each dimension. The near-neighbor mesh and the 3-cube are actually two- and three-dimensional hypercubes, respectively. The cube-connected-cycle is a deviation of the hypercube. For example, the 3-cube-connected-cycle shown in Figure 4j is obtained by replacing each node of the 3-cube by a 3-node cycle. Each node in the cycle is connected to the corresponding node in another cycle.

Dynamic. There are three topological classes in the dynamic category: single-stage, multistage, and crossbar (see Figure 5).

Single-stage. A single-stage network is composed of a stage of switching elements cascaded to a link connection pattern. The shuffle-exchange network[23] is a single-stage network based on a perfect-shuffle connection cascaded to a stage of switching elements as shown in Figure 5a. The single-stage network is also called a recirculating network because data items may have to recirculate through the single stage several times before reaching their final destination.

Multistage. A multistage network consists of more than one stage of switching elements and is usually capa-

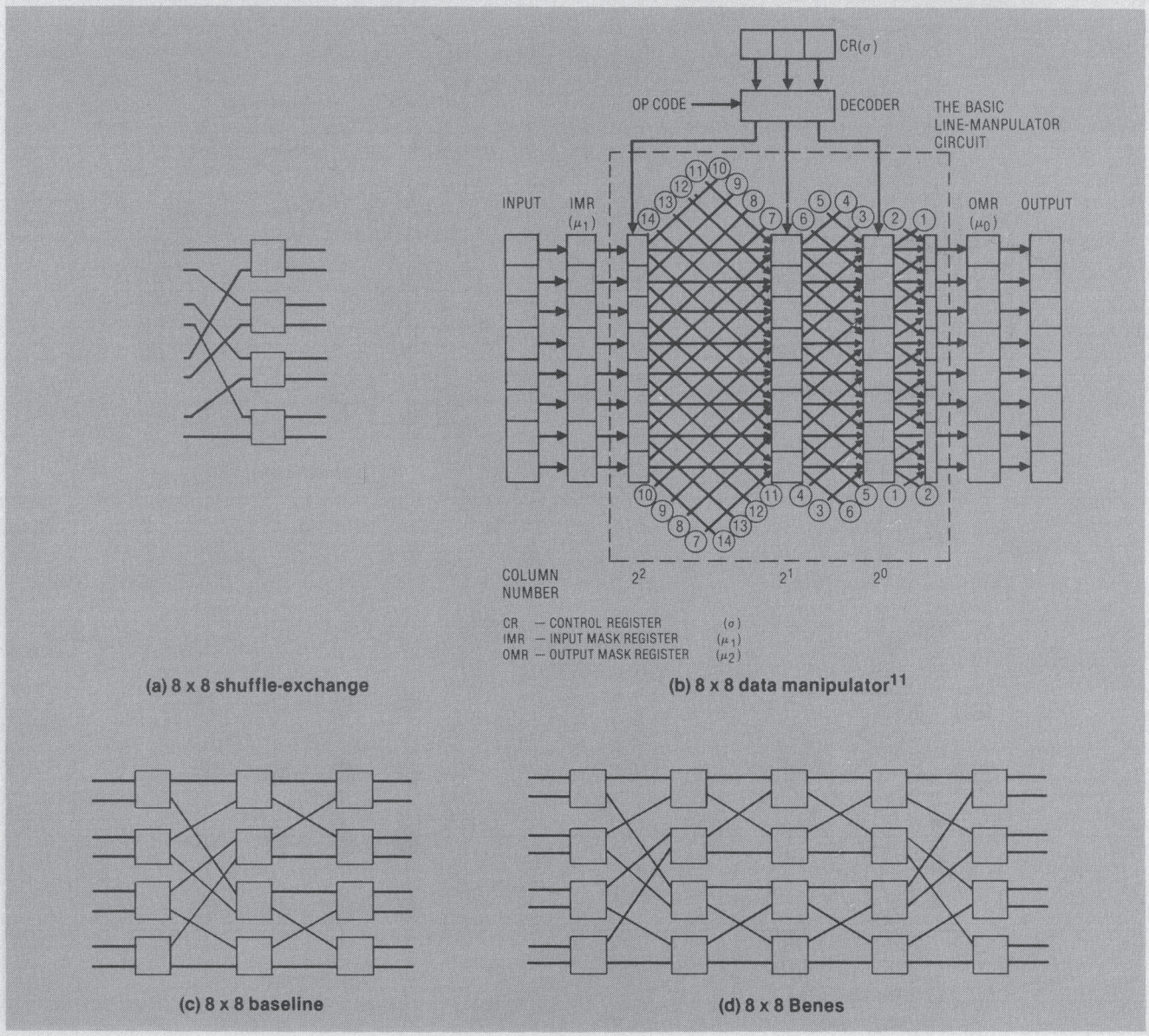
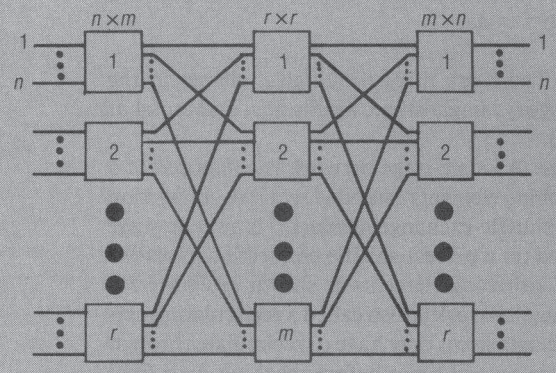
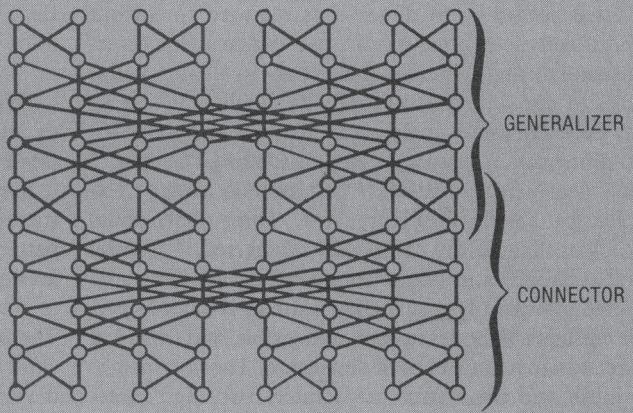


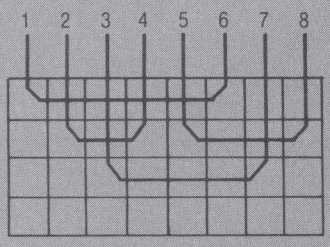
Figure 5. Examples of dynamic network topologies: (a) single stage; (b-i) multistage; and (j) crossbar. **(Cont'd on p. 16.)**



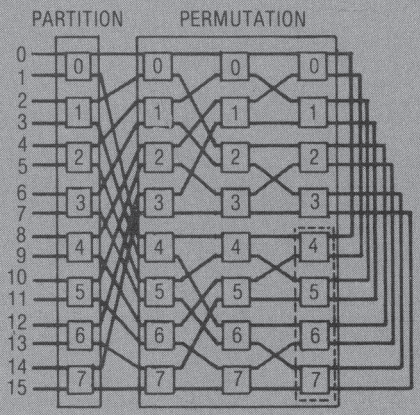
(e) $m \geq 2n - 1$ Clos[38]



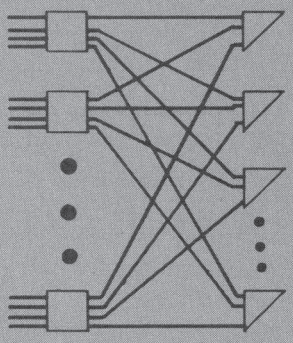
(f) One-to-many nonblocking[39]



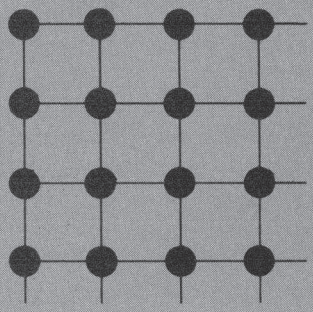
(g) One-sided cellular



(h) One-sided baseline



(i) One-sided Clos



(j) Crossbar

Figure 5 (cont'd from p.15). Examples of multistage and crossbar (j) dynamic network topologies.

ble of connecting an arbitrary input terminal to an arbitrary output terminal. Multistage networks can be one-sided or two-sided. The one-sided networks, sometimes called full switches, have input-output ports on the same side. The two-sided multistage networks, which usually have an input side and an output side, can be divided into three classes: blocking, rearrangeable, and nonblocking.

In blocking networks, simultaneous connections of more than one terminal pair may result in conflicts in the use of network communication links. Examples of this type of network, which has been extensively investigated, include data manipulator,[24] baseline,[25,26] SW banyan,[27] omega,[28] flip,[29] indirect binary n-cube,[30] and delta.[31] A topological equivalence relationship has been established for this class of networks in terms of the baseline network.[25,26] A data manipulator and a baseline network are shown in Figure 5b and 5c.

A network is called a rearrangeable nonblocking network if it can perform all possible connections between inputs and outputs by rearranging its existing connections so that a connection path for a new input-output pair can always be established. A well-defined network, the Benes network[12] shown in Figure 5d, belongs to this class. The Benes rearrangeable network topology has been extensively studied for use in synchronous data permutation[32-35] and asynchronous interprocessor communication.[36,37]

A network which can handle all possible connections without blocking is called a nonblocking network. Two cases have been considered in the literature. In the first case, the Clos network[38] shown in Figure 5e, a one-to-one connection is made between an input and an output. The other case considers one-to-many connections.[39] Here, a generalized-connection network topology is generated to pass any of the N^N mapping of inputs onto outputs where N is the number of inputs or outputs (see Figure 5f). In a one-sided network (or full switch), one-to-one connection is possible between all pairs of terminals.[40,41] A cellular implementation, a base-line topology construction, and a Clos construction are shown in Figure 5g-i.

Crossbar. In a crossbar switch every input port can be connected to a free output port without blocking. Figure 5j shows a schematic which is similar to one used in C.mmp.[42] A crossbar switch called a versatile line manipulator has also been designed and implemented.[43,44]

Communication protocols

The switching methodology and the control strategy are implemented in switching elements (or switching points) according to required communication protocols. The communication protocols can be viewed on two levels. The first level concerns switching control algorithms which generate necessary control settings on switching elements to ensure reliable data routings from source to destination. The first-level protocols are referred to as routing techniques here. The second level is concerned with the link control procedure that provides the handshaking process among switching points. The handshaking process is a basic function implemented by switching elements.

Routing techniques. The routing techniques depend on the network topology and the operation mode used. More or less, each multiple-processor system needs a routing algorithm. Here, we use several well-defined routing algorithms for examples.

Near-neighbor mesh. Bitonic sort has been adapted by several authors[45-47] for the routing of an $n \times n$ mesh-connected, single instruction-multiple data stream system. The procedure developed by Nassimi[47] is as follows:

Procedure SORT (n,n)
1) K ← S ← 1
2) **While** K < n **do**
   a) consider the n × n processor array as composed of many adjacent K × 2K subarrays
   b) **do** in parallel for each K × 2K array HORIZONTAL_MERGE(K, 2K)
   c) S ← S + 1
   d) Consider the n × n processor array as composed of many adjacent 2K × 2K subarrays
   e) **do** in parallel for each 2K × 2K subarray VERTICAL_MERGE(2K, 2K)
   f) S ← S + 1; K ← 2 * K

end
end SORT

The HORIZONTAL_MERGE sorts a bitonic sequence arranged in two arrays with the increasing sequence on the

TERMINALS FROM TRANSNET

PURCHASE PLAN • 12-24 MONTH FULL OWNERSHIP PLAN • 36 MONTH LEASE PLAN

	DESCRIPTION	PURCHASE PRICE	PER MONTH 12 MOS.	PER MONTH 24 MOS.	PER MONTH 36 MOS.
DEC	LA36 DECwriter II	$1,095	$105	$ 58	$ 40
	LA34 DECwriter IV	995	95	53	36
	LA34 DECwriter IV Forms Ctrl.	1,095	105	58	40
	LA120 DECwriter III KSR	2,295	220	122	83
	LA120 DECwriter III RO	2,095	200	112	75
	VT100 CRT DECscope	1,695	162	90	61
	VT101 CRT DECscope	1,195	115	67	43
	VT125 CRT Graphics	3,295	315	185	119
	VT131 CRT DECscope	1,745	167	98	63
	VT132 CRT DECscope	1,995	190	106	72
	VT18XAC Personal Computer Option	2,495	240	140	90
TEXAS INSTRUMENTS	TI745 Portable Terminal	1,595	153	85	58
	TI765 Bubble Memory Terminal	2,595	249	138	93
	TI Insight 10 Terminal	695	67	37	25
	TI785 Portable KSR, 120 CPS	2,395	230	128	86
	TI787 Portable KSR, 120 CPS	2,845	273	152	102
	TI810 RO Printer	1,695	162	90	61
	TI820 KSR Printer	2,195	211	117	80
LEAR SIEGLER	ADM3A CRT Terminal	595	57	34	22
	ADM5 CRT Terminal	645	62	36	24
	ADM32 CRT Terminal	1,165	112	65	42
	ADM42 CRT Terminal	1,995	190	106	72
DATAMEDIA	DT80/1 CRT Terminal	1,695	162	90	61
	DT80/3 CRT Terminal	1,295	125	70	48
	DT80/5L APL 15" CRT	2,295	220	122	83
TELEVIDEO	920 CRT Terminal	895	86	48	32
	950 CRT Terminal	1,075	103	57	39
NEC SPINWRITER	Letter Quality, 7715 RO	2,895	278	154	104
	Letter Quality, 7725 KSR	3,295	316	175	119
GENERAL ELECTRIC	2030 KSR Printer 30 CPS	1,195	115	67	43
	2120 KSR Printer 120 CPS	2,195	211	117	80
HAZELTINE	Executive 80/20	1,345	127	75	49
	Executive 80/30	1,695	162	90	61
EPSON	MX-80 F/T Printer	745	71	42	27
	MX-100 Printer	895	86	48	32

FULL OWNERSHIP AFTER 12 OR 24 MONTHS • 10% PURCHASE OPTION AFTER 36 MONTHS

MICROCOMPUTERS
APPLE • COMMODORE • HP85 • DEC LSI 11

ACCESSORIES AND PERIPHERAL EQUIPMENT
ACOUSTIC COUPLERS • MODEMS • THERMAL PAPER • RIBBONS • INTERFACE MODULES • FLOPPY DISK UNITS

TRANSNET CORPORATION
1945 ROUTE 22 • UNION, N.J. 07083 • (201) 688-7800
TWX 710-985-5485

Reader Service Number 6 ►

left array and the decreasing sequence on the right array.

```
ERROR: ioerror
OFFENDING COMMAND: image

STACK:

-mark-
-savelevel-
-mark-
-savelevel-
```