Interconnect Network

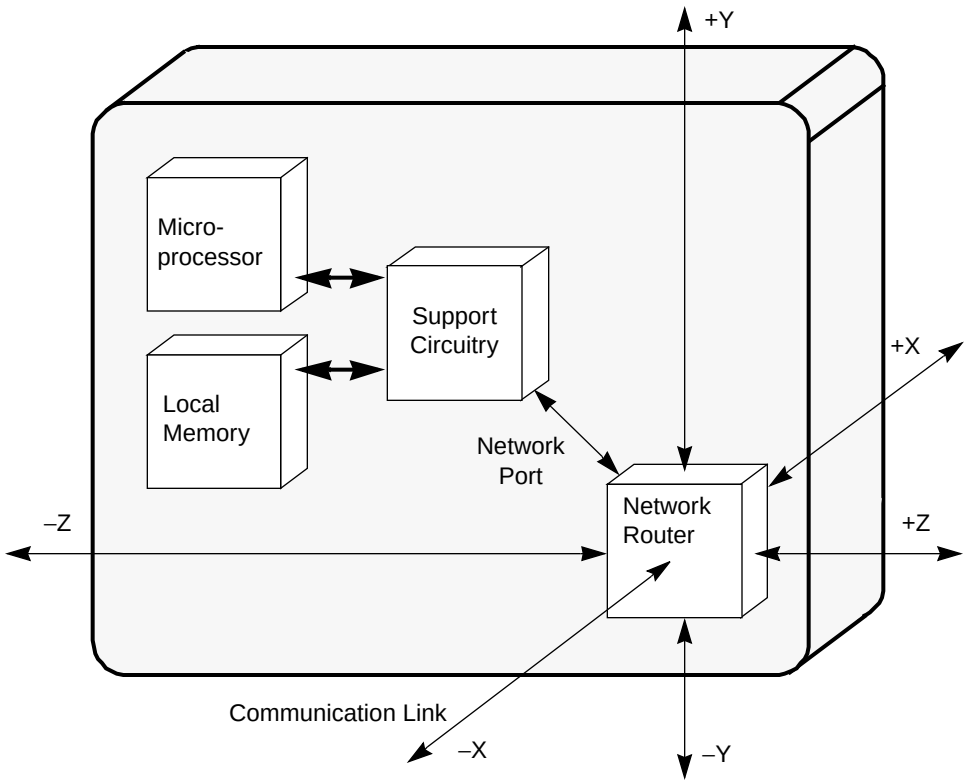
HMM-146-0

Interconnect Network Components

In the CRAY T3E system, the interconnect network provides communication among the PEs. The interconnect network contains three types of components: network ports, network routers, and communication links.

Figure 1 shows how a network port, a network router, and six communication links connect to one PE. The PE and interconnect network components shown in Figure 1 are also collectively referred to as a node.

Figure 1. PE and Interconnect Network Components



Network Port

The network port is a pair of unidirectional channels that connect the support circuitry to a network router. Each PE uses a network port to send and receive packets of information over the interconnect network.

NOTE: With respect to the network router, the network port is also referred to as the PE port.

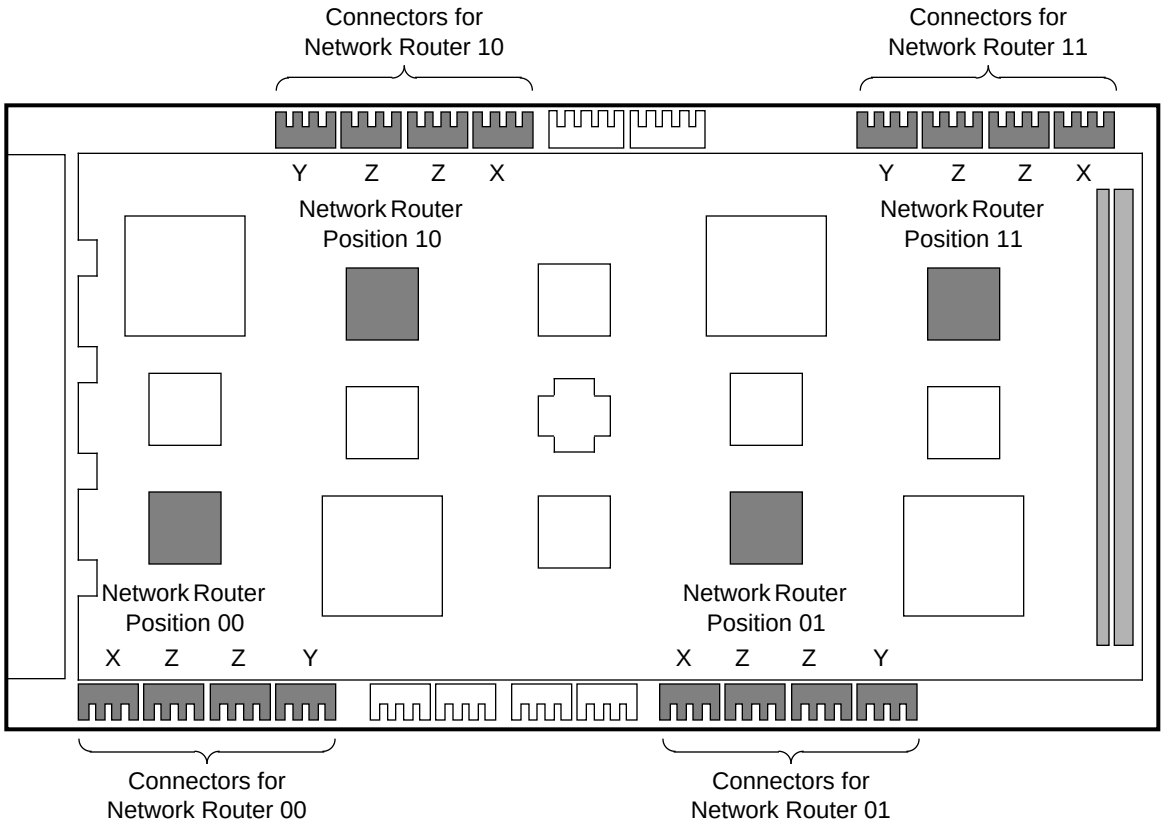
Network Routers

Network routers steer data and control information from a communication link or the network port to another communication link or the network port.

Network Router Physical Locations

Each printed circuit board (PCB) in a CRAY T3E system contains four nodes. Each node has one network router integrated circuit and four connectors for communication links that attach to the network router (refer to Figure 2).

Figure 2. Network Router Physical Locations on a PCB

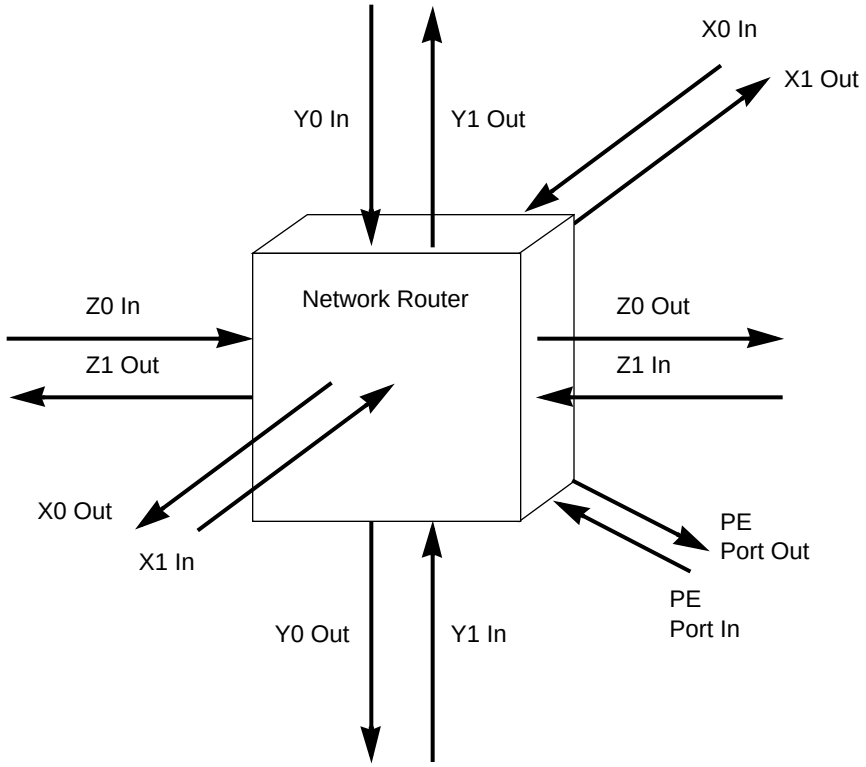


NOTE: Figure 2 shows the side of the PCB that has the integrated circuits. This side of the PCB mounts onto the cold plate of a module. Unless otherwise noted, figures throughout this document show this side of the PCB.

Network Router Ports

In addition to a PE port, each network router has six network router ports: X0 port, X1 port, Y0 port, Y1 port, Z0 port, and Z1 port. Each network router port has an In and an Out (refer to Figure 3).

Figure 3. Network Router Ports

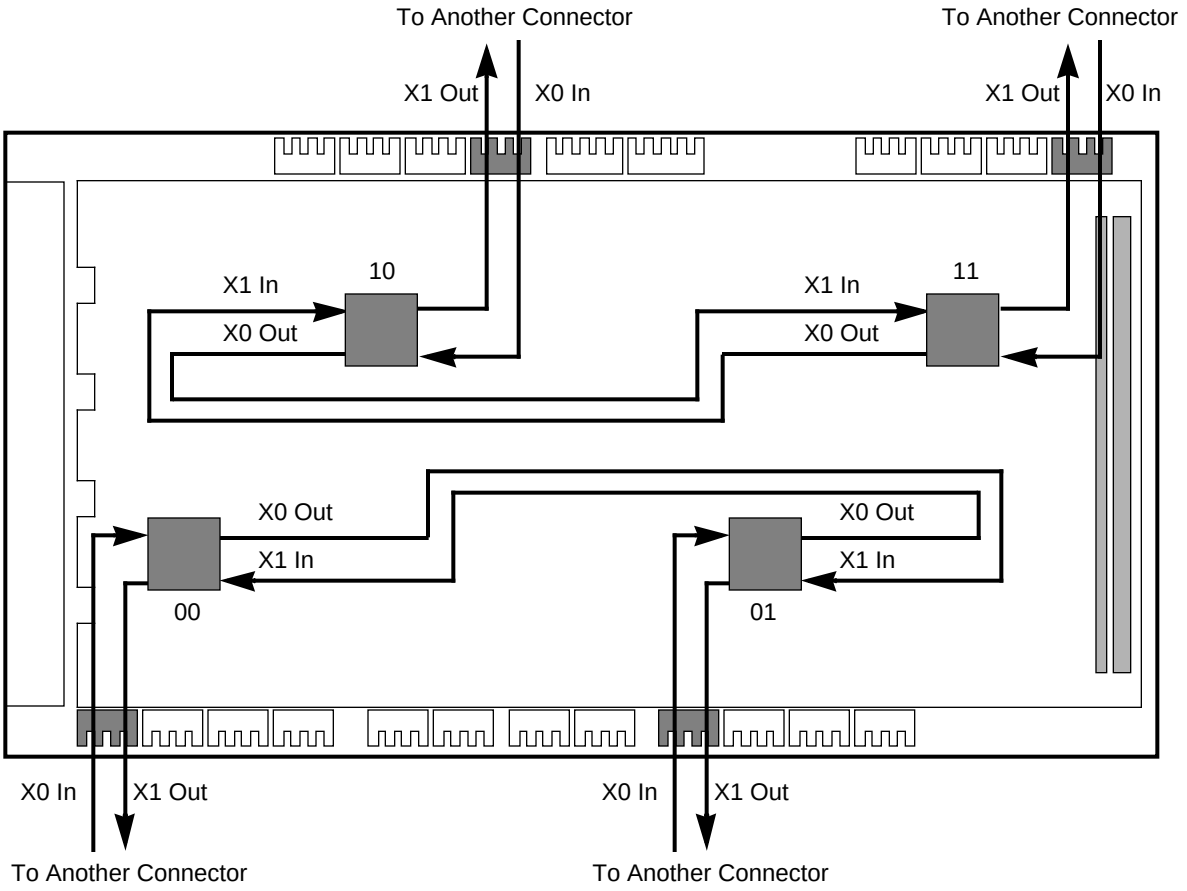


When the Out of a network router port connects to the In of a network router port in the same dimension, but in another network router, a unidirectional channel is created. For example, when the Z0 Out of a network router connects to the Z0 In or Z1 In of another network router, this creates a unidirectional channel.

X-dimension Network Router Port Connections

Figure 4 shows the X-dimension port connections for one PCB. In the X dimension, the In direction of one port (X0 or X1) combined with the Out direction of the other port (X1 or X0) creates a unidirectional channel. For example, the X0 Out port of network router 00 and the X1 In port of network router 01 create a unidirectional channel that transfers information from network router 00 to network router 01.

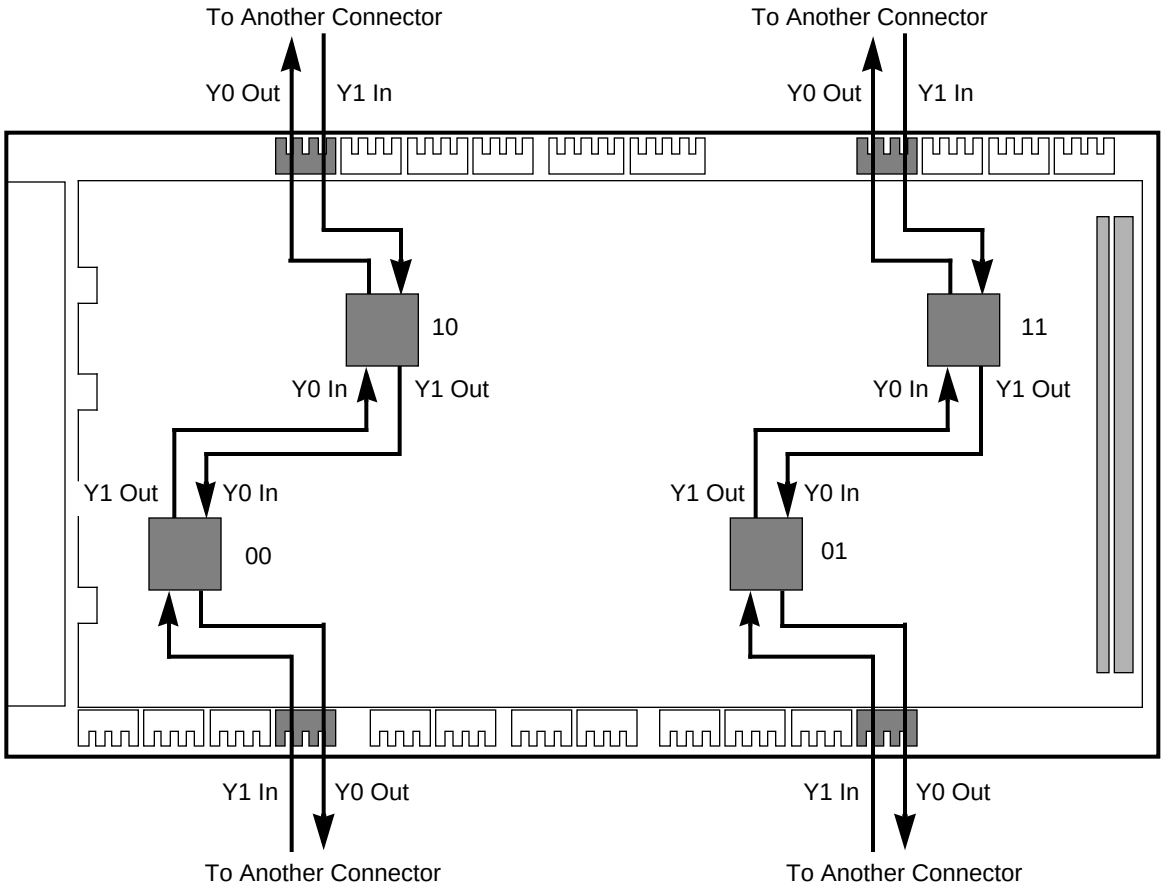
Figure 4. X-dimension Network Router Port Connections for One PCB



Y-dimension Network Router Port Connections

Figure 5 shows the Y-dimension port connections for one PCB. In the Y dimension, the In direction of one port (Y0 or Y1) combined with the Out direction of the other port (Y1 or Y0) creates a unidirectional channel. For example, the Y1 Out port of network router 00 and the Y0 In port of network router 10 create a unidirectional channel that transfers information from network router 00 to network router 10.

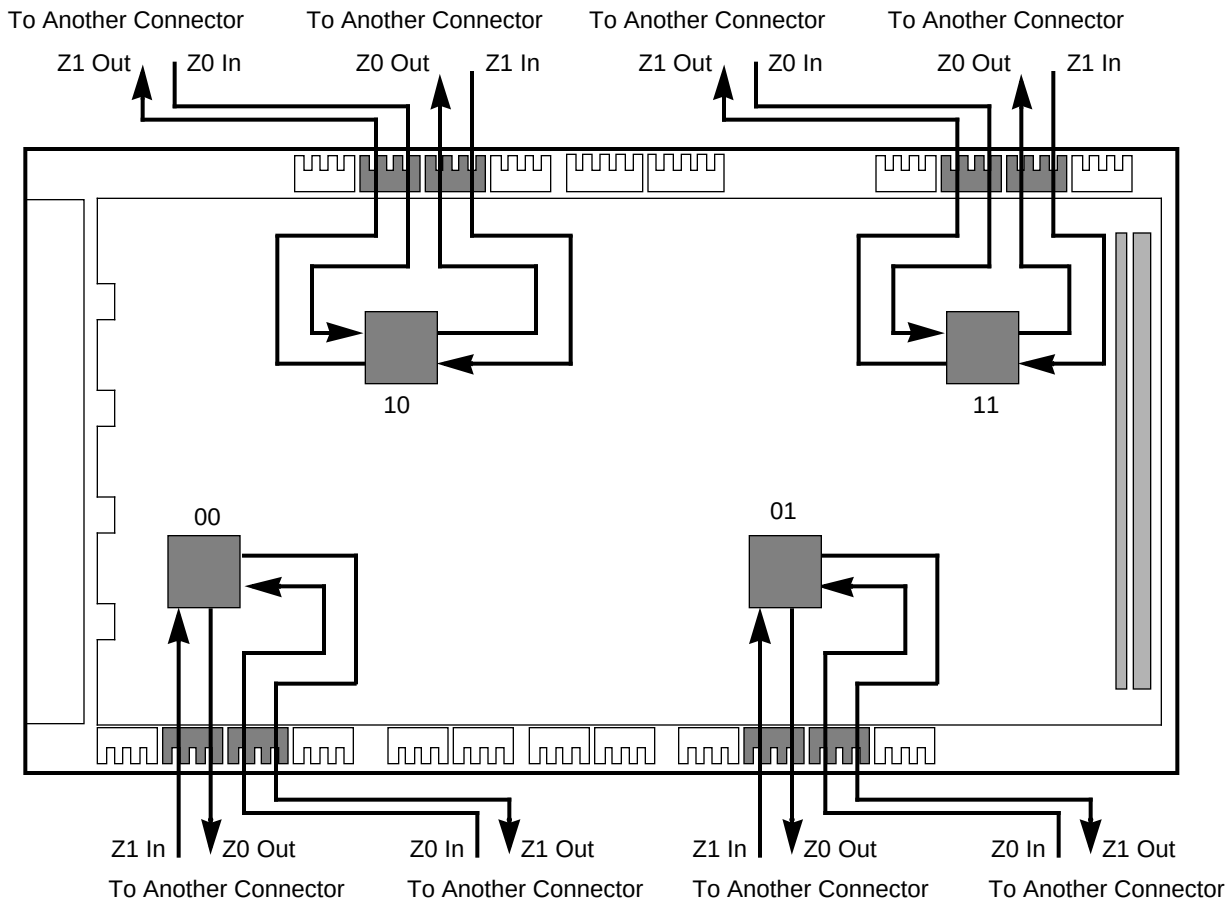
Figure 5. Y-dimension Network Router Port Connections for One PCB



Z-dimension Network Router Port Connections

Figure 6 shows the Z-dimension port connections for one PCB. In the Z dimension, the In direction of one port combined with the Out direction of another port creates a unidirectional channel. For example, the Z0 Out port of network router 00 and the Z0 or Z1 In port of a network router on another PCB create a unidirectional channel that transfers information from network router 00 to the other network router.

Figure 6. Z-dimension Network Router Port Connections for One PCB

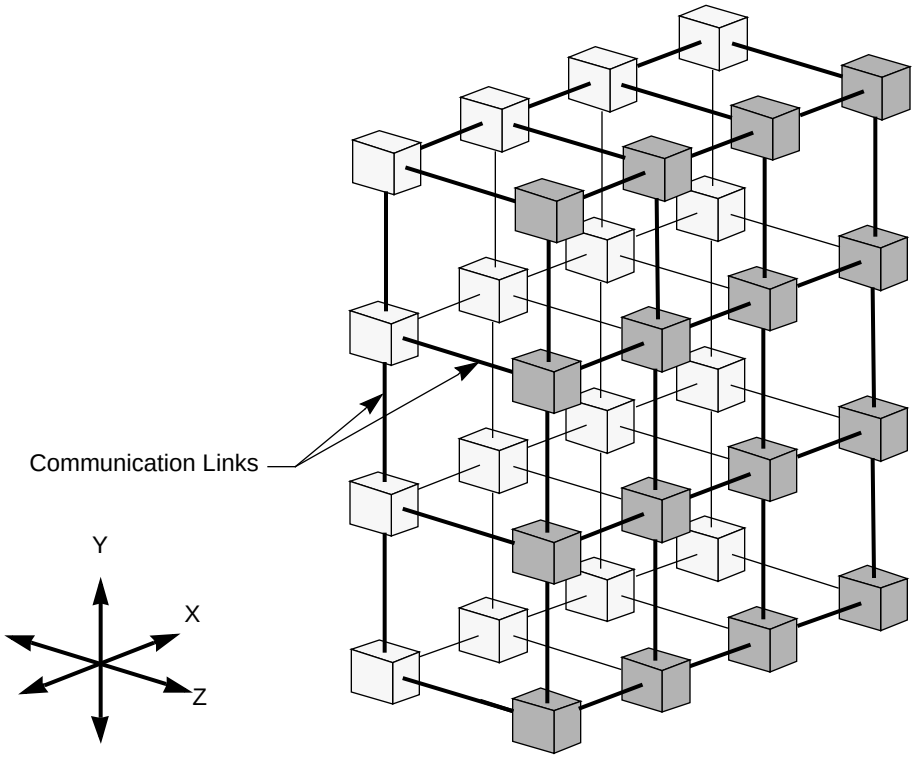


Communication Links

Communication links transfer data and control information between network routers in the interconnect network. Each communication link connects two network routers.

The communication links form a three-dimensional matrix of paths that connect nodes in the X, Y, and Z dimensions (refer to Figure 7). For clarity, Figure 7 shows a simplified interconnect network and may not reflect an actual configuration of the CRAY T3E system.

Figure 7. Communication Links in the Interconnect Network



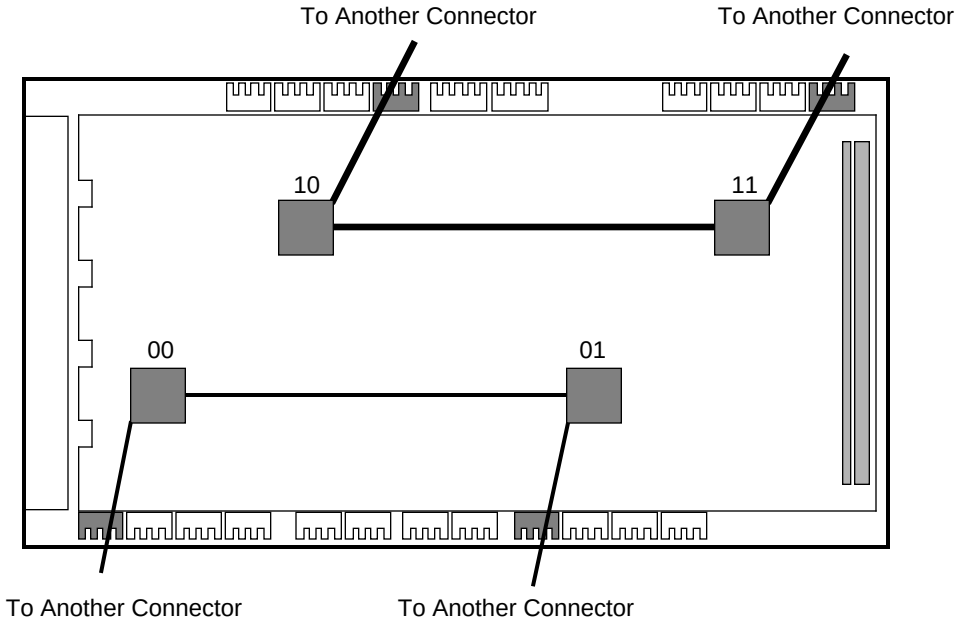
X-dimension Communication Links

Each pair of unidirectional channels that transfer information between two network routers in the X dimension is an X-dimension communication link. Each PCB contains two types of X-dimension communication links: internal and external.

A PCB contains two internal X-dimension communication links. One internal communication link connects network router 00 to network router 01. The other internal communication link connects network router 10 to network router 11 (refer to Figure 8).

A PCB also contains connectors for four external X-dimension communication links. Each network router on the PCB connects to one external X-dimension communication link.

Figure 8. PCB X-dimension Communication Links



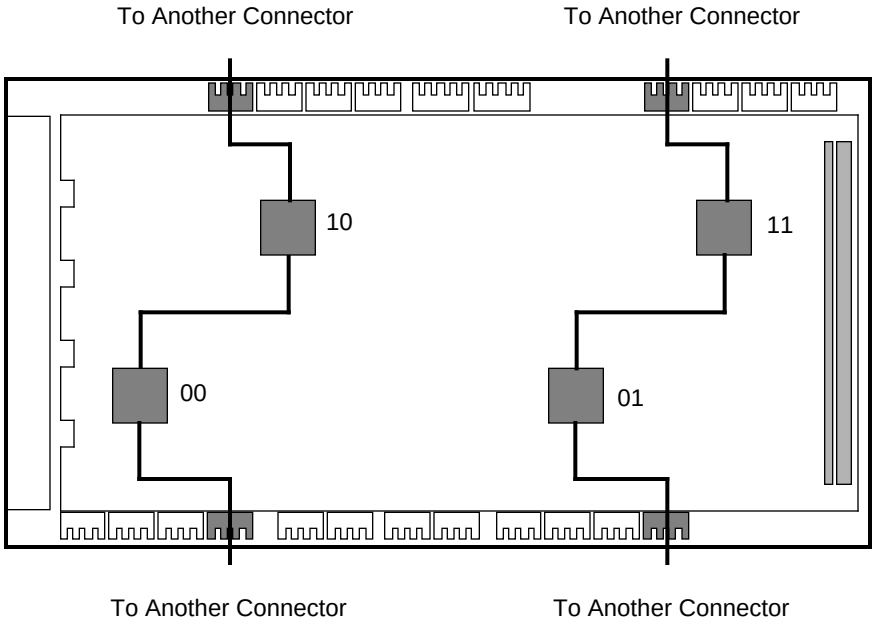
Y-dimension Communication Links

Each pair of unidirectional channels that transfer information between two network routers in the Y dimension is a Y-dimension communication link. Each PCB contains two types of Y-dimension communication links: internal and external.

A PCB contains two internal Y-dimension communication links. One internal communication link connects network router 00 to network router 10. The other internal communication link connects network router 01 to network router 11 (refer to Figure 9).

A PCB also contains connectors for four external Y-dimension communication links. Each network router on the PCB connects to one external Y-dimension communication link.

Figure 9. PCB Y-dimension Communication Links

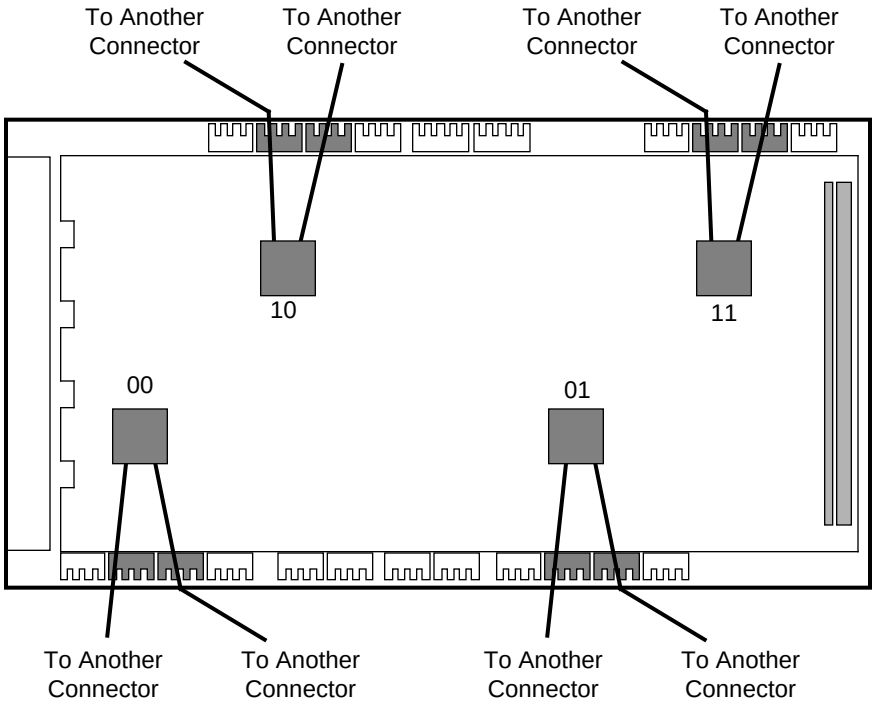


Z-dimension Communication Links

Each pair of unidirectional channels that transfer information between two network routers in the Z dimension is a Z-dimension communication link. Each PCB contains only one type of Z-dimension communication link: an external link.

A PCB contains connectors for eight external Z-dimension communication links. Each network router on the PCB connects to two external Z-dimension communication links (refer to Figure 10).

Figure 10. PCB Z-dimension Communication Links



Direction Assignments

When initializing the system, software assigns a direction to the network router ports. To do this, software writes a value to specified bits of the router orientation (R_ORIENT) register in one physical node. System hardware then automatically sets the R_ORIENT values for all other nodes.

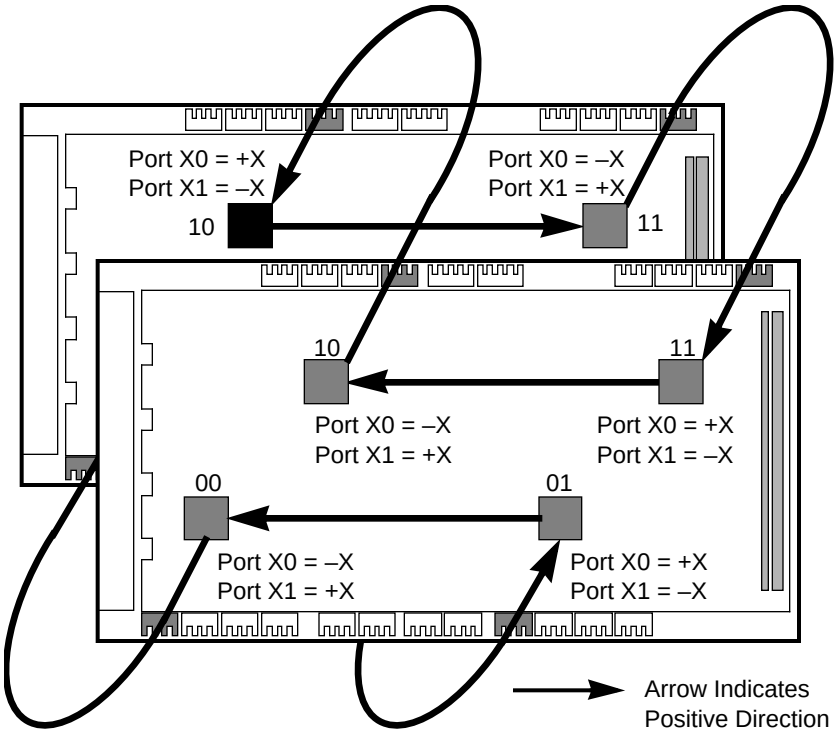
X-dimension Direction Assignments

Bit 0 (PLUS_X bit) of the R_ORIENT register in each node sets the direction for the X-dimension network router ports. Table 1 lists the possible values of the PLUS_X bit and Figure 11 shows sample port direction assignments.

Table 1. PLUS_X bit of R_ORIENT Register

Value	Description
0	X0 port is +X and X1 port is –X
1	X0 port is –X and X1 port is +X

Figure 11. X-dimension Directions



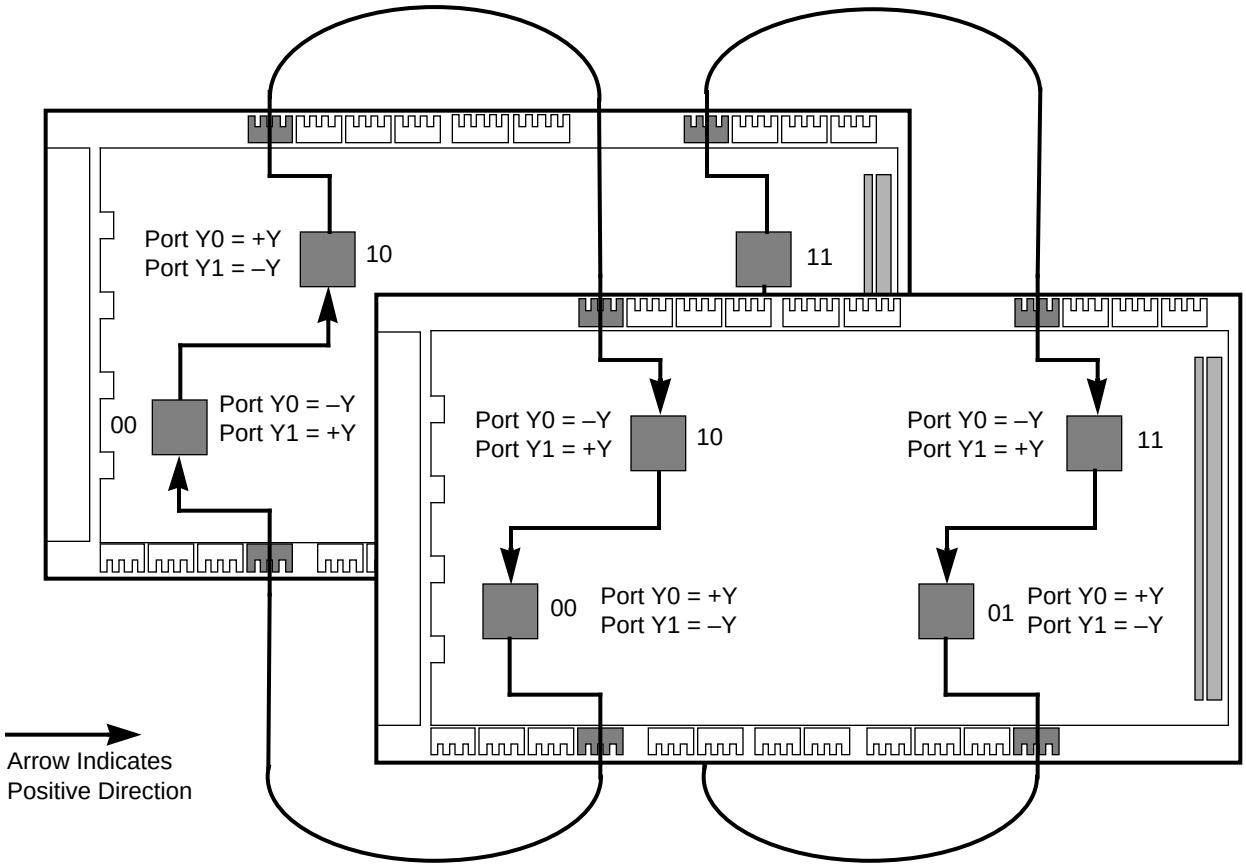
Y-dimension Direction Assignments

Bit 1 (PLUS_Y bit) of the R_ORIENT register in each node sets the direction for the Y-dimension network router ports. Table 2 lists the possible values of the PLUS_Y bit and Figure 12 shows sample port direction assignments.

Table 2. PLUS_Y bit of R_ORIENT Register

Value	Description
0	Y0 port is +Y and Y1 port is –Y
1	Y0 port is –Y and Y1 port is +Y

Figure 12. Y-dimension Directions



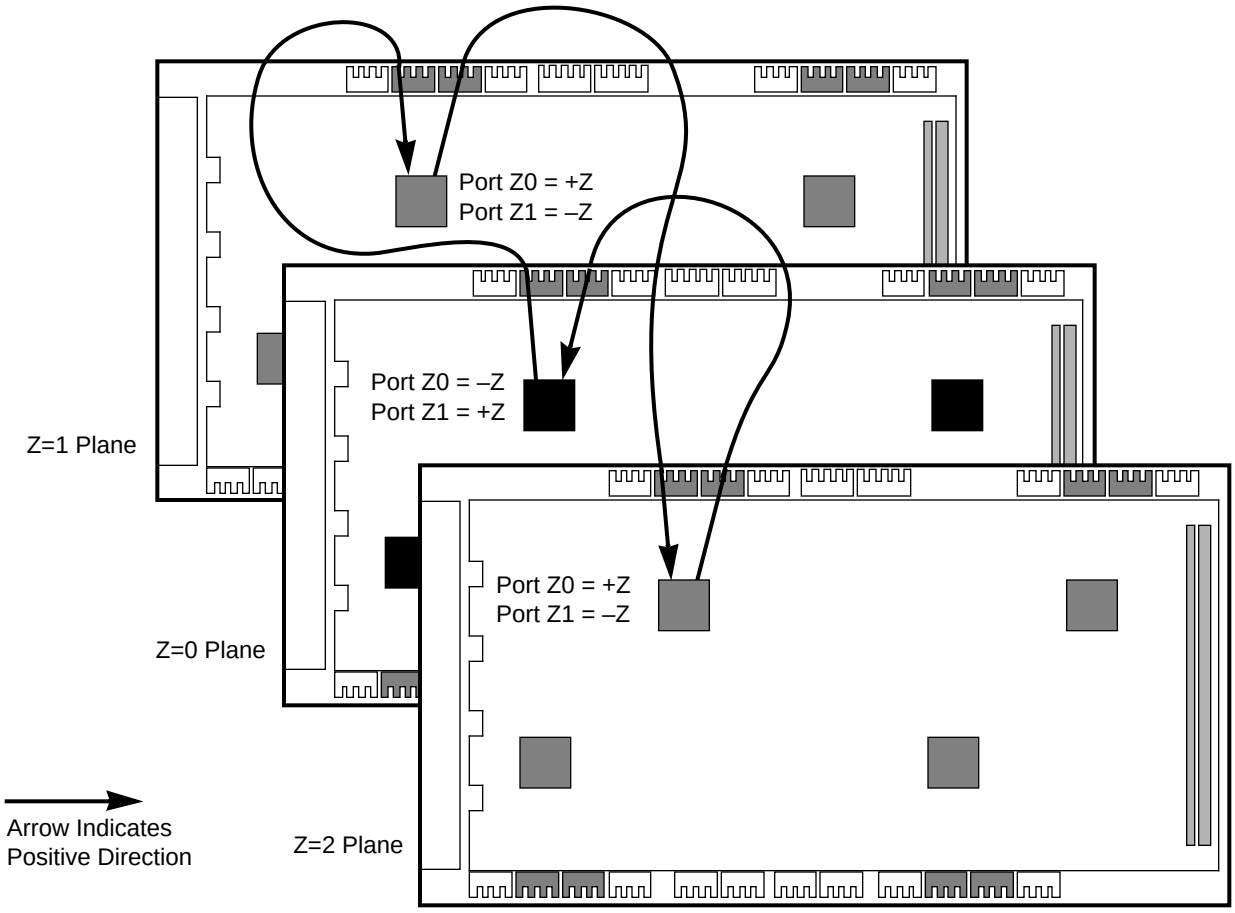
Z-dimension Direction Assignments

Bit 2 (PLUS_Z bit) of the R_ORIENT register in each node sets the direction for the Z-dimension network router ports. Table 3 lists the possible values of the PLUS_Z bit and Figure 13 shows sample port direction assignments.

Table 3. PLUS_Z Bit of R_ORIENT Register

Value	Description
0	Z0 port is +Z and Z1 port is –Z
1	Z0 port is –Z and Z1 port is +Z

Figure 13. Z-dimension Directions



Node and PE Numbers

In the interconnect network, three types of numbers identify nodes and PEs: a physical node number, a logical node (or logical PE) number, and a virtual PE number.

Physical Node Number

When initializing the CRAY T3E system, software assigns every node in the system a unique number that indicates how the node is physically connected in the interconnect network. This number is the physical node number.

R_PWHO Register

To assign a physical number to a node, software writes a value to the network router physical number (R_PWHO) register. Table 4 shows the bit format of the R_PWHO register.

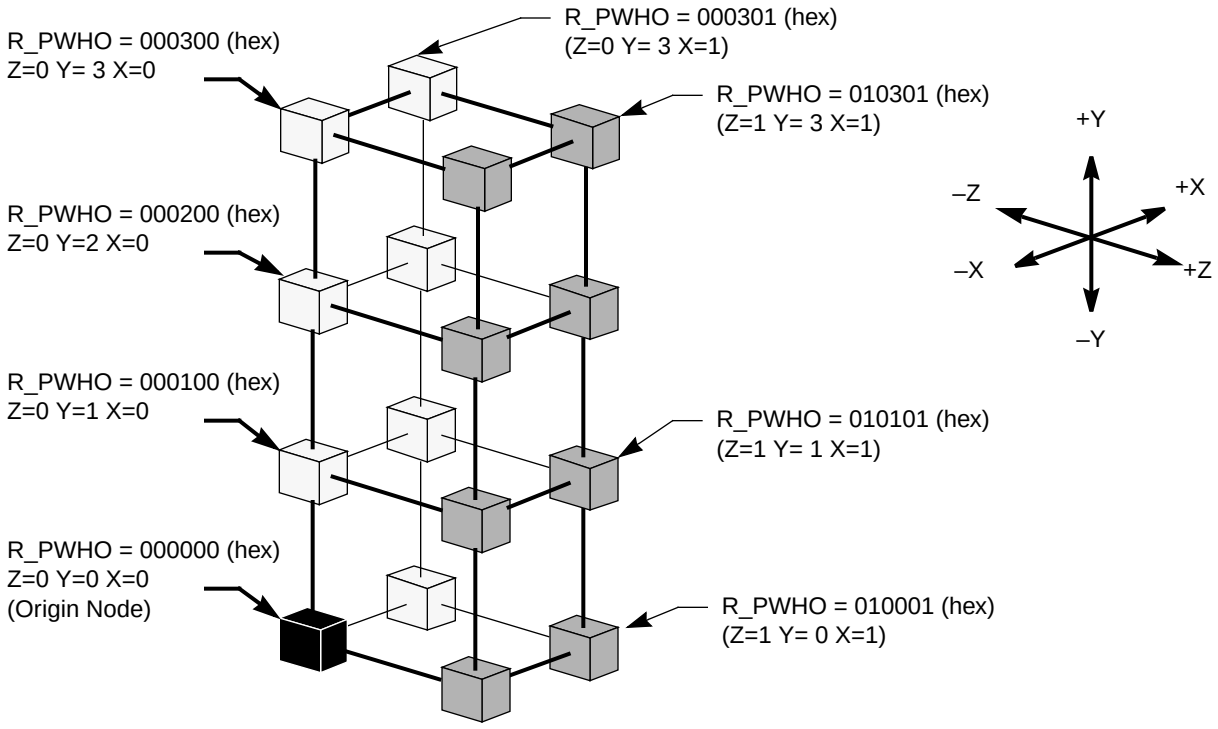
Table 4. R_PWHO Register Bit Format

Bits	Description
<2 : 0>	These bits indicate the position of the node in the X dimension.
<7 : 3>	These bits are not used.
<12 : 8>	These bits indicate the position of the node in the Y dimension.
<15 : 13>	These bits are not used.
<19 : 16>	These bits indicate the position of the node in the Z dimension.
<63 : 20>	These bits are not used.

Sample Physical Node Numbers

Figure 14 shows sample physical node numbers in a CRAY T3E system that has 16 nodes. For clarity, Figure 14 shows a simplified interconnect network and may not reflect an actual configuration of the CRAY T3E system.

Figure 14. Sample Physical Node Numbers



The physical node number indicates the physical position of a node with respect to a node that software designates as the origin (Z=0 Y=0 X=0). For example, the node that software assigns as physical node number 000100 (hex) connects to the origin node through a Y dimension communication link.

Logical Node (or Logical PE) Number

Before running applications, software assigns every node in the system a unique number that indicates how the node is sequentially distributed in the interconnect network. This number is the logical node number. Logical node numbers enable software to sequentially distribute the nodes in the interconnect network in a manner that best suits the user applications.

Software also assigns each PE in the system a logical PE number; however, in the CRAY T3E system, the logical node number and the logical PE number are always set to the same value. Because of this characteristic, the term logical PE number is used exclusively throughout the rest of this and other documents, ah, except for the next paragraph.

R_LWHO and L_WHOAMI Registers

To assign a logical node number, software writes a value to the network router logical number (R_LWHO) register. To assign a logical PE number, software writes a value (the same value as the R_LWHO register) to the logical PE number (L_WHOAMI) register. Table 5 shows the bit format of the R_LWHO and L_WHOAMI registers.

Table 5. R_LWHO and L_WHOAMI Register Bit Format

Bits	Description
<11 : 0>	These bits contain the logical PE (or logical node) number.
<63 : 12>	These bits are not used.

NOTE: The support circuitry uses the L_WHOAMI register when determining if a request is for local or remote memory. The network router uses the R_LWHO register when generating the routing tag for a packet.

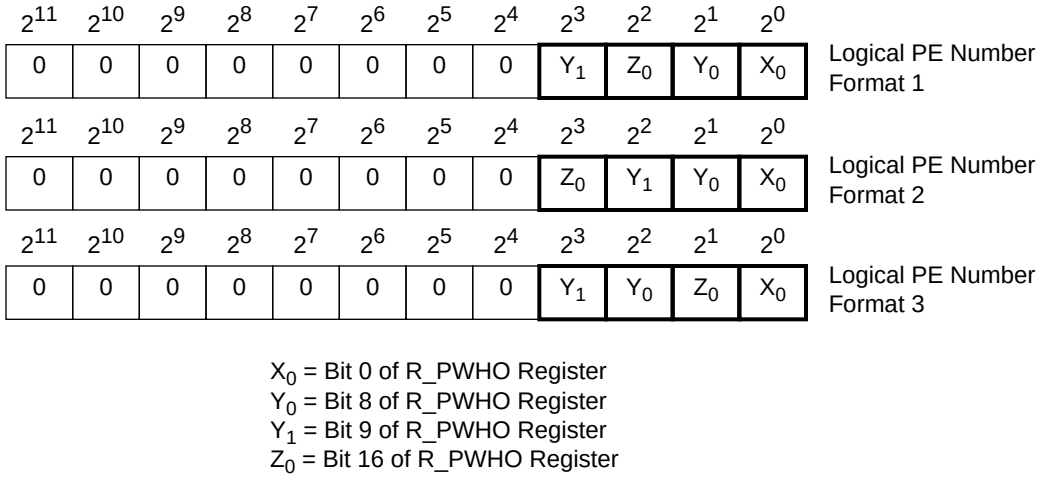
Software uses only as many bits of the logical PE number as are needed for a system. For example, in a 16-PE CRAY T3E system, software uses bits <3 : 0> of the logical PE number to provide a unique logical number for each of the 16 PEs in the system. Software sets the remaining bits to 0's.

The next subsection provides sample formats of the logical PE number. These samples show how different formats of the logical PE number cause different distribution patterns of the logical PE numbers among the physical nodes. These samples may not reflect the actual format that software uses for the logical PE number.

Sample Logical PE Numbers

Figure 15 and Table 6 show sample logical PE number formats and list logical PE numbers for each physical node in a 16-PE CRAY T3E system.

Figure 15. Sample Logical PE Number Formats



X_0 = Bit 0 of R_PWHO Register
Y_0 = Bit 8 of R_PWHO Register
Y_1 = Bit 9 of R_PWHO Register
Z_0 = Bit 16 of R_PWHO Register

Table 6. Sample Logical PE Numbers

R_PWHO (hex)	L_WHOAMI (hex) Format 1	L_WHOAMI (hex) Format 2	L_WHOAMI (hex) Format 3	L_WHOAMI (hex) Random Format
000000	0	0	0	6
000001	1	1	1	C
000100	2	2	4	7
000101	3	3	5	5
000200	8	4	8	2
000201	9	5	9	D
000300	A	6	C	8
000301	B	7	D	1
010000	4	8	2	B
010001	5	9	3	3
010100	6	A	6	E
010101	7	B	7	0
010200	C	C	A	F
010201	D	D	B	4
010300	E	E	E	A
010301	F	F	F	9

Virtual PE Number

When an application initiates, software determines the resources needed for the application and creates a partition for the application. A partition is a group of PEs that are assigned to one application.

The application uses a unique number to reference each PE in a partition. This number is the virtual PE number.

Software defines a partition by indicating a base logical PE number and a virtual PE limit. The base logical PE number is the logical PE number that corresponds to virtual PE number 0 (this is the smallest-numbered logical PE in the partition). The virtual PE limit is the largest-numbered virtual PE in the partition.

As an example, Table 7 lists all of the logical PE numbers in a 16-PE CRAY T3E system and lists sample partitions and virtual PE numbers for the system.

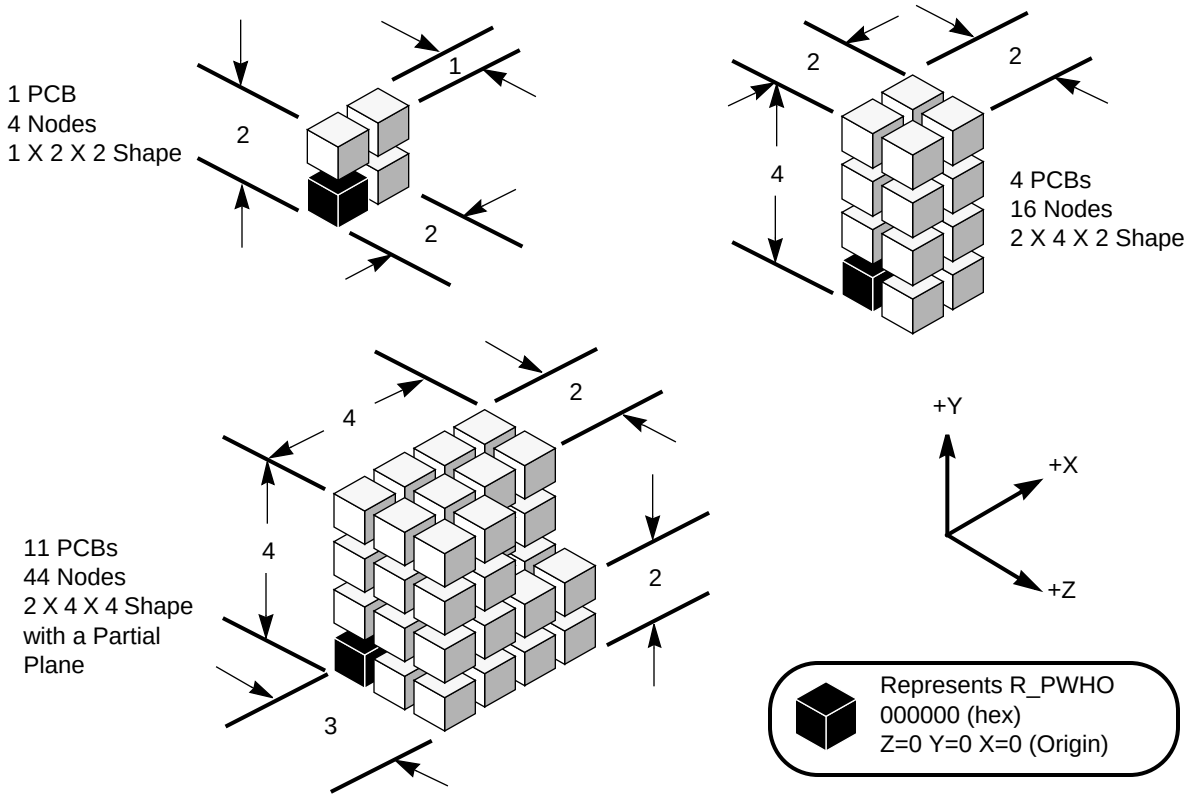
Table 7. Sample Logical PE Numbers

L_WHOAMI (hex)	Virtual PE Numbers	Partition
0	0	Partition A
1	1	Logical PE Base = 0
2	2	Virtual Limit = 3
3	3	
4	0	Partition B
5	1	Logical PE Base = 4
6	2	Virtual Limit = 5
7	3	
8	4	
9	5	
A	Not Applicable	These nodes are presently not running an application.
B	Not Applicable	
C	Not Applicable	
D	0	Partition C
E	1	Logical PE Base = D
F	2	Virtual Limit = 2

Node Shapes

The number of physical nodes in each dimension determines the node shape of a system. For example, a 16-PE CRAY T3E AC system has 2 nodes in the Z dimension, 4 nodes in the Y dimension, and 2 nodes in the X dimension. The node shape for this system is 2 X 4 X 2. Figure 16 shows this and 2 other node shapes of CRAY T3E AC systems. For clarity, Figure 16 does not show the communication links.

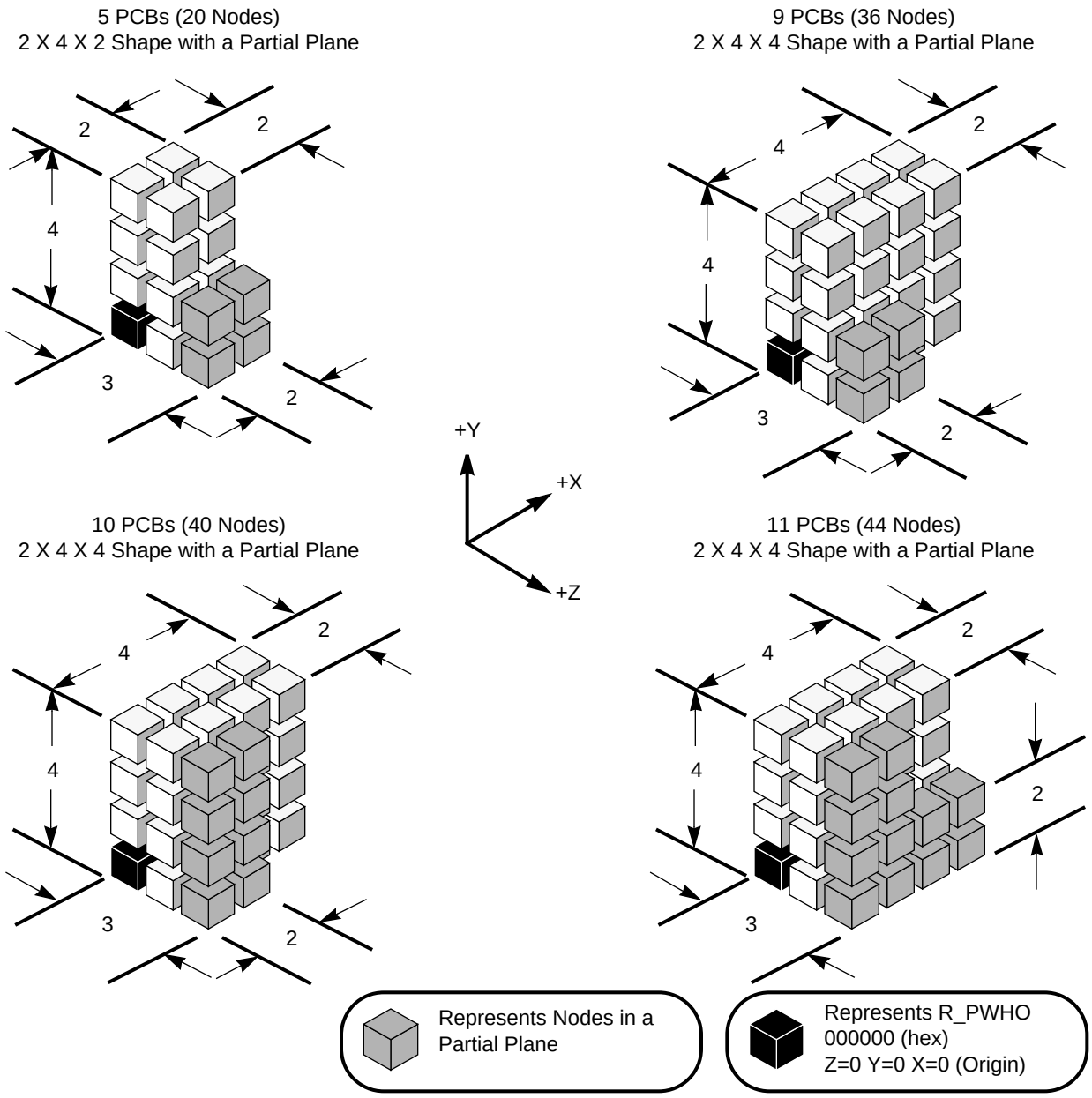
Figure 16. Sample Node Shapes in CRAY T3E AC Systems



The node shapes of several CRAY T3E systems contain partial planes. A partial plane is a set of nodes that connect in the same plane and are less than the maximum number of nodes allowed in the plane. Partial planes are in different locations in AC systems than they are in LC systems.

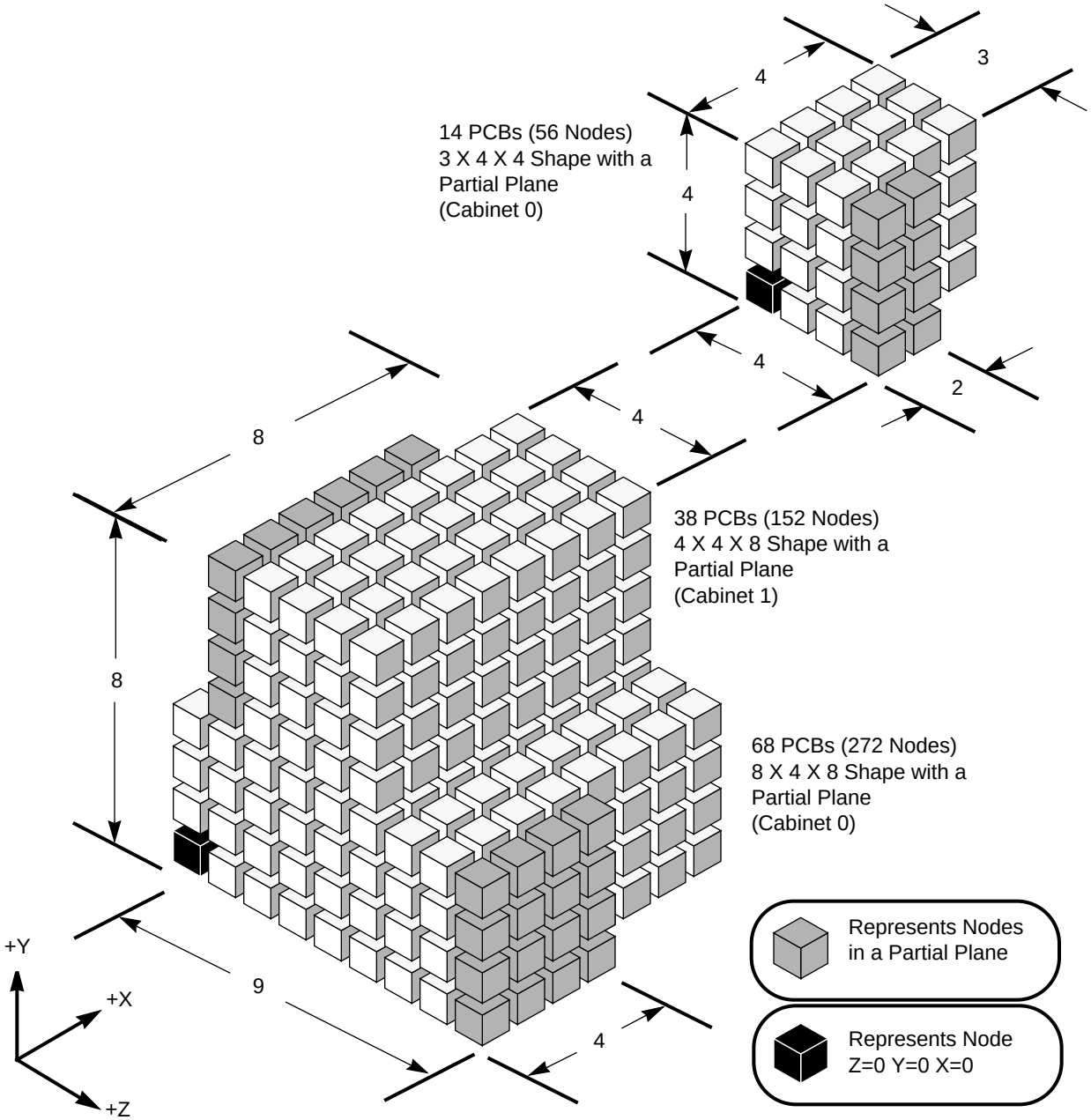
When an AC system contains a partial plane, the partial plane always resides in the largest-numbered Z plane. For example, in a 20-PE CRAY T3E AC system, the partial plane resides in the Z=2 plane (refer to Figure 17).

Figure 17. AC System Partial Planes



When an LC system contains partial planes, the partial planes reside in the largest-numbered Z plane of a full system cabinet and reside in different Z planes in a partially-full system cabinet. For example, in a 424-PE CRAY T3E LC system, the partial planes reside in the Z=8 plane of cabinet 0 and the Z=1 plane of cabinet 1 (refer to Figure 18).

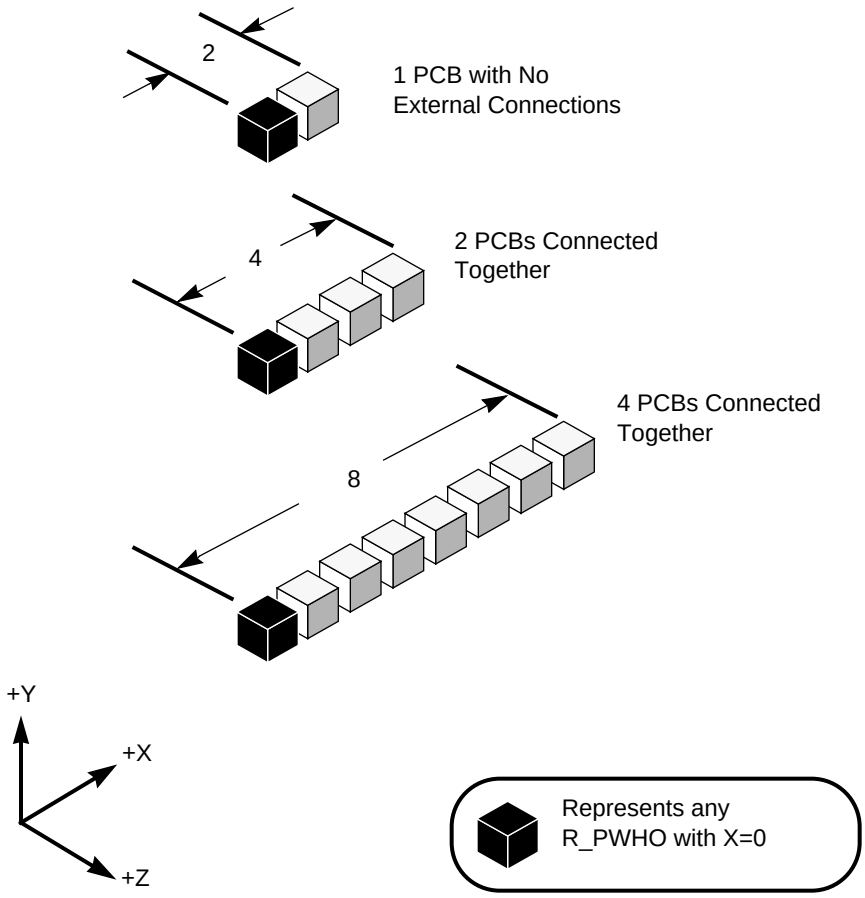
Figure 18. LC System Partial Planes



X-dimension Node Shapes

In the CRAY T3E system, the X-dimension may be one of three shapes: X=2, X=4, or X=8. Figure 19 shows these shapes and lists the numbers of PCBs that connect to form each shape.

Figure 19. X-dimension Node Shapes

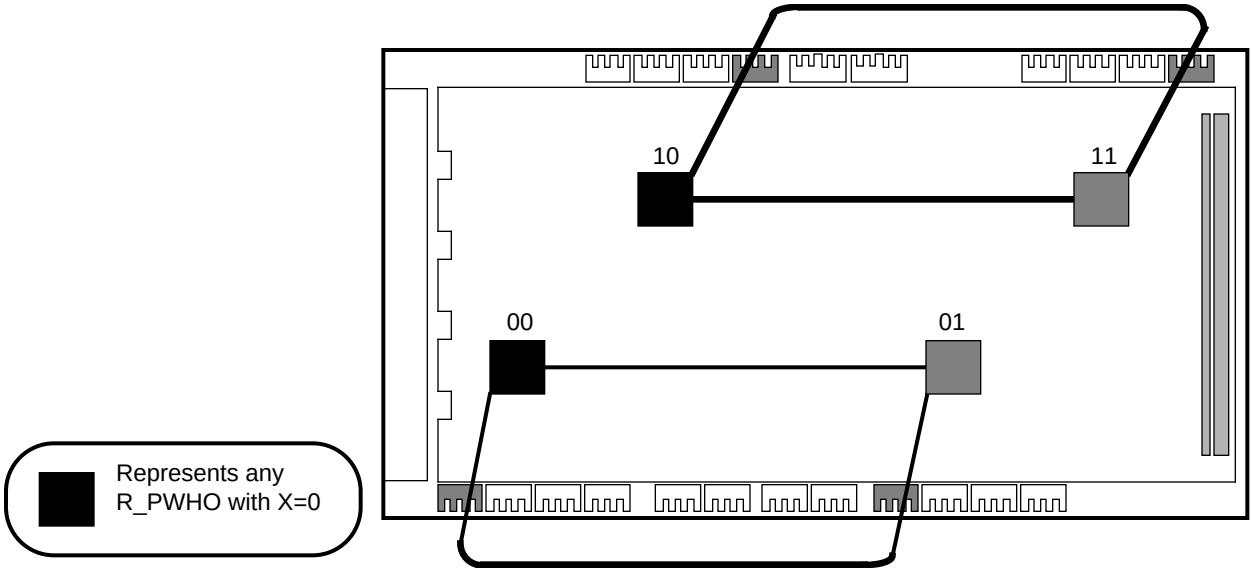


The following subsections show the communication link connections for each X-dimension node shape.

X=2 Communication Link Connections

A PCB that does not connect to another PCB creates an X=2 node shape. The three smallest CRAY T3E AC systems contain one, two, or three PCBs that do not connect to other PCBs. For these systems, the external X-dimension communication links on each PCB connect together. Figure 20 shows these X-dimension connections.

Figure 20. X=2 Communication Link Connections



The external X-dimension communication links connect to form a torus connection. A torus connection connects the smallest-numbered node in a dimension directly to the largest-numbered node in the same dimension. This type of connection forms a ring where information can transfer from one node, through all of the nodes in the same dimension, and back to the original node.

NOTE: In the CRAY T3E AC systems, a single PCB may be used to create a Y=2 and X=2 node shape partial plane. For these systems, the external X-dimension communication links connect together to complete the torus in the X dimension; however, the Y-dimension communication links do not connect together.

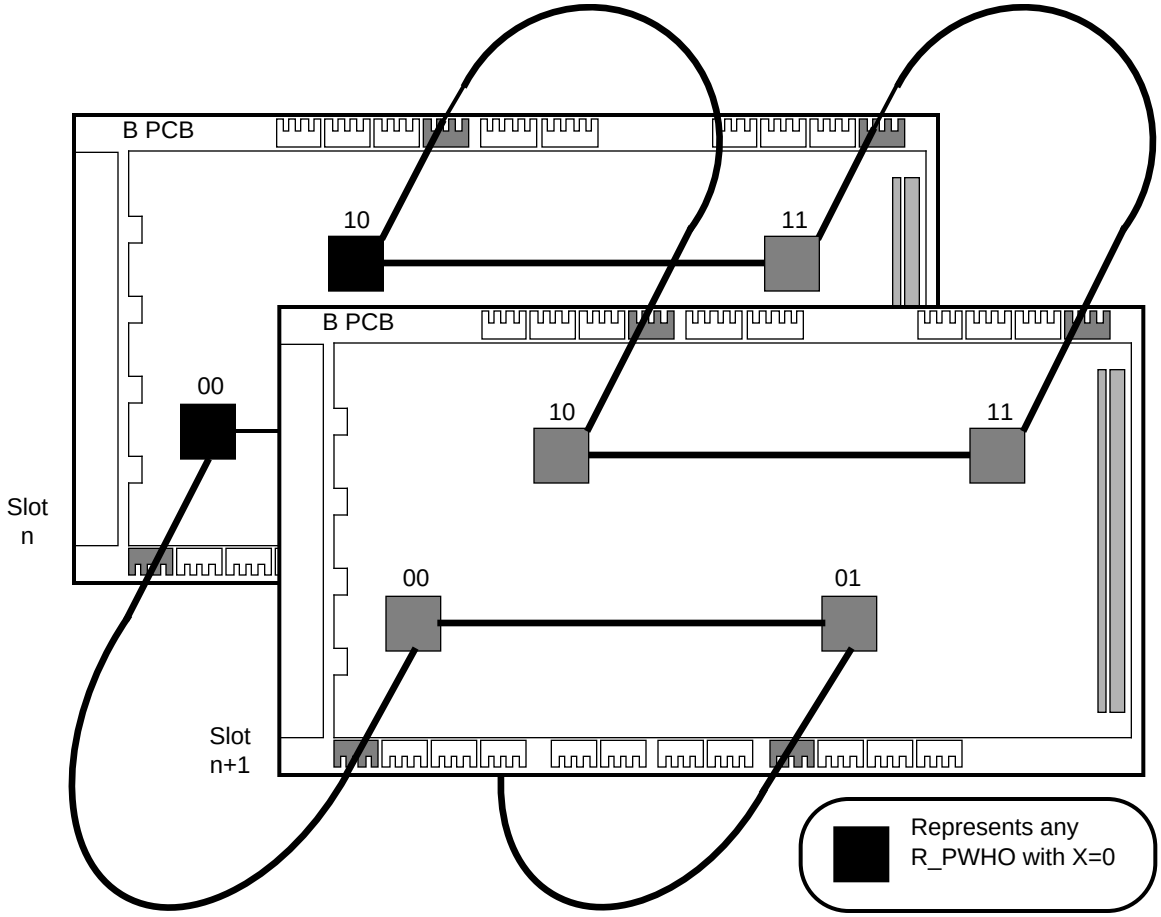
In the CRAY T3E LC systems, a single module (two PCBs) may be used to create a Y=4 and X=2 node shape partial plane. For these systems, the external X-dimension communication links for each PCB connect together to complete the torus in the X dimension.

X=4 Communication Link Connections

Two PCBs connected together create an X=4 node shape. Several CRAY T3E system configurations use this shape. The communication link connections for this shape are different in the LC systems than they are in the AC systems.

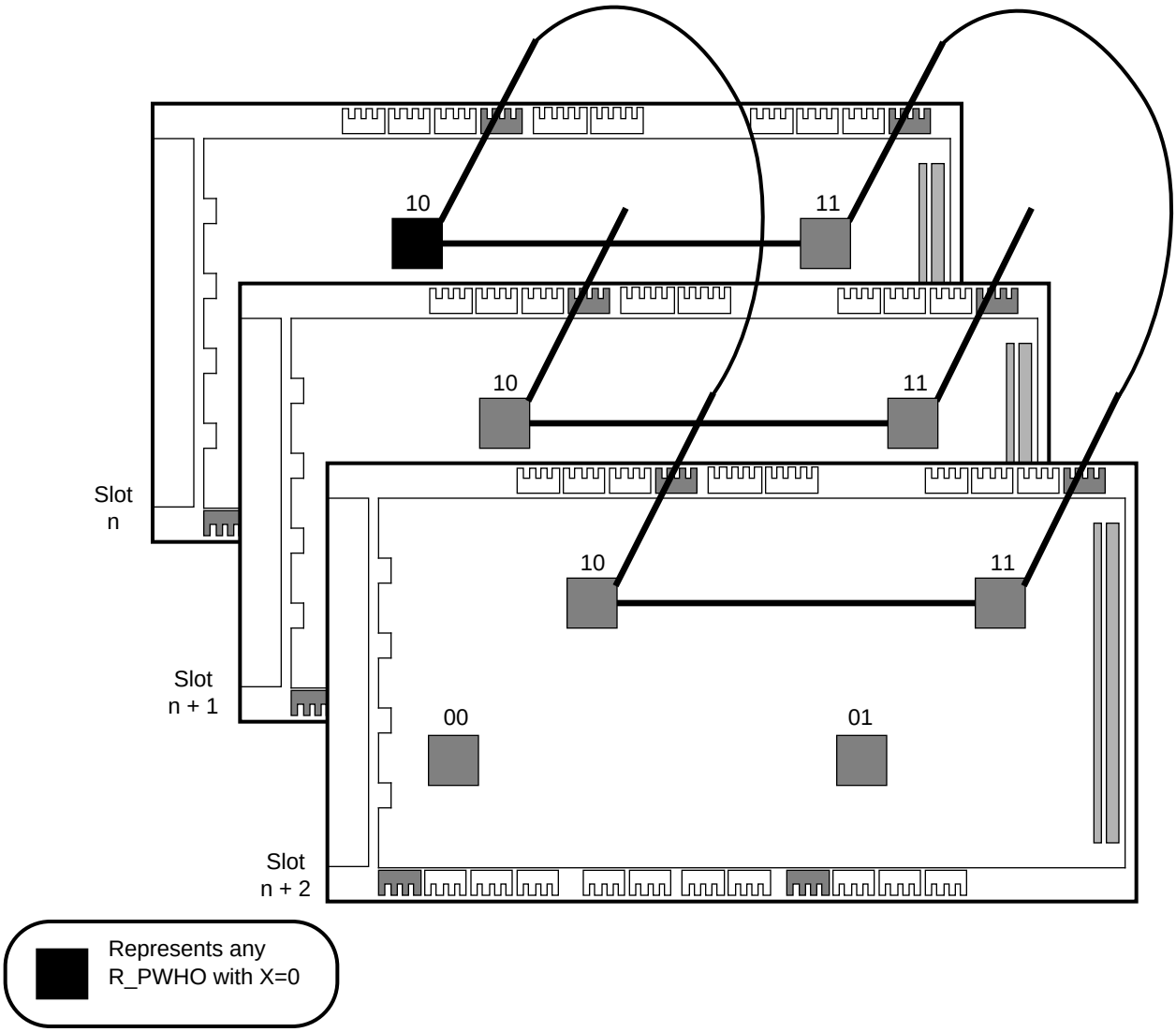
When two PCBs connect in the X dimension of an LC system, the A-side PCB of one module connects to the A-side PCB of the module one slot above. Likewise, the B-side PCB of one module connects to the B-side PCB of the module one slot above (refer to Figure 21).

Figure 21. LC X=4 Communication Link Connections (B Side)



When two PCBs connect together in the X dimension in an AC system, one PCB connects to the PCB two slots away (refer to Figure 22). For clarity, Figure 22 does not show the connections for the network routers in positions 00 and 01 on each PCB.

Figure 22. AC X=4 Communication Link Connections

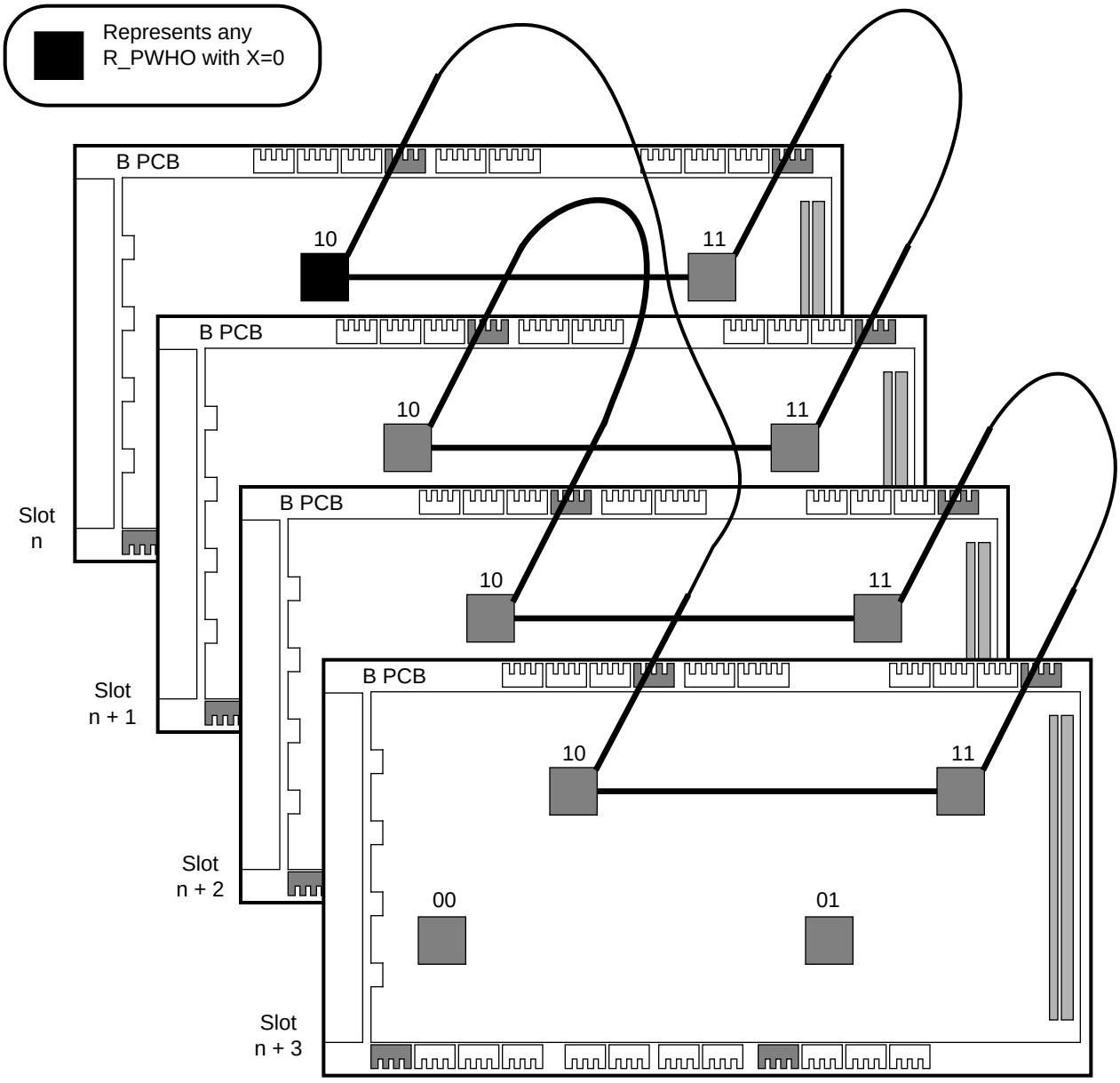


NOTE: When the PCB located in slot n + 1 in Figure 22 is part of a complete plane, the external X-dimension communication links of this PCB connect to another PCB. When the PCB located in slot n + 1 in Figure 22 is part of a partial plane, the external X-dimension communication links of this PCB connect to each other.

X=8 Communication Link Connections

Four PCBs connected create an X=8 node shape. All CRAY T3E LC systems that have 128 or more PEs use this shape. For these connections, the A-side PCBs of four contiguously numbered slots connect in the X dimension. Likewise, the B-side PCBs of four contiguously numbered slots connect in the X dimension (refer to Figure 23). For clarity, Figure 23 does not show the connections for the network routers in positions 00 and 01 on each PCB.

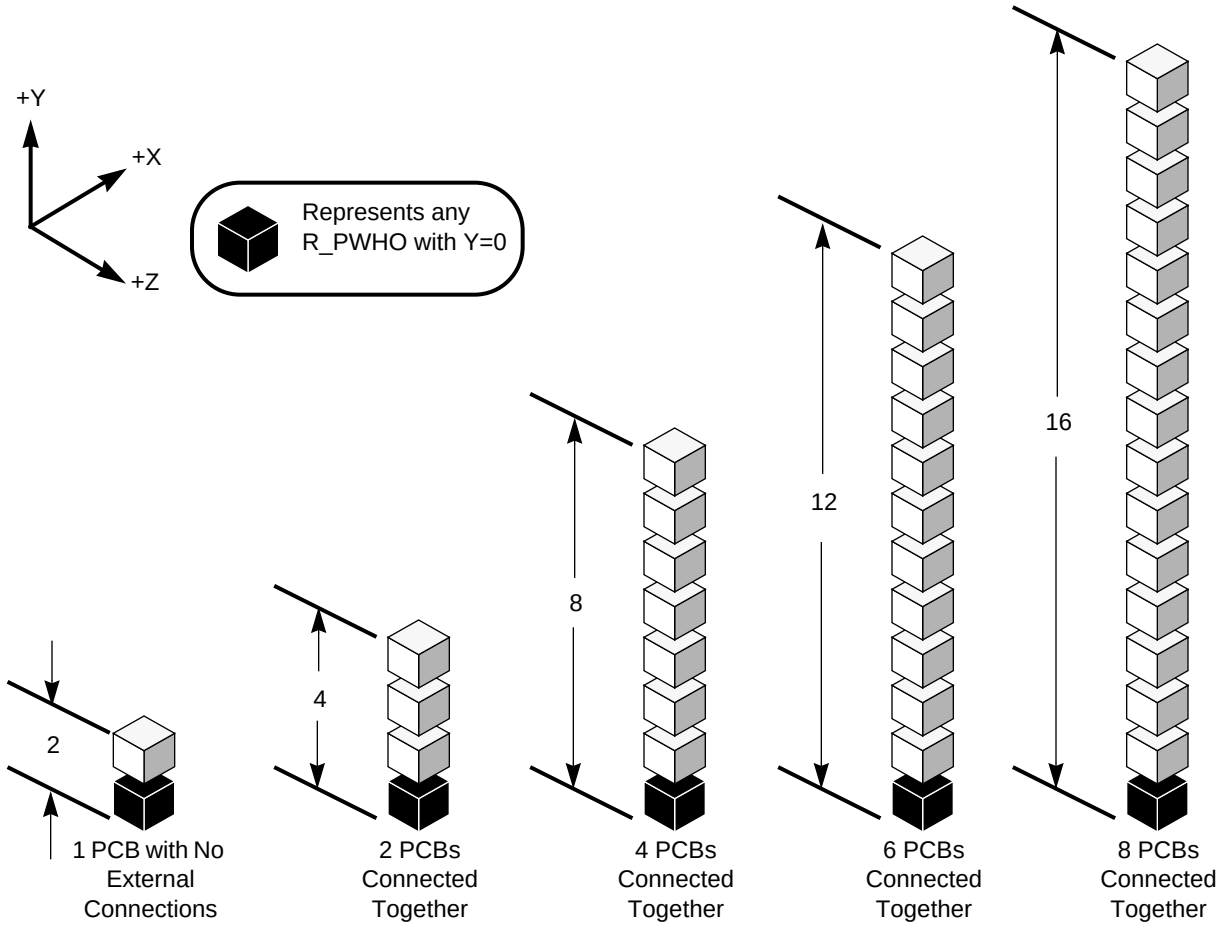
Figure 23. X=8 Communication Link Connections



Y-dimension Node Shapes

In the CRAY T3E system, the Y-dimension may be one of nine shapes: Y=2, Y=4, Y=8, Y=12, Y=16, Y=20, Y=24, Y=28, and Y=32. Figure 24 shows some of these shapes and lists the number of PCBs that connect to form each shape.

Figure 24. Y-dimension Node Shapes

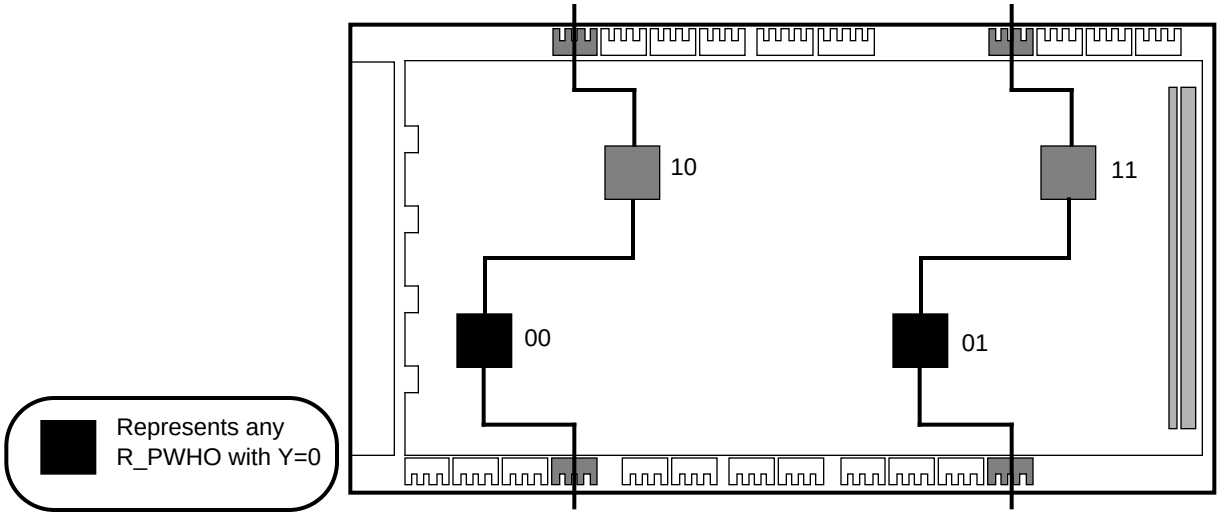


The following subsections show the communication link connections for each Y-dimension node shape.

Y=2 Communication Link Connections

A PCB that does not connect to another PCB creates a Y=2 node shape. The smallest CRAY T3E AC systems contain one, two, or three PCBs that do not connect to other PCBs. For these systems, the external Y-dimension communication links on each PCB are unconnected (refer to Figure 25).

Figure 25. Y=2 Communication Link Connections

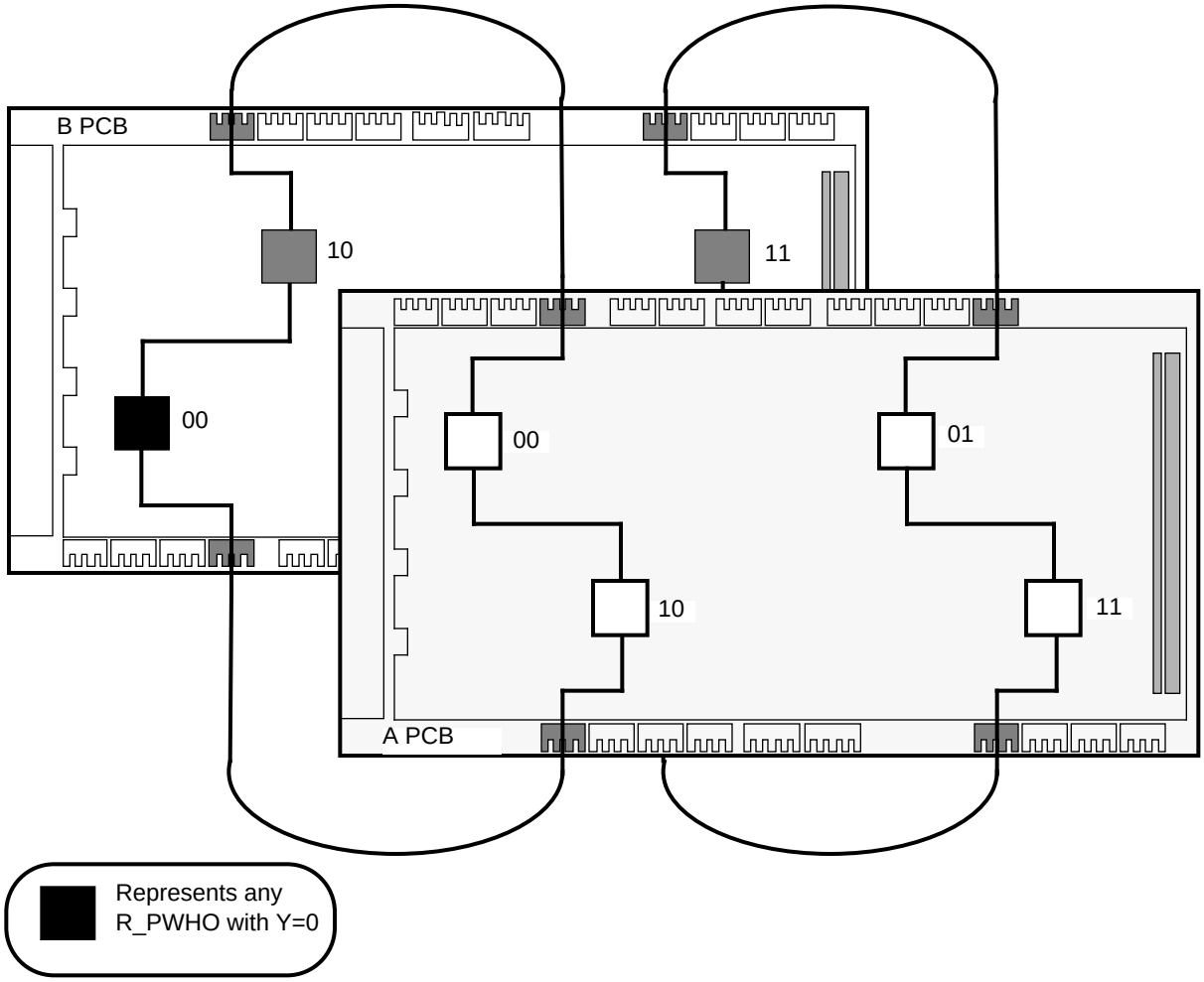


Y=4 Communication LInk Connections

Two PCBs connected create a Y=4 node shape. Several CRAY T3E system configurations use this shape. The communication link connections for this shape are different in the LC systems than they are in the AC systems.

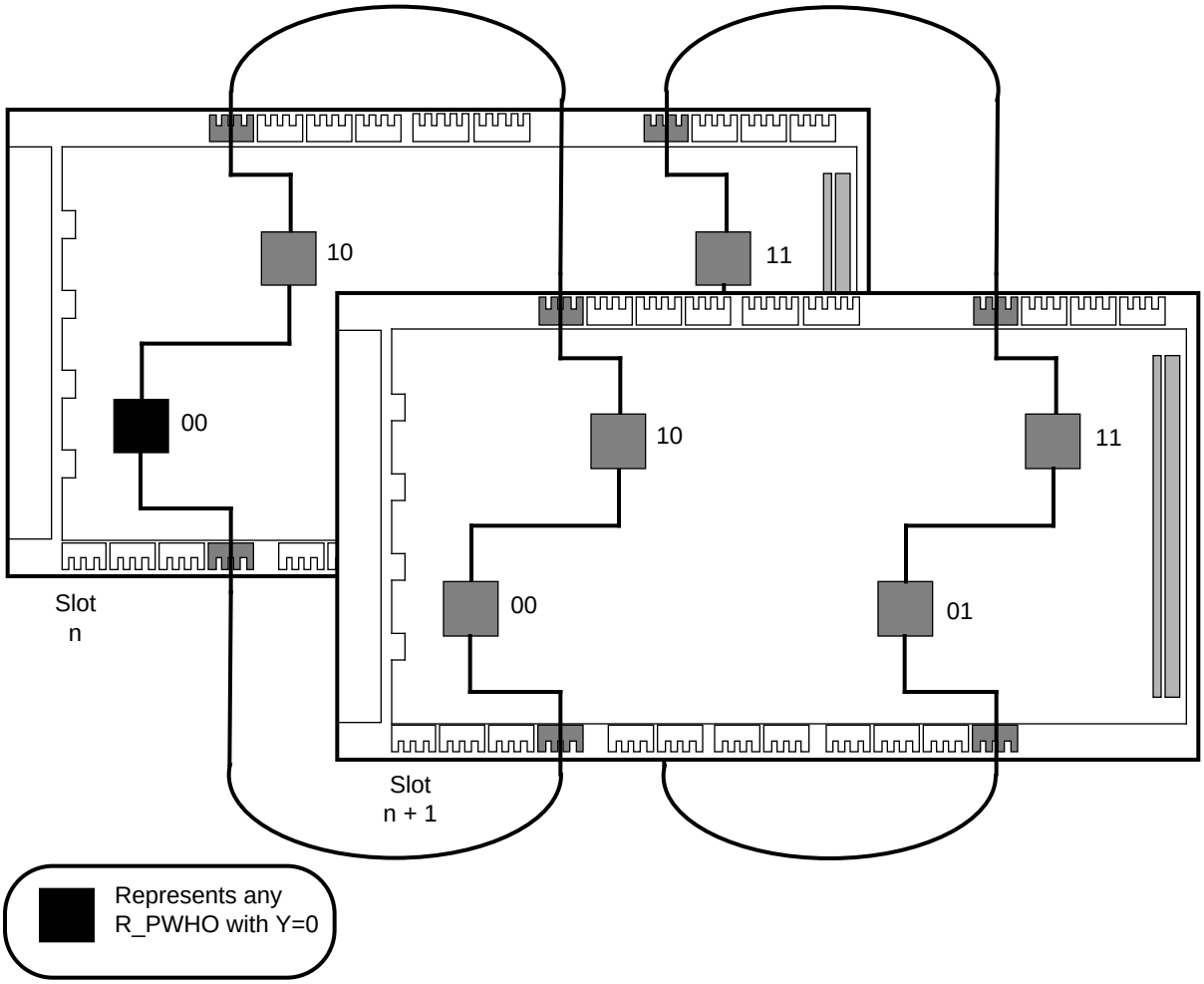
When two PCBs connect in the Y dimension of an LC system, the A-side PCB of a module connects to the B-side PCB of the same module (refer to Figure 26).

Figure 26. LC Y=4 Communication Link Connections



When two PCBs connect in the Y dimension of an AC system, one PCB connects to the PCB in the next slot (refer to Figure 27).

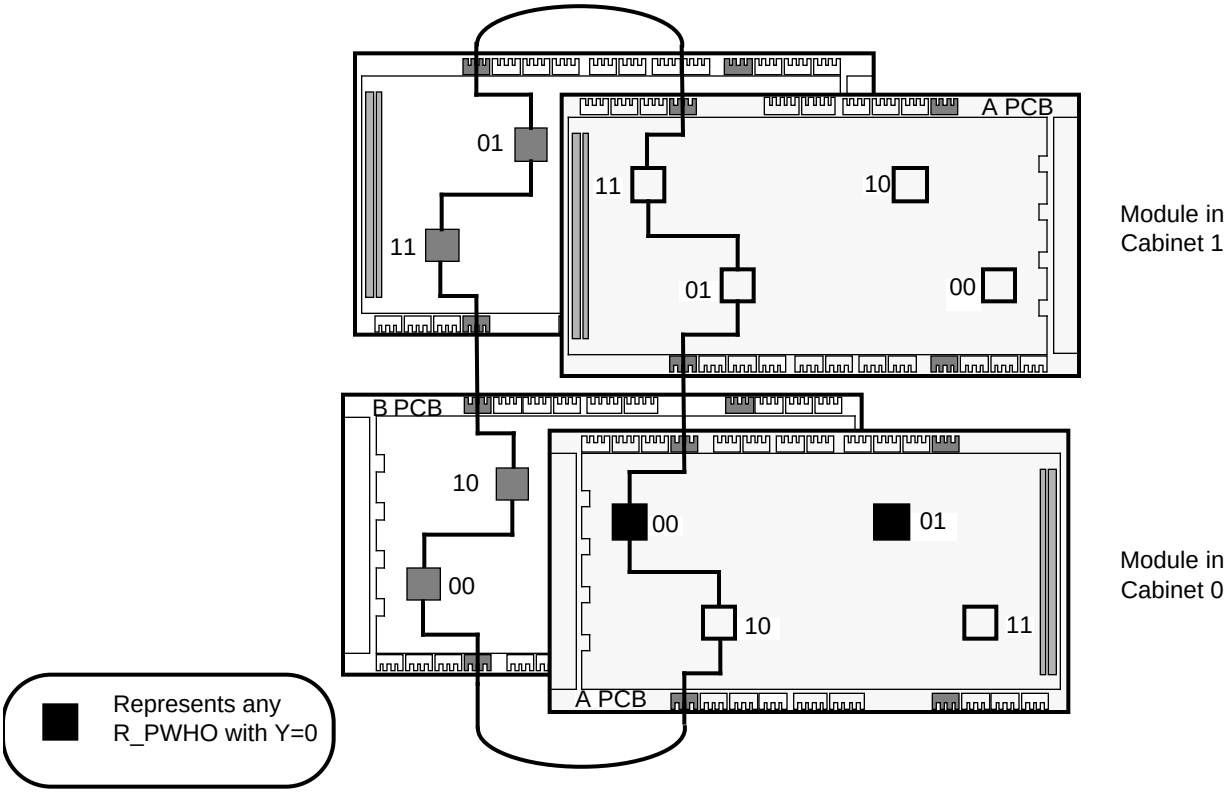
Figure 27. AC Two-PCB External Y-dimension Connections



Y=8 Communication Link Connections

Four PCBs connected create a Y=8 node shape. All CRAY T3E LC systems that have two system cabinets that house modules use this shape. For these connections, the A-side PCB of a module in cabinet 0 connects to the A-side PCB of a module in cabinet 1. Likewise, the B-side PCB of a module in cabinet 0 connects to the B-side PCB of a module in cabinet 1 (refer to Figure 28). For clarity, Figure 28 does not show the connections for half of the network routers.

Figure 28. Y=8 Communication Link Connections



NOTE: In multiple-cabinet liquid-cooled systems, when the X-dimension of the R_PWHO register is 00, 01, 04, or 05, the Y-dimension communication links that connect between cabinets do not connect a module in one cabinet to a module in the same slot of the next cabinet.

As an example, Table 8 shows the cabinet and slot locations for the nodes in the Z=01 plane of a 512-PE CRAY T3E LC System. Notice that when the X-dimension of the R_PWHO register is 00 or 01, the module in slot 17 of cabinet 0 connects to the module in slot 20 of cabinet 1. Also, when the X-dimension of the R_PWHO register is 04 or 05, the module in slot 20 of cabinet 0 connects to the module in slot 17 of cabinet 1.

Table 8. Sample Node Locations for a 512-PE CRAY T3E LC System

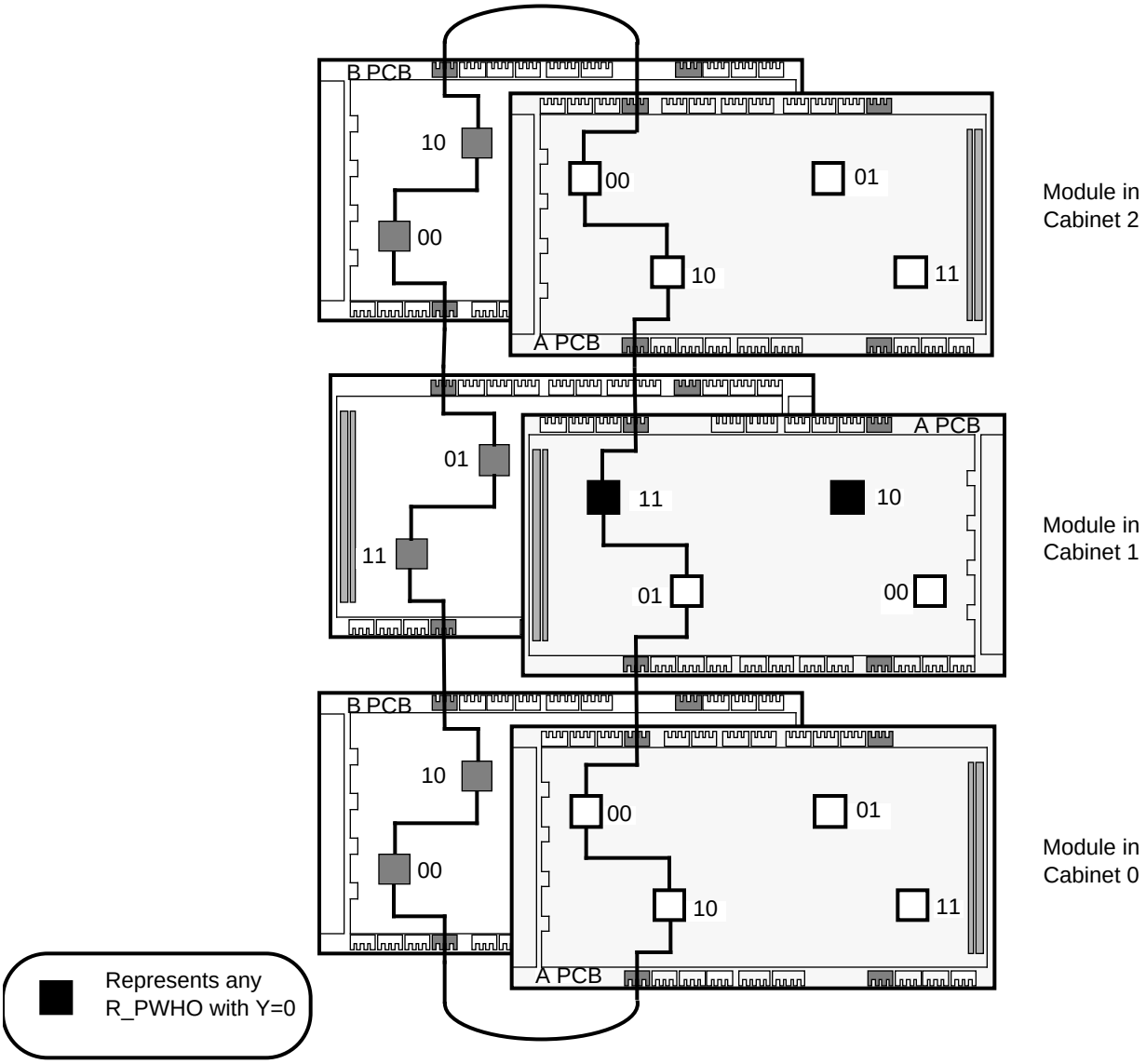
R_PWHO	XX=07	XX=06	XX=05	XX=04	XX=03	XX=02	XX=01	XX=00
0107XX	Cabinet 1 Slot 21	Cabinet 1 Slot 21	Cabinet 1 Slot 17	Cabinet 1 Slot 17	Cabinet 1 Slot 19	Cabinet 1 Slot 19	Cabinet 1 Slot 20	Cabinet 1 Slot 20
0106XX	Cabinet 1 Slot 21	Cabinet 1 Slot 21	Cabinet 1 Slot 17	Cabinet 1 Slot 17	Cabinet 1 Slot 19	Cabinet 1 Slot 19	Cabinet 1 Slot 20	Cabinet 1 Slot 20
0105XX	Cabinet 1 Slot 21	Cabinet 1 Slot 21	Cabinet 1 Slot 17	Cabinet 1 Slot 17	Cabinet 1 Slot 19	Cabinet 1 Slot 19	Cabinet 1 Slot 20	Cabinet 1 Slot 20
0104XX	Cabinet 1 Slot 21	Cabinet 1 Slot 21	Cabinet 1 Slot 17	Cabinet 1 Slot 17	Cabinet 1 Slot 19	Cabinet 1 Slot 19	Cabinet 1 Slot 20	Cabinet 1 Slot 20
0103XX	Cabinet 0 Slot 21	Cabinet 0 Slot 21	Cabinet 0 Slot 20	Cabinet 0 Slot 20	Cabinet 0 Slot 19	Cabinet 0 Slot 19	Cabinet 0 Slot 17	Cabinet 0 Slot 17
0102XX	Cabinet 0 Slot 21	Cabinet 0 Slot 21	Cabinet 0 Slot 20	Cabinet 0 Slot 20	Cabinet 0 Slot 19	Cabinet 0 Slot 19	Cabinet 0 Slot 17	Cabinet 0 Slot 17
0101XX	Cabinet 0 Slot 21	Cabinet 0 Slot 21	Cabinet 0 Slot 20	Cabinet 0 Slot 20	Cabinet 0 Slot 19	Cabinet 0 Slot 19	Cabinet 0 Slot 17	Cabinet 0 Slot 17
0000XX	Cabinet 0 Slot 21	Cabinet 0 Slot 21	Cabinet 0 Slot 20	Cabinet 0 Slot 20	Cabinet 0 Slot 19	Cabinet 0 Slot 19	Cabinet 0 Slot 17	Cabinet 0 Slot 17

When the X-dimension of the R_PWHO register is 02, 03, 06, or 07, the Y-dimension communication links that connect between cabinets do connect a module in one cabinet to a module in the same slot of the next cabinet.

Y=12 Communication Link Connections

Six PCBs connected together create a Y=12 node shape. This shape is used in all CRAY T3E LC systems that have three system cabinets that house modules. For these connections, the A-side PCB of a module in a cabinet connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module in a cabinet connects to the B-side PCB of a module in the next cabinet (refer to Figure 29). For clarity, Figure 29 does not show the connections for half of the network routers.

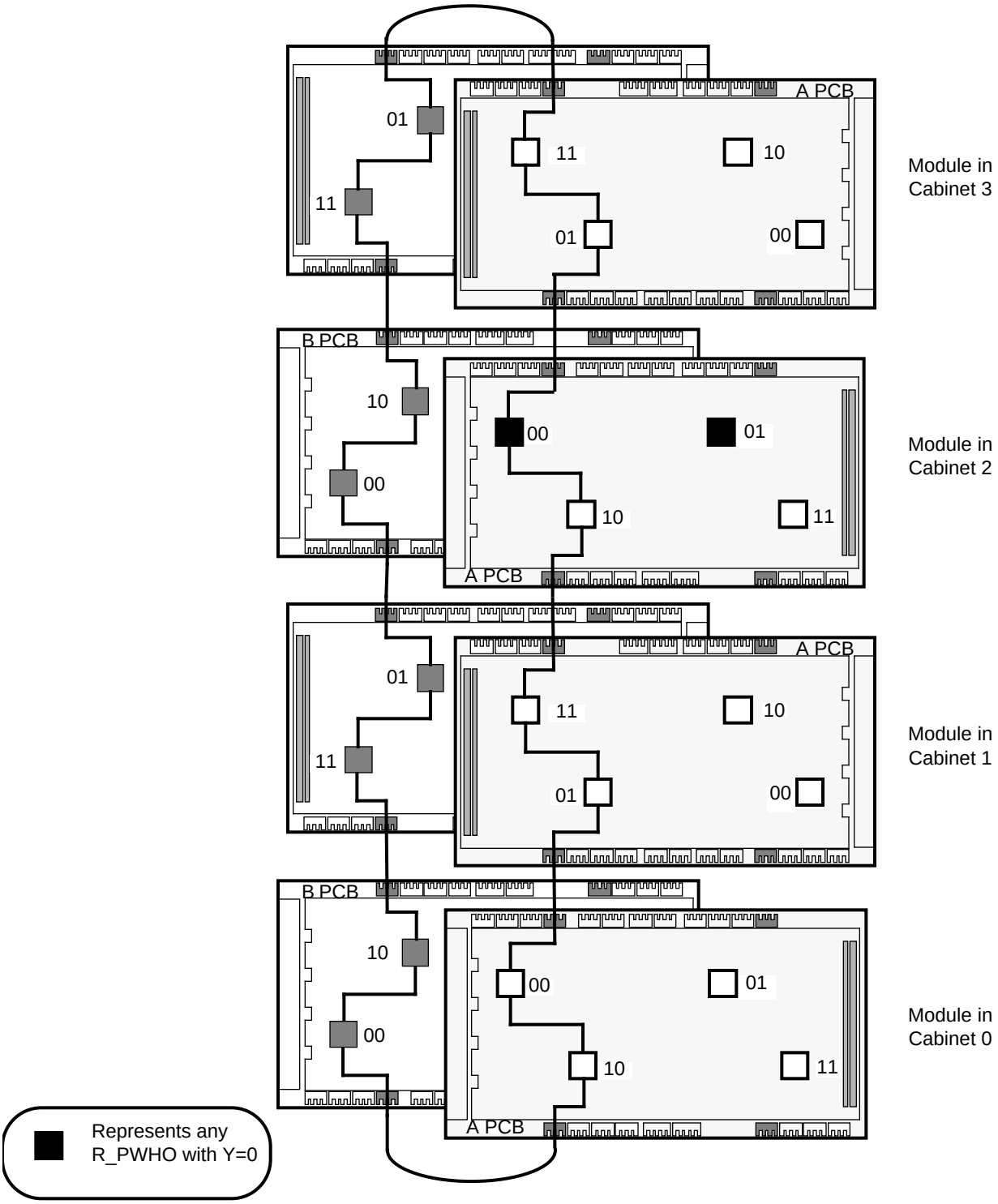
Figure 29. Y=12 Communication Link Connections



Y=16 Communication Link Connections

Eight PCBs connected create a Y=16 node shape. All CRAY T3E LC systems that have four system cabinets that house modules use this shape. For these connections, the A-side PCB of a module connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module connects to the B-side PCB of a module in the next cabinet (refer to Figure 30). For clarity, Figure 30 does not show the connections for half of the network routers.

Figure 30. Y=16 Communication Link Connections



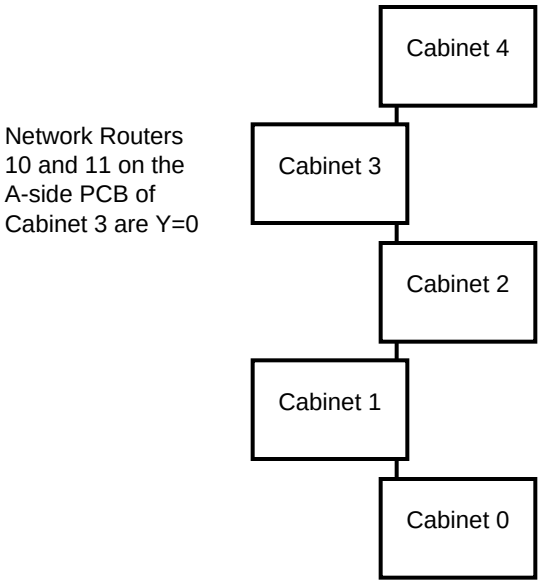
Y=20 Communication Link Connections

Ten PCBs connected create a Y=20 node shape. All CRAY T3E LC systems that have five system cabinets that house modules use this shape. For these connections, the A-side PCB of a module in a cabinet connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module in a cabinet connects to the B-side PCB of a module in the next cabinet.

The Y-dimension communication link connections for a Y=20 node shape are similar to the connections for a Y=16 node shape (refer again to Figure 30); however, the connections are made between 10 PCBs.

Figure 31 shows the cabinet positions and indicates which cabinet contains the Y=0 network routers for a Y=20 node shape.

Figure 31. Y=20 Cabinet Positions



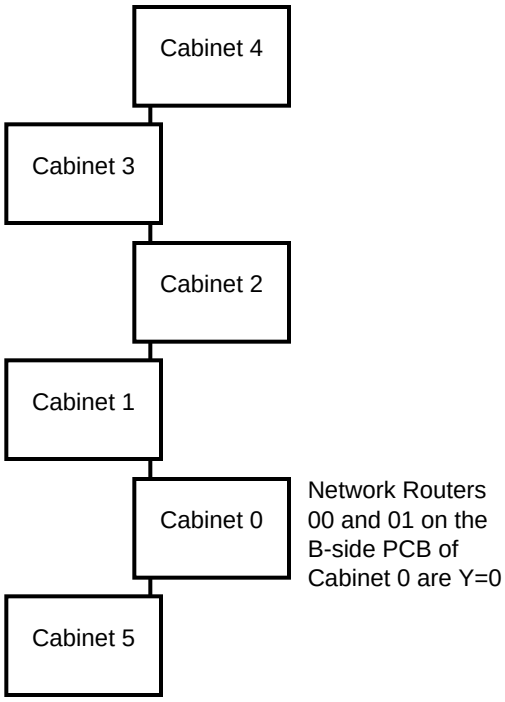
Y=24 Communication Link Connections

Twelve PCBs connected create a Y=24 node shape. All CRAY T3E LC systems that have six system cabinets that house modules use this shape. For these connections, the A-side PCB of a module in a cabinet connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module in a cabinet connects to the B-side PCB of a module in the next cabinet.

The Y-dimension communication link connections for a Y=24 node shape are similar to the connections for a Y=16 node shape (refer again to Figure 30); however, the connections are made between 12 PCBs.

Figure 32 shows the cabinet positions and indicates which cabinet contains the Y=0 network routers for a Y=24 node shape.

Figure 32. Y=24 Cabinet Positions



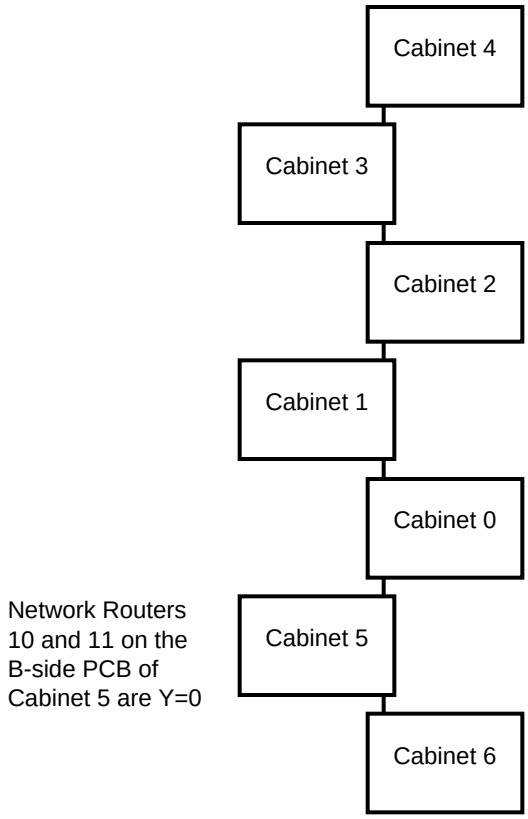
Y=28 Communication Link Connections

Fourteen PCBs connected create a Y=28 node shape. All CRAY T3E LC systems that have seven system cabinets that house modules use this shape. For these connections, the A-side PCB of a module in a cabinet connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module in a cabinet connects to the B-side PCB of a module in the next cabinet.

The Y-dimension communication link connections for a Y=28 node shape are similar to the connections for a Y=16 node shape (refer again to Figure 30); however, the connections are made between 14 PCBs.

Figure 33 shows the cabinet positions and indicates which cabinet contains the Y=0 network routers for a Y=28 node shape.

Figure 33. Y=28 Cabinet Positions



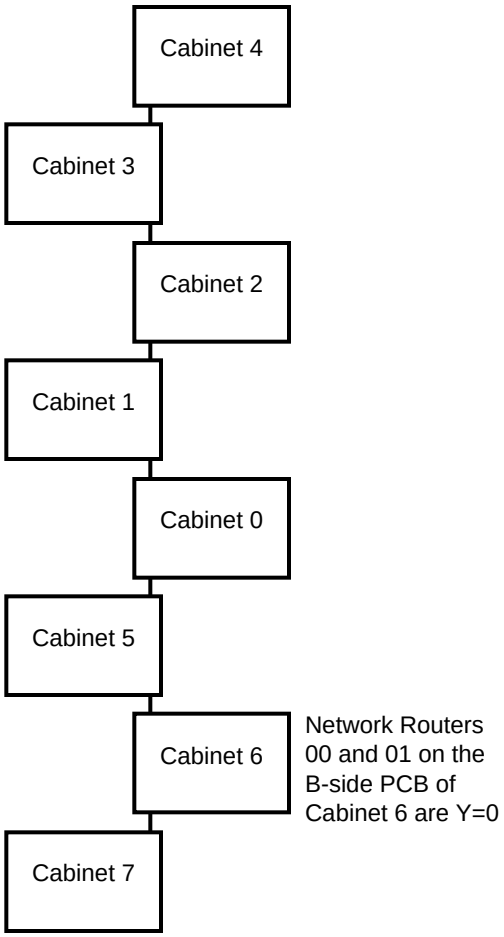
Y=32 Communication Link Connections

Sixteen PCBs connected create a Y=32 node shape. All CRAY T3E LC systems that have eight system cabinets that house modules use this shape. For these connections, the A-side PCB of a module in a cabinet connects to the A-side PCB of a module in the next cabinet. Likewise, the B-side PCB of a module in a cabinet connects to the B-side PCB of a module in the next cabinet.

The Y-dimension communication link connections for a Y=32 node shape are similar to the connections for a Y=16 node shape (refer again to Figure 30); however, the connections are made between 16 PCBs.

Figure 34 shows the cabinet positions and indicates which cabinet contains the Y=0 network routers for a Y=32 node shape.

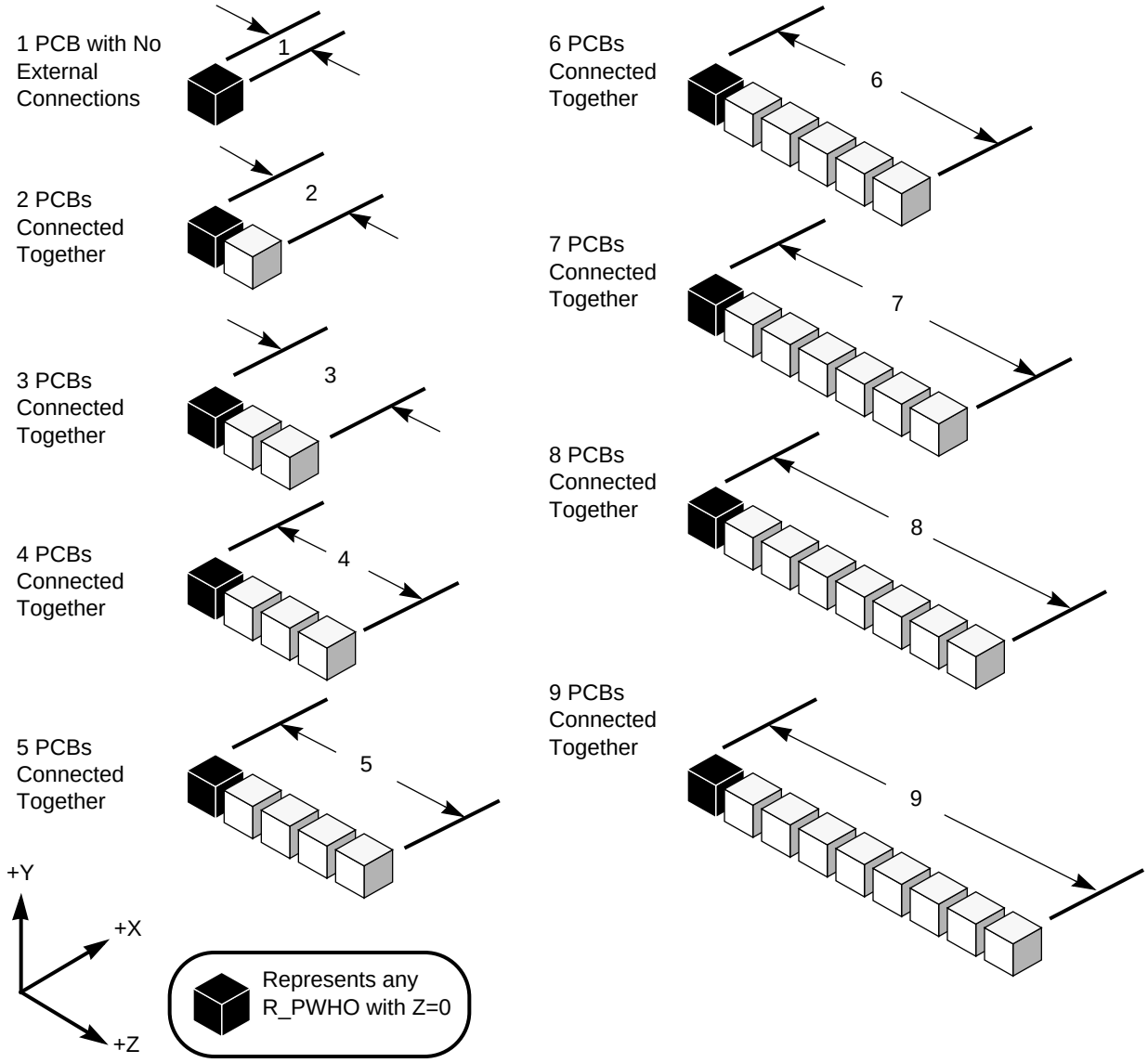
Figure 34. Y=32 Cabinet Positions



Z-dimension Node Shapes

In the CRAY T3E system, the Z-dimension may be one of 9 shapes: Z=1, Z=2, Z=3, Z=4, Z=5, Z=6, Z=7, Z=8, and Z=9. Figure 35 shows these shapes and lists the number of PCBs that connect together to form each shape.

Figure 35. Z-dimension Node Shapes

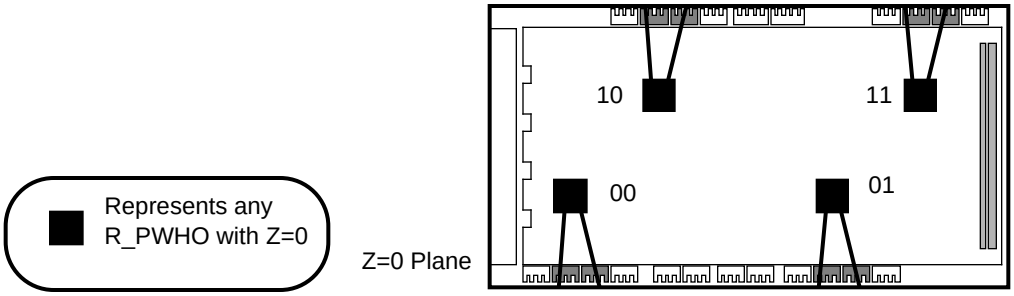


The following subsections show the communication link connections for each Z-dimension node shape. The *CRAY T3E System Configurations* documents show the cabinet slot locations for the Z-dimension.

Z=1 Communication Link Connections

A PCB that does not connect to another PCB creates a Z=1 node shape. The smallest CRAY T3E AC system contains one PCB that does not connect to another PCB. For this system, the external communication links on the PCB are unconnected (refer to Figure 36).

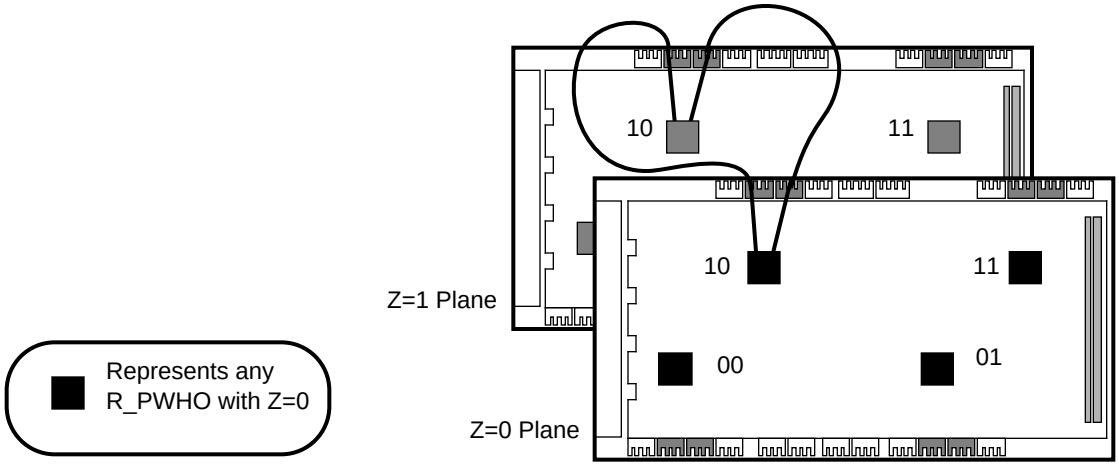
Figure 36. Z=1 Communication Link Connections



Z=2 Communication Link Connections

Two PCBs connected together create a Z=2 node shape (refer to Figure 37). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

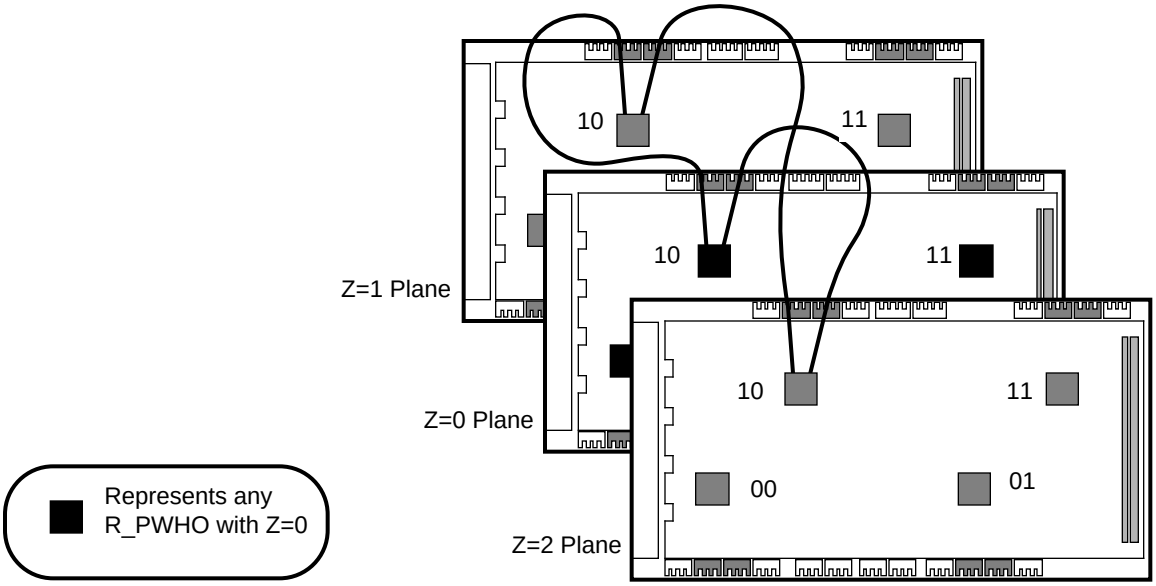
Figure 37. Z=2 Communication Link Connections



Z=3 Communication Link Connections

Three PCBs connected create a Z=3 node shape (refer to Figure 38). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

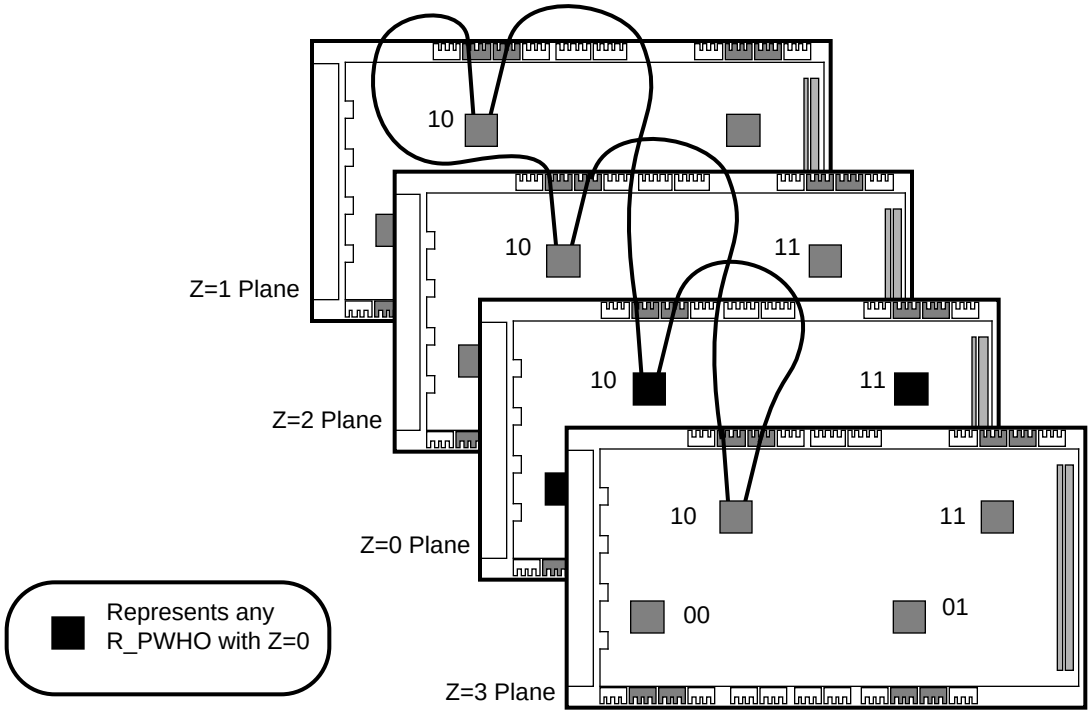
Figure 38. Z=3 Communication Link Connections



Z=4 Communication Link Connections

Four PCBs connected create a Z=4 node shape (refer to Figure 39). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

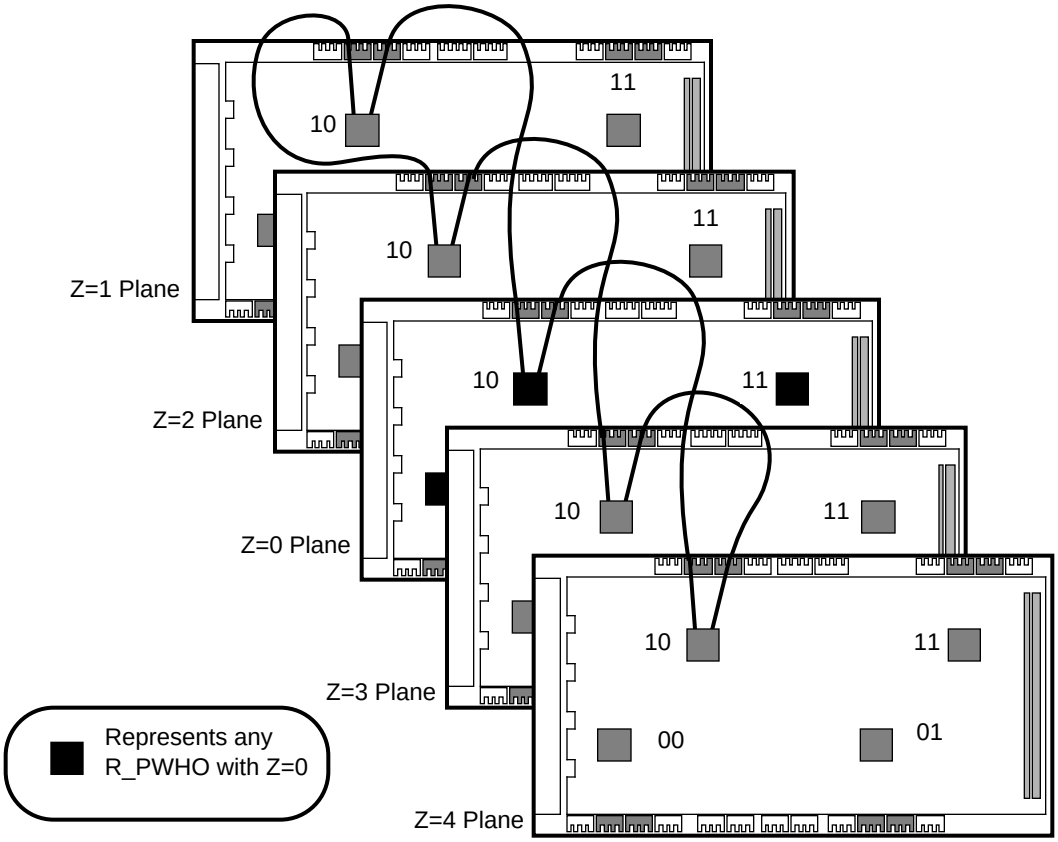
Figure 39. Z=4 Communication Link Connections



Z=5 Communication Link Connections

Five PCBs connected create a Z=5 node shape (refer to Figure 40). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

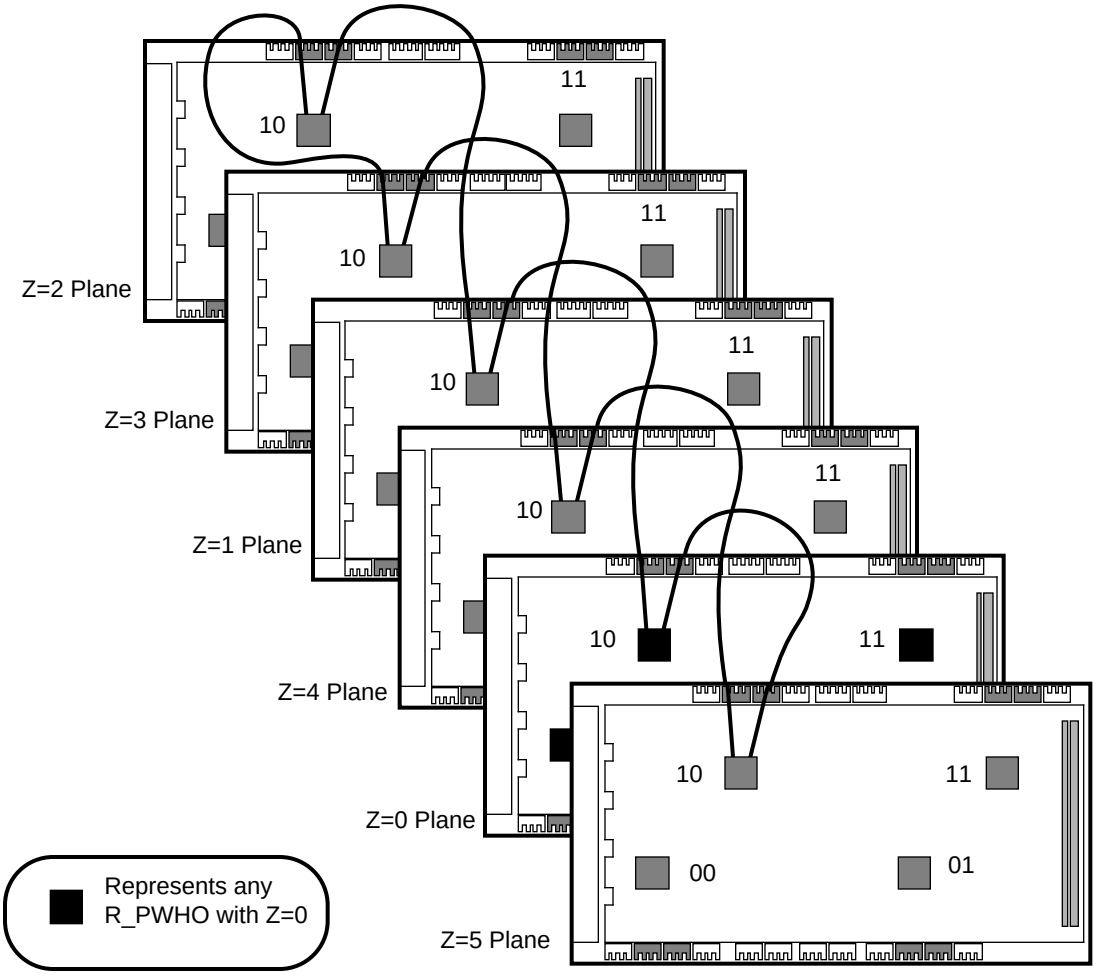
Figure 40. Z=5 Communication Link Connections



Z=6 Communication Link Connections

Six PCBs connected create a Z=6 node shape (refer to Figure 41). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

Figure 41. Z=6 Communication Link Connections



Z=7 Communication Link Connections

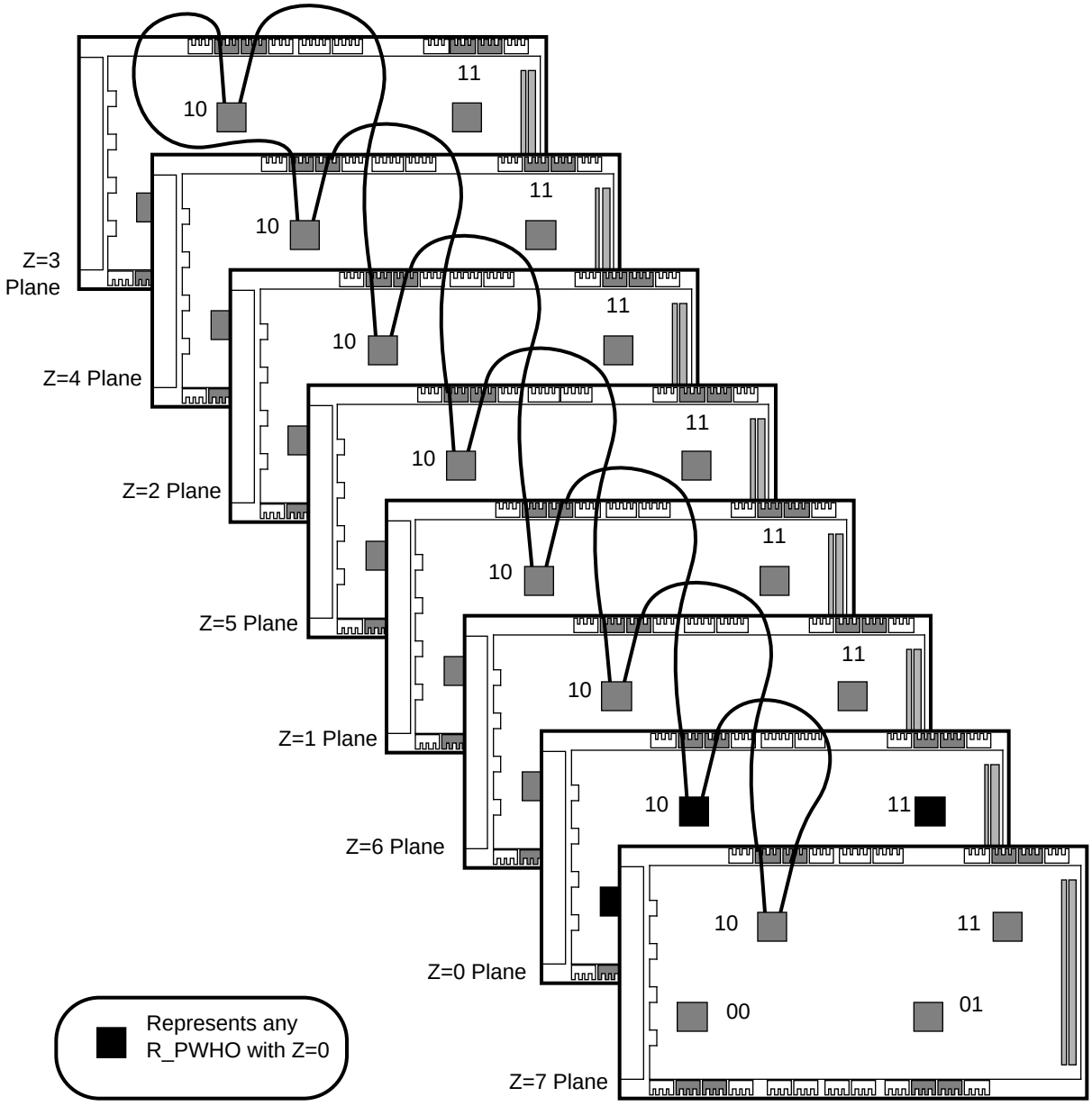
Seven PCBs connected create a Z=7 node shape (refer to Figure 42). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

Figure 42. Z=7 Communication Link Connections

Z=8 Communication Link Connections

Eight PCBs connected create a Z=8 node shape (refer to Figure 43). In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

Figure 43. Z=8 Communication Link Connections

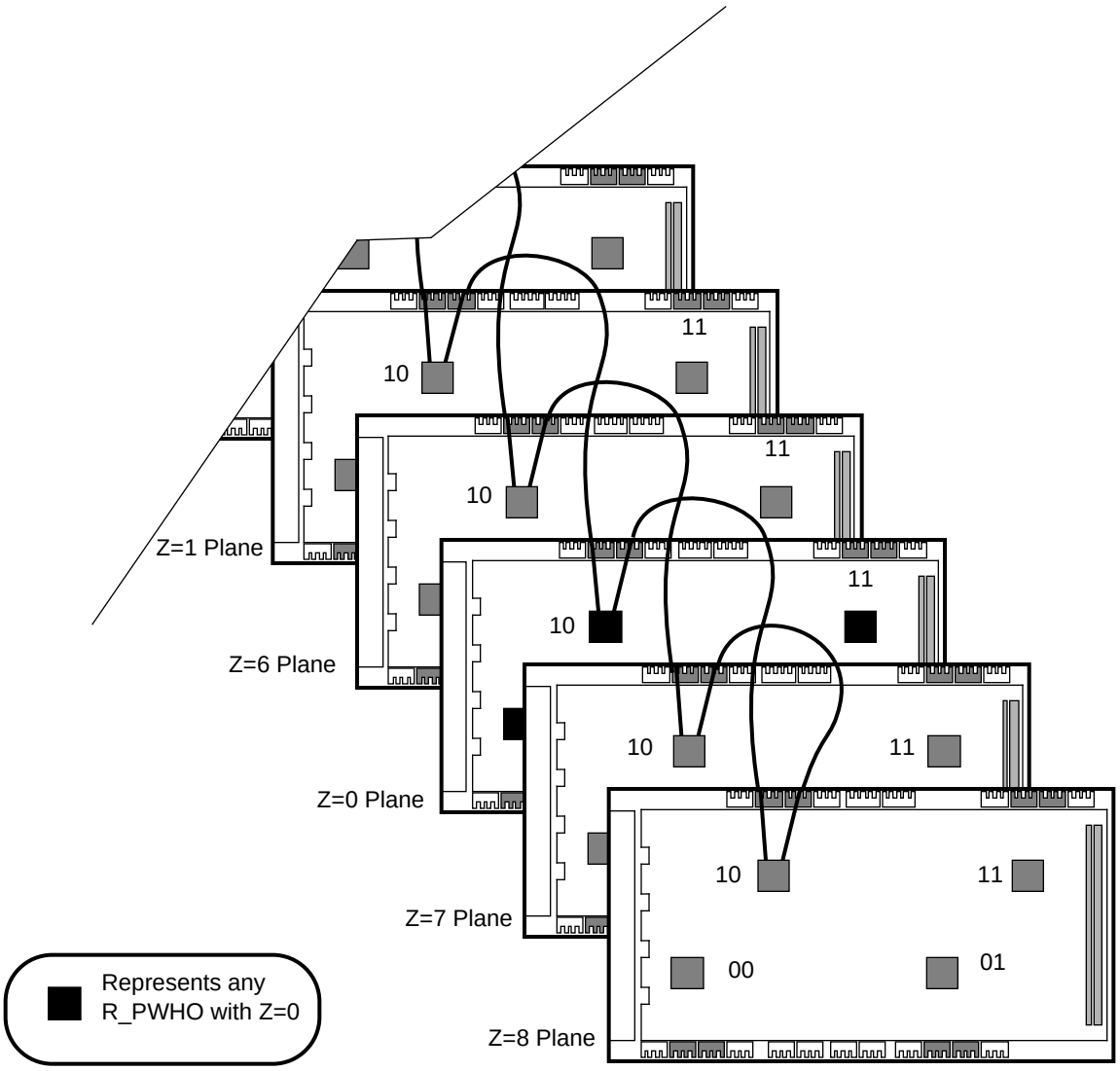


Z=9 Communication Link Connections

Nine PCBs connected create a Z=9 node shape. In an LC system, the A-side PCB of one module connects to the A-side PCB of another module. Likewise in an LC system, the B-side PCB of one module connects to the B-side PCB of another module. In an AC system, the PCB of one module connects to the PCB of another module.

Figure 44 shows nine PCBs connected together in the Z dimension. For clarity, some of the PCBs are hidden in Figure 44. These PCBs are connected and numbered exactly the same as shown in Figure 43.

Figure 44. Z=9 Communication Link Connections



Information Routing

Information travels through the interconnect network using one of three types of routing methods: special routing, direction order routing, or adaptive routing. All three types of routing transfer information over the communication links in the interconnect network; however, each type of routing uses a different set of channel buffers in the network routers.

Special Routing

Software uses special routing for network maintenance, which includes troubleshooting an unhealthy network and initializing certain network registers. To perform special routing, software uses the special get (SGET) and special put (SPUT) E-register commands.

Special routing transfers packets of information through a nonblocking network. A nonblocking network is free of potential deadlock conditions; however, if two packets of information access the same network router port at the same time, one packet will overwrite the other.

Special routing uses delta values to route packets of information through the interconnect network. Delta values enable software to send information to physical nodes before assigning them physical node numbers.

Six delta values define the path of a special route and each delta value is positive or negative. These values are:

1. Primary delta X Value
2. Primary delta Y Value
3. Primary delta Z Value
4. Secondary delta X Value
5. Secondary delta Y Value
6. Secondary delta Z Value

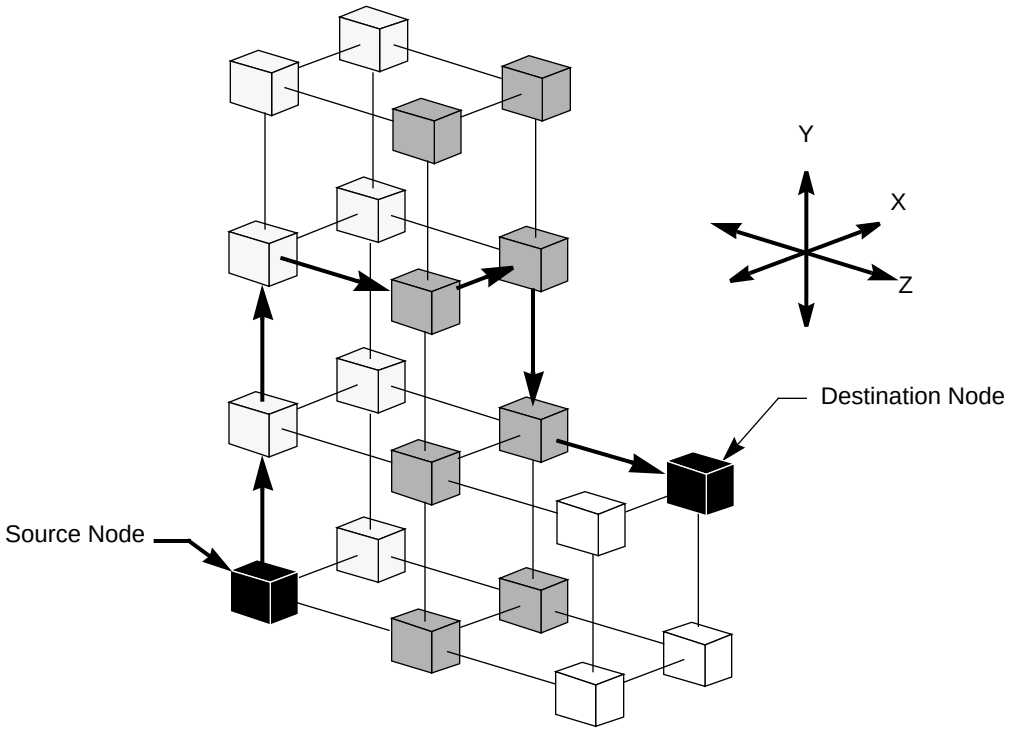
In the interconnect network, each transfer of information over a communication link is called a hop. When following a special routing path, information makes the same number of hops as a delta value. For example, if the primary delta X value is set to +3, the information completes three hops in the positive X direction. Also when following a special routing path, information travels through the directions in the order shown in the list above.

Special Routing Path

Figure 45 shows one of several possible special routing paths through the interconnect network from a source node to a destination node in a 20-PE CRAY T3E AC system. For this example, the delta values are set as shown in the following list. For clarity, Figure 45 does not show all of the communication links in the system.

- Primary delta X Value = 0
- Primary delta Y Value = +2
- Primary delta Z Value = +1
- Secondary delta X Value = +1
- Secondary delta Y Value = –1
- Secondary delta Z Value = +1

Figure 45. Sample Special Route



Special Routing Examples

As an example of special routing, the following list describes how a source node writes a value into the network router SGET return (R_SGET_RTRN) register of a destination node. The destination node uses the values stored in the R_SGET_RTRN register as the delta values for an SGET response packet.

1. The PE in the source node stores values for the R_SGET_RTRN register of the destination node in a source E register.

2. The PE in the source node issues an SPUT command. While issuing the command, the PE generates the delta values, the address for the R_SGET_RTRN register, and the source E-register number.

3. The network router in the source node creates an SPUT packet and sends the packet to the destination node over the nonblocking network.

4. The network router in the destination node receives the SPUT packet and stores the delta values in the R_SGET_RTRN register.

As another example of special routing, the following list describes how a source node reads a value from the network router software (R_SW) register of a destination node. Software writes values to the R_SW register while determining the node shape of the system. Software can write any value to the R_SW register and the value does not affect the hardware.

1. The PE in the source node issues an SGET command. While issuing the command, the PE generates the delta values, the address for the R_SW register, and the destination E-register number (in the source node).

2. The network router in the source node creates an SGET request packet and sends the packet to the destination node over the nonblocking network.

3. The network router in the destination node retrieves the value from the R_SW register and uses the value stored in the R_SGET_RTRN register to create an SGET response packet.

4. The network router in the destination node sends the SGET response packet to the source node over the nonblocking network.

5. The source node stores the value from the R_SW register into the destination E register in the source node.

Direction Order Routing

Software uses direction order routing to transfer data through the interconnect network in a predetermined path. When a packet of information follows direction order routing, the packet travels through the network in the following order:

1. Any initial hop in the +X, +Y, or +Z direction
2. Any travel in the +X direction
3. Any travel in the +Y direction
4. Any travel in the +Z direction
5. Any travel in the –X direction
6. Any travel in the –Y direction
7. Any travel in the –Z direction
8. Any final hop in the –Z direction

Before direction order routing can occur, software must store configuration and addressing values into network router registers. These registers include:

- R_ORIENT register
- R_PWHO register
- R_LWHO register
- R_CHAN[7 : 0] registers
- R_NET_LUT registers
- R_LUT_OVERRIDE register
- R_VCSEL[5 : 0] registers

The “Direction Assignments” subsection in this document provides more information on the R_ORIENT register. The “Node and PE Numbers” subsection in this document provides more information on the R_PWHO and R_LWHO registers. The following subsections provide more information on the other registers in the list.

R_CHAN[7 : 0] Registers

Each network router has eight network router channel (R_CHAN[7 : 0]) registers. Each R_CHAN register is assigned to one channel (refer to Table 9).

Table 9. R_CHAN Registers

Register	Assigned to
R_CHAN0	+X channel (The network router port assigned to +X)
R_CHAN1	+Y channel (The network router port assigned to +Y)
R_CHAN2	+Z channel (The network router port assigned to +Z)
R_CHAN3	–X channel (The network router port assigned to –X)
R_CHAN4	–Y channel (The network router port assigned to –Y)
R_CHAN5	–Z channel (The network router port assigned to –Z)
R_CHAN6	Support circuitry channel (The network port)
R_CHAN7	I/O channel

Software uses the R_CHAN[7 : 0] registers to logically disconnect a communication link from a node. This process prevents corrupted packets of information from propagating through the interconnect network. Table 10 shows the bit format of the R_CHAN[7 : 0] registers.

Table 10. R_CHAN[7 : 0] Bit Format

Bits	Name	Description
0	IGNORE_NORMAL	When set to 1, this bit signals the network router that when it receives any packets on this channel that use the direction order routing buffers, the network router will not pass the packet to the next node in the network.
1	IGNORE_SPCL	When set to 1, this bit signals the network router that when it receives any packets on this channel that use the special routing buffers, the network router will not pass the packet to the next node in the network.
2	NO_ADAPT	When set to 1, this bit indicates that the adaptive routing channel out of this channel is disabled. This bit is not valid for R_CHAN6 and R_CHAN7.
3	NO_FAST	When set to 1, this bit disables the fast path through the network router port. This bit is not valid for R_CHAN6 and R_CHAN7.
<63 : 4>	Not applicable	These bits are not used.

R_NET_LUT Registers

Each network router has a set of network router look-up table (R_NET_LUT) registers. When software running in a source PE signals the support circuitry to transfer information to a destination logical PE, the support circuitry uses the R_NET_LUT registers to create a routing tag for a packet of information. The following subsections describe the routing tag, describe the R_NET_LUT register format, and provide sample values for the R_NET_LUT registers.

Routing Tag

The routing tag determines the path a packet follows when traveling from the source physical node to the destination physical node. The routing tag consists of the following fields:

- X-dimension address and direction
- Y-dimension address and direction
- Z-dimension address and direction
- Initial +X, +Y, or +Z hop
- Final –Z hop
- Adaptive routing bit

The three dimension address and direction fields each contain part of a physical address and a direction value. For example, if the X-dimension address is 3 and the X-dimension direction is positive, a packet completes hops in the +X direction until it arrives at a node with a physical node number with X=3.

The initial +X, +Y, or +Z hop field indicates whether a packet makes an initial hop in the +X, +Y, or +Z direction. A packet may complete one initial hop before continuing with the path determined by the routing tag. Software may use an initial hop to make routing easier for packets that travel to physical nodes that reside in a partial plane.

The final –Z hop field indicates whether a packet makes a final hop in the –Z direction after completing the path determined by the routing tag. Software may also use the final hop for routing to physical nodes that are in a partial plane.

The adaptive routing bit indicates whether the packet follows direction order routing or adaptive routing when traveling through the network.

R_NET_LUT Format

There are a total of 544 R_NET_LUT registers. For systems with 544 or fewer PEs, software may assign one R_NET_LUT register to each logical PE in the system. For example, in a 16-PE CRAY T3E system, software may use the first 16 contiguous R_NET_LUT registers (refer to mode 0 in Table 11).

For all systems, software may assign one R_NET_LUT register to four logical PEs in the system. For example, in a 16-PE CRAY T3E system, software may use the first 4 contiguous R_NET_LUT registers (refer to mode 1 or 2 in Table 11).

NOTE: For systems with more than 544 PEs, software cannot use mode 0. Software selects the mode by setting bits in the R_LUT_OVERRIDE register.

Table 11. Logical PE to R_NET_LUT Register Assignments (16 PEs)

R_LWHO (hex)	Register (Mode 0)	Register (Mode 1 or 2)
0	R_NET_LUT[0]	R_NET_LUT[0]
1	R_NET_LUT[1]	R_NET_LUT[0]
2	R_NET_LUT[2]	R_NET_LUT[0]
3	R_NET_LUT[3]	R_NET_LUT[0]
4	R_NET_LUT[4]	R_NET_LUT[1]
5	R_NET_LUT[5]	R_NET_LUT[1]
6	R_NET_LUT[6]	R_NET_LUT[1]
7	R_NET_LUT[7]	R_NET_LUT[1]
8	R_NET_LUT[8]	R_NET_LUT[2]
9	R_NET_LUT[9]	R_NET_LUT[2]
A	R_NET_LUT[10]	R_NET_LUT[2]
B	R_NET_LUT[11]	R_NET_LUT[2]
C	R_NET_LUT[12]	R_NET_LUT[3]
D	R_NET_LUT[13]	R_NET_LUT[3]
E	R_NET_LUT[14]	R_NET_LUT[3]
F	R_NET_LUT[15]	R_NET_LUT[3]

When software running in a source PE signals the support circuitry to transfer information to a destination logical PE, the support circuitry references the R_NET_LUT register that corresponds to the destination logical PE number. The network router then reads routing tag information out of that R_NET_LUT register (refer to Table 12) and uses the information to create a packet.

Table 12. R_NET_LUT Register Bit Format

Bits	Name	Description
<2 : 0>	X_ADR	These bits contain the X-dimension physical address of the destination node. For mode 1 and mode 2, the value of bit 0 does not matter. When software sets the mode to mode 1, the support circuitry replaces bit 0 of the R_NET_LUT register with bit 0 of the logical PE number. When software sets the mode to mode 2, the support circuitry replaces bit 0 of the R_NET_LUT register with bit 1 of the logical PE number.
<6 : 3>	Not applicable	These bits are not used.
7	X_SGN	This bit indicates the X-dimension direction. 0 = +X 1 = –X
<12 : 8>	Y_ADR	These bits contain the Y-dimension physical address of the destination node. For mode 1 and mode 2, the value of bit 8 does not matter. When software sets the mode to mode 1, the support circuitry replaces bit 8 of the R_NET_LUT register with bit 1 of the logical PE number. When software sets the mode to mode 2, the support circuitry replaces bit 8 of the R_NET_LUT register with bit 0 of the logical PE number.
<14 : 13>	Not applicable	These bits are not used.
15	Y_SGN	This bit indicates the Y-dimension direction. 0 = +Y 1 = –Y
<19 : 16>	Z_ADR	These bits contain the Z-dimension physical address of the destination node.
<22 : 20>	Not applicable	These bits are not used.
23	Z_SGN	This bit indicates the Z-dimension direction. 0 = +Z 1 = –Z
<31 : 24>	Not applicable	These bits are not used.
<33 : 32>	INITIAL_HOP	These bits indicate the initial hop direction. 00 = +X 10 = +Z 01 = +Y 11 = No initial hop
<35 : 34>	Not applicable	These bits are not used.
36	FINAL_HOP	This bit indicates the final hop direction. When set to 1, the Z_SGN and ADAPT bits must be set to 0. 0 = No final hop 1 = –Z
37	ADAPT	This bit indicates adaptive routing. 0 = Direction order routing 1 = Adaptive routing
<63 : 38>	Not applicable	These bits are not used.

Mode 0 R_NET_LUT Sample Values

Figure 46 shows the physical nodes for a 16-PE CRAY T3E system. For clarity, Figure 46 does not show the communication links that complete the torus in each dimension. For the following example, software assigned each physical node in the system the unique logical PE numbers shown in Figure 46.

As an example of mode 0 R_NET_LUT register values, Figure 46 shows one possible direction order routing path from logical PE 2 [R_PWHO 000100 (hex)] to logical PE 4 [R_PWHO 010000 (hex)]. In physical node R_PWHO 000100 (hex), the R_NET_LUT[4] register values for this routing path are:

- No initial hop (INITIAL_HOP = 3)
- X direction = + (X_SGN = 0)
- X address = 0 (X_ADR = 0)
- Z direction = + (Z_SGN = 0)
- Z address = 1 (Z_ADR = 1)
- Y direction = – (Y_SGN = 1)
- Y address = 0 (Y_ADR = 0)
- No final hop (FINAL_HOP = 0)

Figure 46. Sample Mode 0 Direction Order Routing Path

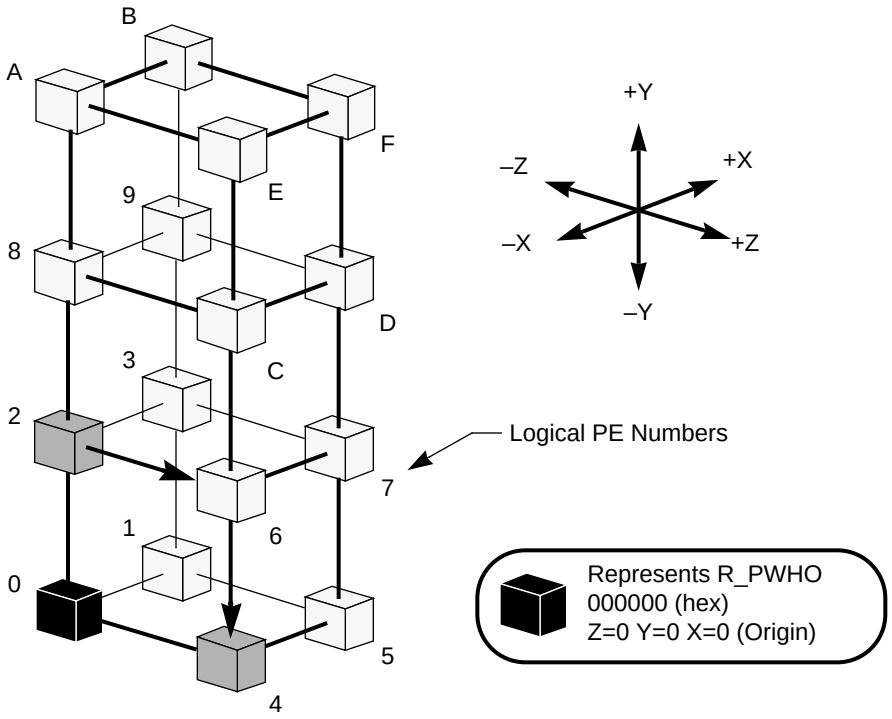


Table 13 shows possible values for all 16 R_NET_LUT registers used in physical node R_PWHO 000100 (hex) shown in Figure 46.

Table 13. R_NET_LUT Register Values for R_PWHO 000100 (hex)

Register	X Dimension		Y Dimension		Z Dimension		Initial Hop	Final Hop
	Address	Direction	Address	Direction	Address	Direction		
R_NET_LUT[0]	0	+	0	–	0	+	None	None
R_NET_LUT[1]	1	+	0	–	0	+	None	None
R_NET_LUT[2]	0	+	1	+	0	+	None	None
R_NET_LUT[3]	1	+	1	+	0	+	None	None
R_NET_LUT[4]	0	+	0	–	1	+	None	None
R_NET_LUT[5]	1	+	0	–	1	+	None	None
R_NET_LUT[6]	0	+	1	+	1	+	None	None
R_NET_LUT[7]	1	+	1	+	1	+	None	None
R_NET_LUT[8]	0	+	2	+	0	+	None	None
R_NET_LUT[9]	1	+	2	+	0	+	None	None
R_NET_LUT[10]	0	+	3	+	0	+	None	None
R_NET_LUT[11]	1	+	3	+	0	+	None	None
R_NET_LUT[12]	0	+	2	+	1	+	None	None
R_NET_LUT[13]	1	+	2	+	1	+	None	None
R_NET_LUT[14]	0	+	3	+	1	+	None	None
R_NET_LUT[15]	1	+	3	+	1	+	None	None

Mode 1 R_NET_LUT Sample Values

Figure 47 shows the physical nodes for a 16-PE CRAY T3E system. For clarity, Figure 47 does not show the communication links that complete the torus in each dimension. For the following example, software assigned each physical node in the system the unique logical PE numbers shown in Figure 47.

As an example of mode 1 R_NET_LUT register values, Figure 47 shows one possible direction order routing path from logical PE 2 [R_PWHO 000100 (hex)] to logical PE 4 [R_PWHO 010000 (hex)]. In physical node R_PWHO 000100 (hex), the R_NET_LUT[1] register values for this routing path are:

- No initial hop (INITIAL_HOP = 3)
- X direction = + (X_SGN = 0)
- X address = 00x (binary) (X_ADR = 00x)
- Z direction = + (Z_SGN = 0)
- Z address = 0001 (binary) (Z_ADR = 0001)
- Y direction = – (Y_SGN = 1)
- Y address = 0000x (binary) (Y_ADR = 0000x)
- No final hop (FINAL_HOP = 0)

Figure 47. Sample Mode 1 Direction Order Routing Path

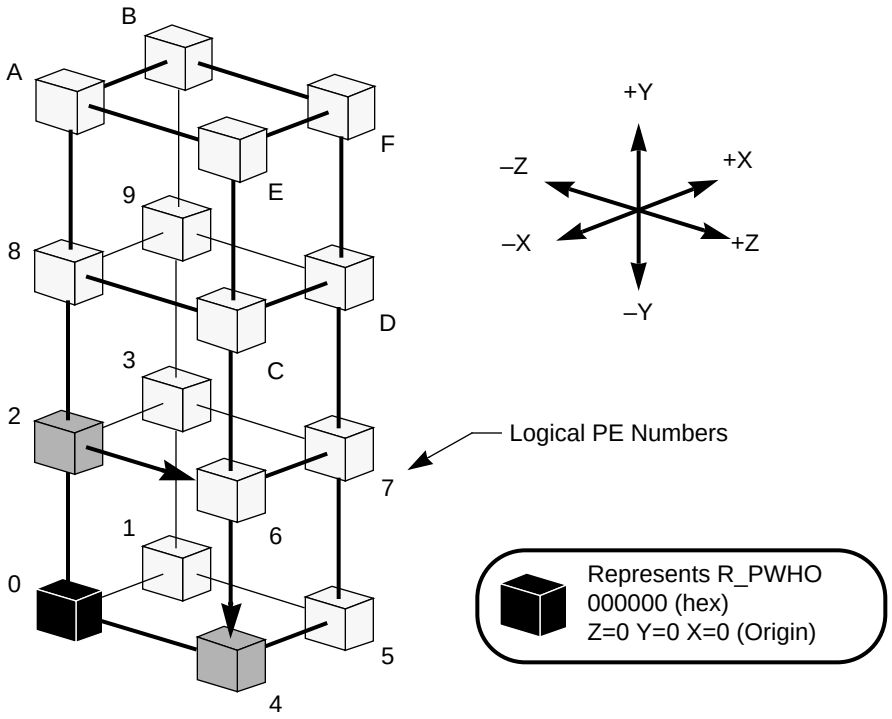


Table 14 shows possible values for all 4 R_NET_LUT registers used in physical node R_PWHO 000100 (hex) shown in Figure 47.

Table 14. R_NET_LUT Register Values for R_PWHO 000100 (hex)

Register	X Dimension		Y Dimension		Z Dimension		Initial Hop	Final Hop
	Address	Direction	Address	Direction	Address	Direction		
R_NET_LUT[0]	00x †	+	0000x	–	0000	+	None	None
R_NET_LUT[1]	00x	+	0000x	–	0001	+	None	None
R_NET_LUT[2]	00x	+	0001x	+	0000	+	None	None
R_NET_LUT[3]	00x	+	0001x	+	0001	+	None	None

† All addresses are binary.

Mode 2 R_NET_LUT Sample Values

Figure 48 shows the physical nodes for a 16-PE CRAY T3E system. For clarity, Figure 48 does not show the communication links that complete the torus in each dimension. For the following example, software assigned each physical node in the system the unique logical PE numbers shown in Figure 48.

As an example of mode 2 R_NET_LUT register values, Figure 48 shows one possible direction order routing path from logical PE 1 [R_PWHO 000100 (hex)] to logical PE 4 [R_PWHO 010000 (hex)]. In physical node R_PWHO 000100 (hex), the R_NET_LUT[1] register values for this routing path are:

- No initial hop (INITIAL_HOP = 3)
- X direction = + (X_SGN = 0)
- X address = 00x (binary) (X_ADR = 00x)
- Z direction = + (Z_SGN = 0)
- Z address = 0001 (binary) (Z_ADR = 0001)
- Y direction = – (Y_SGN = 1)
- Y address = 0000x (binary) (Y_ADR = 0000x)
- No final hop (FINAL_HOP = 0)

Figure 48. Sample Mode 2 Direction Order Routing Path

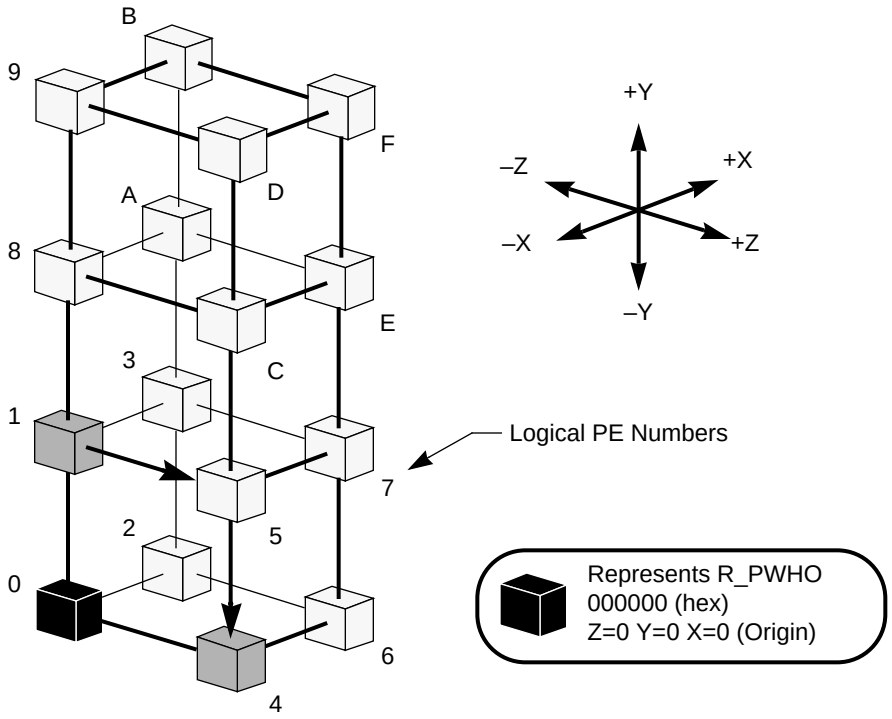


Table 15 shows possible values for all 4 R_NET_LUT registers used in physical node R_PWHO 000100 (hex) shown in Figure 48.

Table 15. R_NET_LUT Register Values for R_PWHO 000100 (hex)

Register	X Dimension		Y Dimension		Z Dimension		Initial Hop	Final Hop
	Address	Direction	Address	Direction	Address	Direction		
R_NET_LUT[0]	00x †	+	0000x	–	0000	+	None	None
R_NET_LUT[1]	00x	+	0000x	–	0001	+	None	None
R_NET_LUT[2]	00x	+	0001x	+	0000	+	None	None
R_NET_LUT[3]	00x	+	0001x	+	0001	+	None	None

† All addresses are binary.

R_LUT_OVERRIDE Register

Each network router has one network router look-up table override (R_LUT_OVERRIDE) register. Software may signal the support circuitry to read routing-tag information for every packet it creates from this register instead of an incomplete or incorrect look-up table stored in the R_NET_LUT registers. This process is called mode 3. Table 16 shows the bit format of the R_LUT_OVERRIDE register.

In addition to storing the routing-tag information for mode 3 operation, software uses the R_LUT_OVERRIDE register to set the mode (mode 0, 1, 2, or 3) for the R_NET_LUT registers (refer again to Table 16).

Table 16. R_LUT_OVERRIDE Bit Format

Bits	Name	Description
<2 : 0>	X_ADR	These bits contain the X-dimension physical address of the destination node.
<6 : 3>	Not applicable	These bits are not used.
7	X_SGN	This bit indicates the X-dimension direction. 0 = +X 1 = –X
<12 : 8>	Y_ADR	These bits contain the Y-dimension physical address of the destination node.
<14 : 13>	Not applicable	These bits are not used.
15	Y_SGN	This bit indicates the Y-dimension direction. 0 = +Y 1 = –Y
<19 : 16>	Z_ADR	These bits contain the Z-dimension physical address of the destination node.
<22 : 20>	Not applicable	These bits are not used.
23	Z_SGN	This bit indicates the Z-dimension direction. 0 = +Z 1 = –Z
<31 : 24>	Not applicable	These bits are not used.
<33 : 32>	INITIAL_HOP	These bits indicate the initial hop direction. 00 = +X 10 = +Z 01 = +Y 11 = No initial hop
<35 : 34>	Not applicable	These bits are not used.
36	FINAL_HOP	This bit indicates the final hop direction. When set to 1, the Z_SGN and ADAPT bits must be set to 0. 0 = No final hop 1 = –Z
37	ADAPT	This bit indicates adaptive routing. 0 = Direction order routing 1 = Adaptive routing
<61 : 38>	Not applicable	These bits are not used.
<63 : 62>	MODE	00 = Mode 0 – Each R_NET_LUT register corresponds to one logical PE number. (Not valid for systems with more than 544 PEs.) 01 = Mode 1 – The support circuitry replaces bit 0 of the R_NET_LUT register with bit 0 of the logical PE number and replaces bit 8 of the R_NET_LUT register with bit 1 of the logical PE number. 10 = Mode 2 – The support circuitry replaces bit 0 of the R_NET_LUT register with bit 1 of the logical PE number and replaces bit 8 of the R_NET_LUT register with bit 0 of the logical PE number. 11 = Mode 3 – The support circuitry uses bits <37 : 0> of the R_LUT_OVERRIDE register instead of the R_NET_LUT registers.

R_VCSEL[5 : 0] Registers

Each network router has six virtual channel select (R_VCSEL[5 : 0]) registers (refer to Table 17). Software uses the virtual channel select registers to assign the dateline physical node in each direction and to optimize the virtual channel buffer usage in the interconnect network.

Table 17. R_VCSEL Register Assignments

Register	Assigned to
R_VCSEL[0]	+X Direction
R_VCSEL[1]	+Y Direction
R_VCSEL[2]	+Z Direction
R_VCSEL[3]	–X Direction
R_VCSEL[4]	–Y Direction
R_VCSEL[5]	–Z Direction

Virtual Channel Buffers

A virtual channel is created when information travels over the same physical communication link but is stored in different buffers. In the CRAY T3E system, the virtual channel buffers in each network router create virtual channels.

Each network router contains two sets of virtual channel buffers: set 0 and set 1. Each set contains two virtual channel buffers: one for request information and one for response information. The virtual channel buffers eliminate communication deadlock conditions that could occur in the interconnect network.

Hardware automatically controls the use of the request and response buffers in each virtual channel buffer set. Software may control the use of virtual channel set 0 or set 1.

To control the use of virtual channel set 0 or set 1, software sets a dateline physical node and may set other values in the VCSEL[5 : 0] registers.

Dateline Physical Node

Software designates one node in each direction of the interconnect network as the dateline physical node. When a packet of information that is using virtual channel set 0 travels through the dateline physical node, hardware automatically switches the virtual channel set for the packet to set 1. The packet then uses virtual channel set 1 as it completes hops in that direction.

When a packet of information that is using virtual channel set 1 travels through the dateline physical node, hardware automatically stops the transfer of the packet. If this occurs, it indicates that software stored incorrect values in the SELECT bits of a VCSEL[5 : 0] register, or another type of packet routing error occurred. The next page provides more information on the SELECT bits.

NOTE: When an initial hop is in the routing tag, the initial hop should not be to a dateline physical node.

R_VCSEL[5 : 0] Bit Format

Each R_VCSEL register has two types of bits: the DATELINE bit and the SELECT bits (refer to Table 18 on the next page). Software uses the DATELINE bit to indicate which physical node in each direction is the dateline physical node. Software uses the SELECT bits to modify the virtual channel buffer usage for optimum interconnect network performance.

The SELECT bits are used only to optimize interconnect network performance. Software may set these bits to all 0's. If software sets a SELECT bit so that a packet of information that is using virtual channel set 1 travels through the dateline physical node, the hardware automatically stops the transfer of the packet and the packet is lost.

Hardware may not use all of the SELECT bits for each direction. For example, in a system with an X=4 node shape, hardware uses only bits <3 : 0> of the R_VCSEL[0] register (+X direction register).

The support circuitry only references the R_VCSEL[5 : 0] registers when a packet changes into a new direction or when a packet changes from adaptive routing to direction order routing.

Table 18. R_VCSEL[5 : 0] Bit Format

Bits	Name	Description
0	SELECT	When set to 0, this bit indicates that a packet traveling in this direction from this physical node to a physical node with bits <18:16> bits <10:8>, or bits <2 : 0> of the physical node number set to 0 uses virtual channel set 0. When set to 1, this bit indicates that a packet traveling in this direction from this physical node to a physical node with bits <18:16> bits <10:8>, or bits <2 : 0> of the physical node number set to 0 uses virtual channel set 1.
1	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 1 in this direction.
2	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 2 in this direction.
3	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 3 in this direction.
4	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 4 in this direction.
5	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 5 in this direction.
6	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 6 in this direction.
7	SELECT	This bit corresponds to physical nodes in this direction that have a physical node number with bits <18:16> bits <10:8>, or bits <2 : 0> set to 7 in this direction.
<31 : 8>	Not applicable	These bits are not used.
32	DATELINE	When set to 0, this bit indicates that this physical node is not the dateline physical node for this direction. When set to1, this bit indicates that this physical node is the dateline physical node for this direction.
<63 : 33>	Not applicable	These bits are not used.

Adaptive Routing

Software uses adaptive routing to transfer data through the interconnect network in an adaptive routing path. When a packet of information follows an adaptive routing path, the packet completes hops in any direction as long as it travels toward the destination node using a minimal-hops path.

A packet that follows adaptive routing uses the same routing tag as a packet that uses direction order routing; however, it does not use the initial hop value and the final hop value. In addition, the packet may not complete all the hops in one direction before switching into another direction.

Adaptive Routing Registers

Before adaptive routing can occur, software must do the following:

- Set the ADAPT bit of the appropriate R_NET_LUT registers to 1
- Set the NO_ADAPT bit of the appropriate R_CHAN[5 : 0] registers to 0. This ensures that a packet following an adaptive routing path through the interconnect network does not try to travel over a communication link that is unconnected. A communication link may be unconnected because it is part of a partial plane or part of a small system configuration.

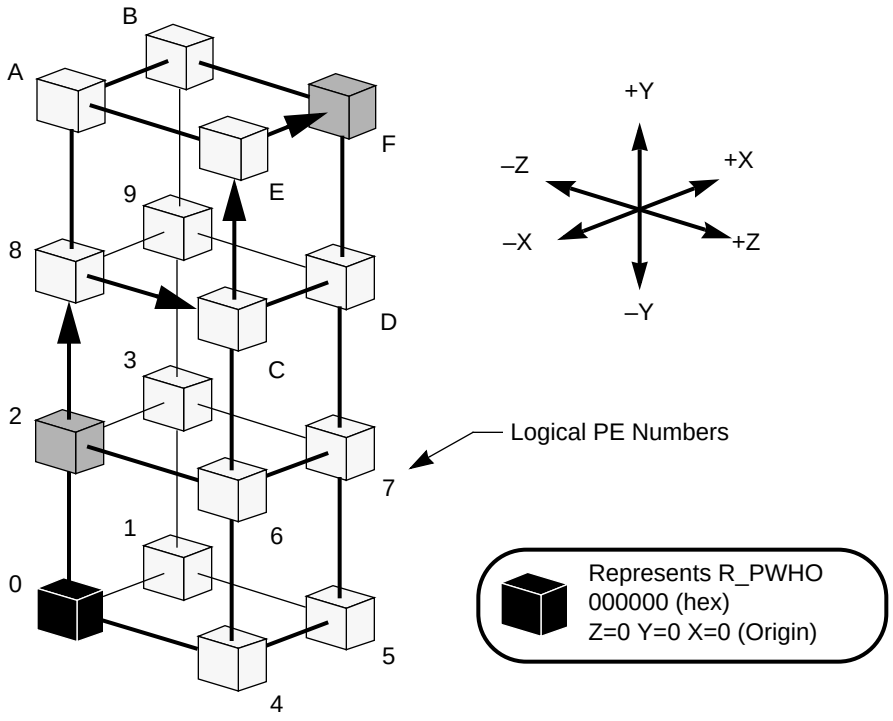
Adaptive Routing Path Example

Figure 49 shows the physical nodes for a 16-PE CRAY T3E system. For clarity, Figure 49 does not show the communication links that complete the torus in each dimension. For the following example, software assigned each physical node in the system the unique logical PE numbers shown in Figure 49.

Figure 49 shows one adaptive routing path from logical PE 2 [R_PWHO 000100 (hex)] to logical PE F (R_PWHO 010301 (hex)]. In physical node R_PWHO 000100 (hex), R_NET_LUT[15] register values for this routing are:

- No initial hop (INITIAL_HOP = 3)
- X direction = + (X_SGN = 0)
- X address = 1 (X_ADR = 1)
- Y direction = + (Y_SGN = 0)
- Y address = 3 (Y_ADR = 3)
- Z direction = + (Z_SGN = 1)
- Z address = 1 (Z_ADR = 1)
- No final hop (FINAL_HOP = 0)
- ADAPT = 1

Figure 49. Sample Adaptive Routing Path



Index

N

O

P

R

S

T

U

V

X

Y

Z