CRAY T3D ™ PE Configurations Overview

HMM-088-0

Cray Research Proprietary

Cray Research, Inc.

Record of Revision

REVISION   DESCRIPTION

May 1994. Original printing.

CRAY T3D PE Configurations Overview

The CRAY T3D system is available to customers in three types of configurations: multiple-cabinet (MC) systems, single-cabinet (SC) systems, and multiple-cabinet air-cooled (MCA) systems. The following subsections provide an overview of these systems.

1   System Cabinets

In a multiple-cabinet system, the CRAY T3D system is housed in one or more cabinets that each contain only CRAY T3D modules. The host system is housed in one or more separate cabinets. Figure 1 shows a maximum configuration, 2,048-processing element (PE) multiple-cabinet system.

Figure 1. Multiple-cabinet System Maximum Configuration

In a single-cabinet system, the CRAY T3D system is housed in the same cabinet as the host system. Figure 2 shows a maximum configuration, 256-PE single-cabinet system.

- **33 PE Modules (1 is a Spare)**
- **2 I/O Modules**
- **1 Clock Module**

- **4 CRAY Y-MP CPUs**
- **SRAM Memory (64 Mwords)**
- **4 I/O Clusters**
- **1 SSD-E/32i**
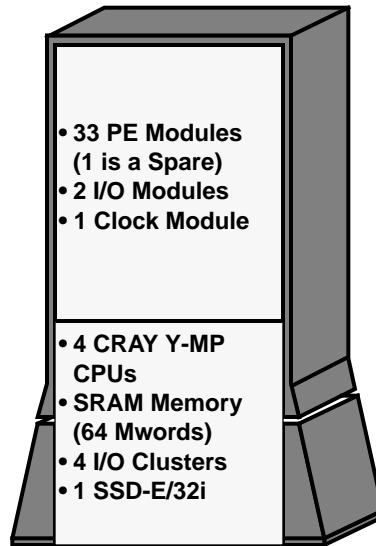
Figure 2. Single-cabinet System Maximum Configuration

In a multiple-cabinet air-cooled system, the CRAY T3D system is housed in one air-cooled cabinet that contains only CRAY T3D modules. The host system is housed in one or more separate cabinets. Figure 3 shows a maximum configuration, 128-PE multiple-cabinet air-cooled system.
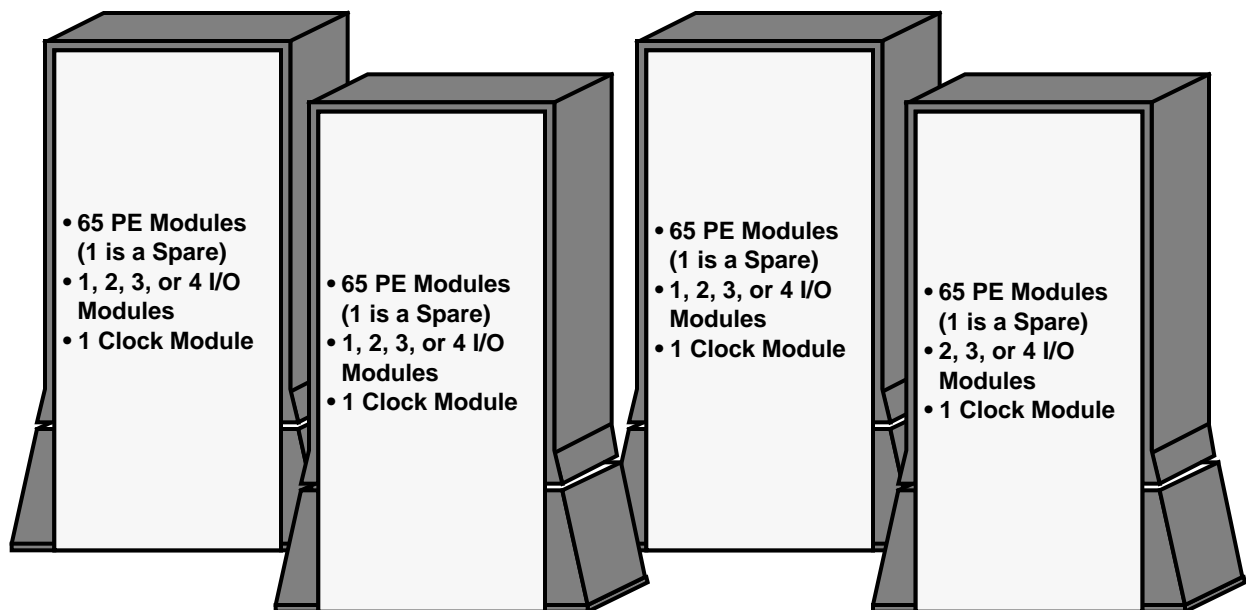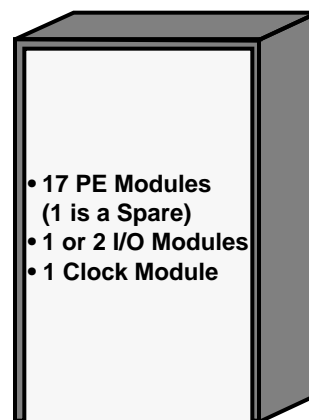

- **17 PE Modules (1 is a Spare)**
- **1 or 2 I/O Modules**
- **1 Clock Module**

Figure 3. Multiple-cabinet Air-cooled System Maximum Configuration

2   System Specifications

The following four tables list the specifications for each of the CRAY T3D system configurations. The following subsections define each specification.

Table 1. CRAY T3D Single-cabinet Systems Specifications (S/N 6101 – 6104)

CRAY T3D System	PE Node Shape in X, Y, Z	Spare PE Nodes	System Memory		Peak Performance (GFLOPS)	Global Bisection Bandwidth (Gbytes/s)	I/O Gateways
			4 Mbit DRAMs	16 Mbit DRAMs			
SC128	4, 4, 4	4	2 Gbytes (256 Mwords)	8 Gbytes (1 Gword)	19.2	19.2	2 or 4
SC256	4, 4, 8	4	4 Gbytes (512 Mwords)	16 Gbytes (2 Gwords)	38.4	19.2	2 or 4

Table 2. CRAY T3D Single-cabinet Systems Specifications (S/N 6105 – 6199)

CRAY T3D System	PE Node Shape in X, Y, Z	Spare PE Nodes	System Memory		Peak Performance (GFLOPS)	Global Bisection Bandwidth (Gbytes/s)	I/O Gateways
			4 Mbit DRAMs	16 Mbit DRAMs			
SC128	4, 4, 4	4	2 Gbytes (256 Mwords)	8 Gbytes (1 Gword)	19.2	19.2	2 or 4
SC160	4, 4, 5	4	2.5 Gbytes (320 Mwords)	10 Gbytes (1.25 Gwords)	24	19.2	2 or 4
SC192	4, 4, 6	4	3 Gbytes (384 Mwords)	12 Gbytes (1.5 Gwords)	28.8	19.2	2 or 4
SC224	4, 4, 7	4	3.5 Gbytes (448 Mwords)	14 Gbytes (1.75 Gwords)	33.6	19.2	2 or 4
SC256	4, 4, 8	4	4 Gbytes (512 Mwords)	16 Gbytes (2 Gwords)	38.4	19.2	2 or 4

Table 3. CRAY T3D Multiple-cabinet Air Cooled Systems Specifications

CRAY T3D System	PE Node Shape in X, Y, Z	Spare PE Nodes	System Memory		Peak Performance (GFLOPS)	Global Bisection Bandwidth (Gbytes/s)	I/O Gateways
			4 Mbit DRAMs	16 Mbit DRAMs			
MCA32	2, 4, 2	4	512 Mbytes (64 Mwords)	2 Gbytes (256 Mwords)	4.8	4.8	2
MCA64	4, 4, 2	4	1 Gbytes (128 Mwords)	4 Gbytes (512 Mwords)	9.6	9.6	2
MCA96	4, 4, 3	4 or 0 †	1.5 Gbytes (192 Mwords)	6 Gbytes (768 Mwords)	14.4	14.4	2 or 4 †
MCA128	4, 4, 4	4 or 0 †	2 Gbytes (256 Mwords)	8 Gbytes (1 Gword)	19.2	19.2	2 or 4 †

† If the system has four I/O gateways, the spare PE module must be removed.

Table 4. CRAY T3D Multiple-cabinet Systems Specifications (S/N 6001 – 6015)

CRAY T3D System	PE Node Shape in X, Y, Z	Spare PE Nodes	System Memory		Peak Performance (GFLOPS)	Global Bisection Bandwidth (Gbytes/s)	I/O Gateways
			4 Mbit DRAMs	16 Mbit DRAMs			
MC128	8, 4, 2	4	2 Gbytes (256 Mwords)	8 Gbytes (1 Gword)	19.2	9.6	2 or 4
MC256	8, 4, 4	4	4 Gbytes (512 Mwords)	16 Gbytes (2 Gwords)	38.4	19.2	2, 4, 6, or 8
MC512	8, 4, 8	4	8 Gbytes (1 Gword)	32 Gbytes (4 Gwords)	76.8	38.4	2, 4, 6, or 8
MC1024	8, 8, 8	8	16 Gbytes (2 Gwords)	64 Gbytes (8 Gwords)	153.6	76.8	2, 4, 6, 8, 10, 12, 14, or 16
MC2048	8, 16, 8	16	32 Gbytes (4 Gwords)	128 Gbytes (16 Gwords)	307.2	76.8	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32

Table 5. CRAY T3D Multiple-cabinet Systems Specifications (S/N 6016 – 6099)

CRAY T3D System	PE Node Shape in X, Y, Z	Spare PE Nodes	System Memory		Peak Performance (GFLOPS)	Global Bisection Bandwidth (Gbytes/s)	I/O Gateways
			4 Mbit DRAMs	16 Mbit DRAMs			
MC128	4, 4, 4	4	2 Gbytes (256 Mwords)	8 Gbytes (1 Gword)	19.2	19.2	2 or 4
MC160	4, 4, 5	4	2.5 Gbytes (320 Mwords)	10 Gbytes (1.25 Gwords)	24	19.2	2 or 4
MC192	4, 4, 6	4	3 Gbytes (384 Mwords)	12 Gbytes (1.5 Gwords)	28.8	19.2	2, 4, or 6
MC224	4, 4, 7	4	3.5 Gbytes (448 Mwords)	14 Gbytes (1.75 Gwords)	33.6	19.2	2, 4, or 6
MC256	4, 4, 8	4	4 Gbytes (512 Mwords)	16 Gbytes (2 Gwords)	38.4	19.2	2, 4, 6, or 8
MC320	8, 4, 5	4	5 Gbytes (640 Mwords)	20 Gbytes (2.5 Gwords)	48	24	2, 4, 6, or 8
MC384	8, 4, 6	4	6 Gbytes (768 Mwords)	24 Gbytes (3 Gwords)	57.6	28.8	2, 4, 6, or 8
MC448	8, 4, 7	4	7 Gbytes (896 Mwords)	28 Gbytes (3.5 Gwords)	67.2	33.6	2, 4, 6, or 8
MC512	8, 4, 8	4	8 Gbytes (1 Gword)	32 Gbytes (4 Gwords)	76.8	38.4	2, 4, 6, or 8
MC640	8, 8, 5	8	10 Gbytes (1.25 Gwords)	40 Gbytes (5 Gwords)	96	48	2, 4, 6, 8, 10, 12, 14, or 16
MC768	8, 8, 6	8	12 Gbytes (1.5 Gwords)	48 Gbytes (6 Gwords)	115.2	57.6	2, 4, 6, 8, 10, 12, 14, or 16
MC896	8, 8, 7	8	14 Gbytes (1.75 Gwords)	56 Gbytes (7 Gwords)	134.4	67.2	2, 4, 6, 8, 10, 12, 14, or 16
MC1024	8, 8, 8	8	16 Gbytes (2 Gwords)	64 Gbytes (8 Gwords)	153.6	76.8	2, 4, 6, 8, 10, 12, 14, or 16
MC1280	8, 16, 5	16	20 Gbytes (2.5 Gwords)	80 Gbytes (10 Gwords)	192	48	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32
MC1536	8, 16, 6	16	24 Gbytes (3 Gwords)	96 Gbytes (12 Gwords)	230.4	57.6	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32
MC1792	8, 16, 7	16	28 Gbytes (3.5 Gwords)	112 Gbytes (14 Gwords)	268.8	67.2	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32
MC2048	8, 16, 8	16	32 Gbytes (4 Gwords)	128 Gbytes (16 Gwords)	307.2	76.8	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32

2.1   Processing Element Node Shape

The processing element node shape indicates how the communication links connect the processing element nodes. For example, Figure 4 shows a processing element node shape of X=4 Y=4 Z=4. For clarity the figure does not show the communication links that complete the torus in each dimension and does not show the Y-dimension communication links.
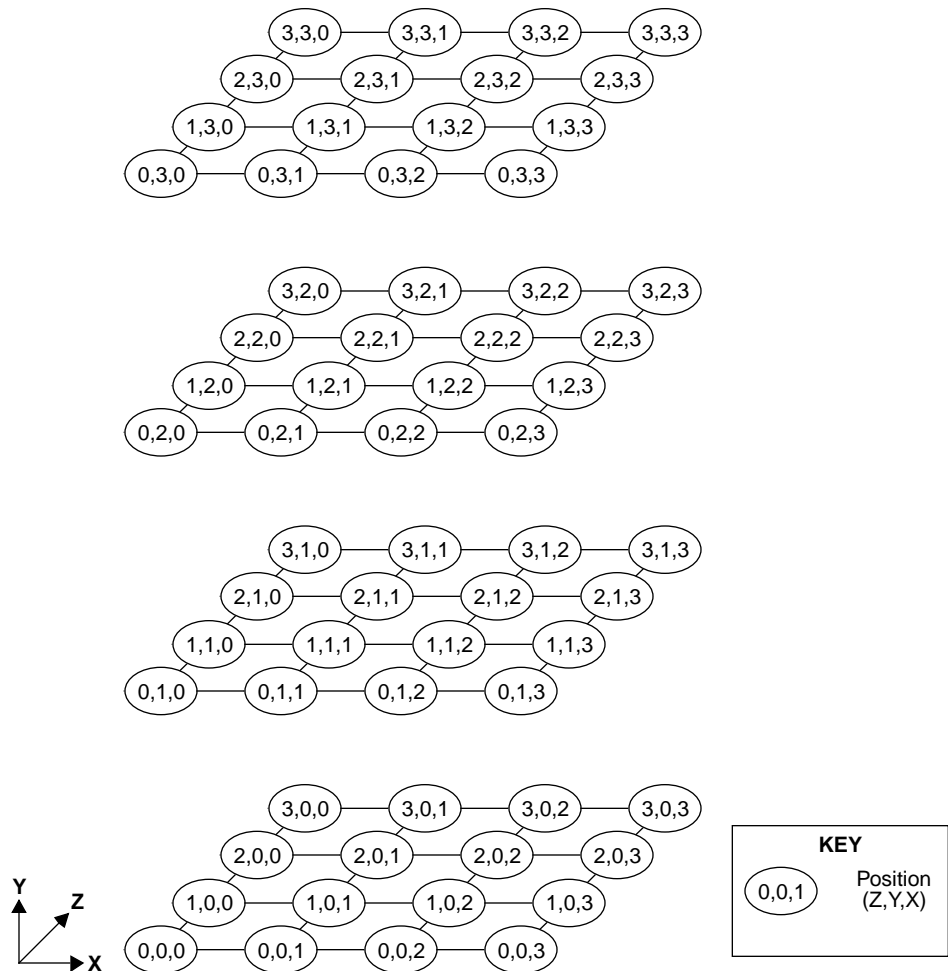
Figure 4. Processing Element Node Shape of X=4 Y=4 Z=4

2.2   Spare Processing Element Nodes

Each cabinet in a CRAY T3D system contains one spare processing element module. Each spare processing element module contains four spare processing element nodes.

2.3 **System Memory**

The size of system memory is the size of local memory in one PE multiplied by the total number of PEs in the system. The size of local memory in each PE is 16 Mbytes (2 Mwords) if the PE contains 4-Mbit dynamic random access memory (DRAM) integrated circuits, or it is 64 Mbytes (8 Mwords) if the PE contains 16-Mbit DRAM integrated circuits.

For example, a CRAY T3D system with 512 PEs, each with 16 Mbytes of local memory, has a total system memory of 8 Gbytes (1 Gword).

2.4 **Peak Performance**

In the CRAY T3D system, the peak performance of one microprocessor is 150 million floating-point operations per second (MFLOPS). The peak performance of a system is the peak performance of one microprocessor multiplied by the total number of microprocessors in the system.

For example, a CRAY T3D system with 512 PEs has a peak performance of 150 MFLOPS x 512 PEs = 76.8 GFLOPS.

2.5   Global Bisection Bandwidth

Calculating the global bisection bandwidth is a six-step process. The global bisection bandwidth of a system configuration should never be calculated by multiplying the global bisection bandwidth of a smaller configuration by a constant value. The following steps describe how to calculate the global bisection bandwidth for a system configuration.

1.   Determine the bandwidth of one unidirectional channel that connects two nodes. In the CRAY T3D system, the bandwidth of one unidirectional channel is given in Equation 1.

$$\text{Unidirectional channel bandwidth} = 300 \text{ Mbytes/s} \qquad \textbf{(EQ 1)}$$

2.   Calculate the bandwidth of one communication link. In the CRAY T3D system, each communication link contains two unidirectional channels (refer to Equation 2).

$$\text{Communication link bandwidth} = 300 \text{ Mbytes/s} \times 2 = 600 \text{ Mbytes/s} \qquad \textbf{(EQ 2)}$$

3.   Determine the maximum number of processing element nodes in each dimension of the system. For example, in the CRAY T3D MC2048 system, these values are Z=8 Y=16 X=8.

4.   Determine the location of the system's bisection plane. A bisection plane is an imaginary plane that bisects the *largest* dimension in the system. For example, Figure 5 shows the bisection plane for a CRAY T3D MC2048 system.
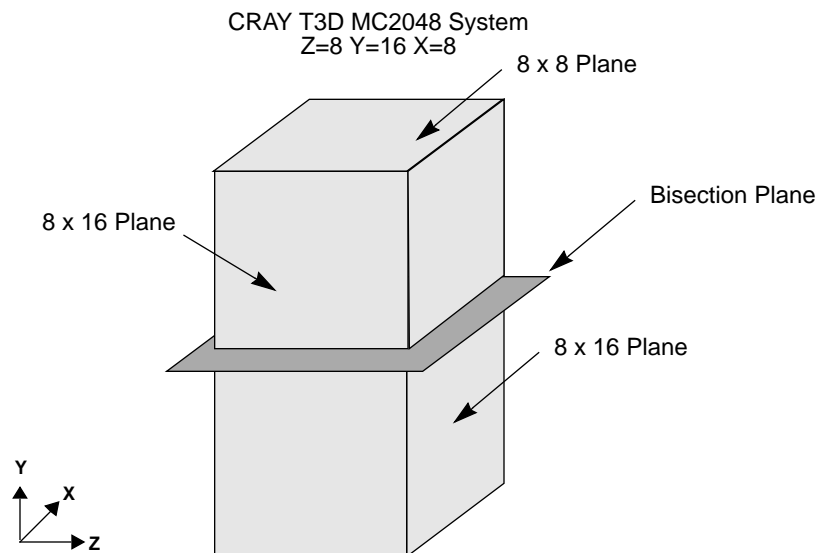
Figure 5. Bisection Plane

5. Calculate the number of communication links that cross the bisection plane. Calculate this value by multiplying the maximum number of processing element nodes in the smallest dimension by the maximum number of processing element nodes in the next smallest dimension, and multiplying the result by two. For example, Equation 3 shows how to calculate the number of communication links that cross the bisection plane in a CRAY T3D MC2048 system.

$$Z \times X \times 2 = 8 \times 8 \times 2 = 128 \text{ Communication Links} \qquad \textbf{(EQ 3)}$$

NOTE: Multiply the maximum number of nodes in the Z and X by two to include the communication links that complete the torus in the Z and X dimensions. As a simple example, Figure 6 shows how two of the communication links in the Y dimension cross the bisection plane in a CRAY T3D MC1024 system.
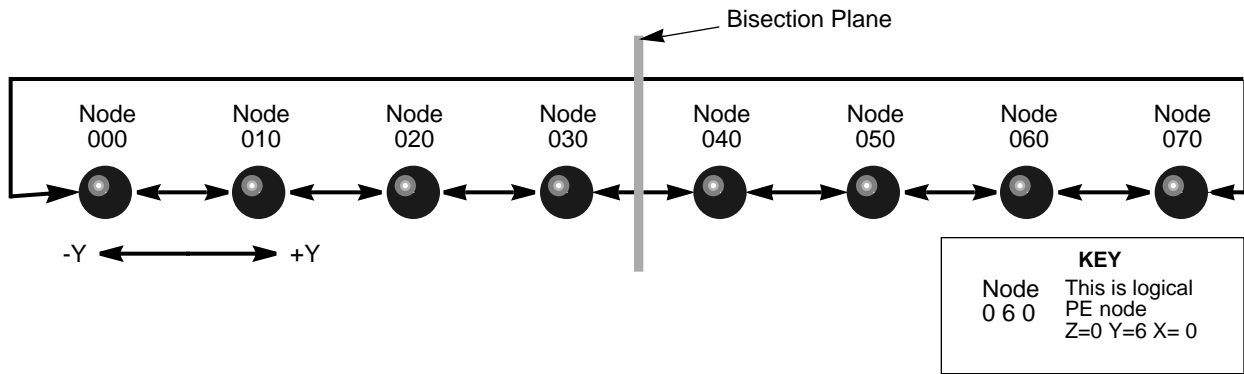

Figure 6. Torus Communication Link and Bisection Plane

6. Calculate the global bisection bandwidth by multiplying the total number of communication links that cross the bisection plane by the bandwidth of one communication link. For example, Equation 4 shows how to calculate the global bisection bandwidth for a CRAY T3D MC2048 system.

$$128 \text{ Communication Links} \times 600 \text{ Mbytes/s} = 76.8 \text{ Gbytes/s} \qquad \textbf{(EQ 4)}$$

2.6   I/O Gateways

The I/O gateways connect the CRAY T3D system to the host system in three types of cabling configurations: phase 1, phase 2, and phase 3. The host system manages tape, network, or other nondisk device I/O. The host system manages disk I/O for phases 1 and 2, and the CRAY T3D system manages disk I/O for phase 3.

2.6.1   Phase 1 Configuration

Phase 1, which is provided with initial CRAY T3D systems, connects a master I/O gateway to the host system over the high speed (HISP) channel. All configurations of the CRAY T3D system must have at least one phase 1 channel configuration (at least one master I/O gateway).

The HISP and LOSP channels from the master I/O gateway connect to the circuitry on a CPU module or shared I/O module in the host system. Figure 7 shows the phase 1 channel configuration of a master I/O gateway.
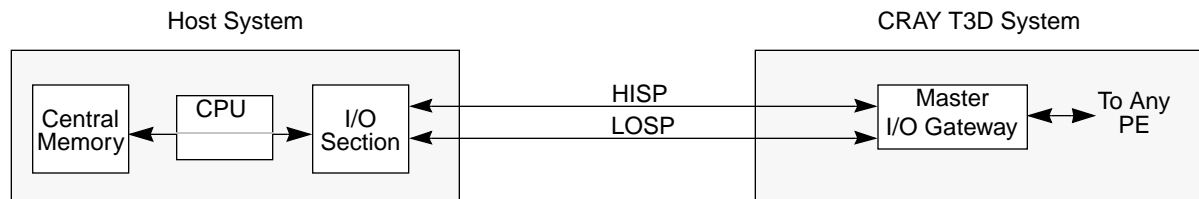
Figure 7. Phase 1 Master I/O Gateway Channel Configurations

When the CRAY T3D system connects to the host system through a master I/O gateway, the CPU in the host system controls all input and output. For example, the CPU controls tape, network, and disk device input and output. For phase 1 configurations, data passes through the CPU and to the master I/O gateway.

2.6.2   Phase 2 Configuration

Phase 2, which will be available in the first half of 1994, connects a slave I/O gateway to an input/output cluster (IOC) that is also connected to a CPU. Figure 8 shows the phase 2 channel configuration of a slave I/O gateway.

Host System   CRAY T3D System
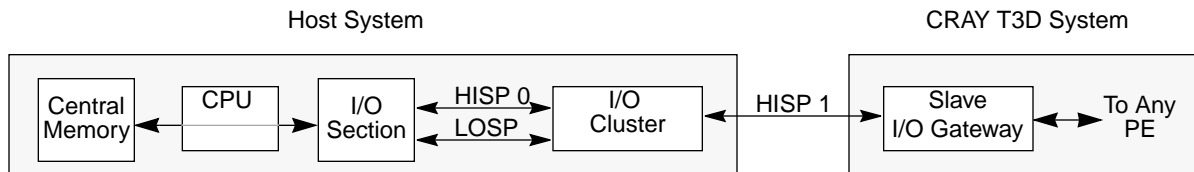

Figure 8. Phase 2 Slave I/O Gateway Channel Configuration

When the CRAY T3D system connects to the host system using the phase 2 channel configuration, the CPU in the host system controls all input and output. For example, the CPU controls tape, network, and disk device I/O.

Although the CPU controls the input and output, a HISP data path connects the CRAY T3D system to the IOC. This provides a path for data to travel between the CRAY T3D system and disk devices without traveling through the circuitry on a CPU module.

The phase 2 configuration uses the back-door HISP software support that is also used for the SSD solid state storage device model E (SSD-E). If the slave I/O gateway is connected to an IOC in this configuration, the SSD-E in the host system cannot be configured with back-door capability to the same IOC.

2.6.3   Phase 3 Configuration

Phase 3, which will be available in 1995, connects a slave I/O gateway to an IOC. Figure 9 shows the phase 3 configuration of a slave I/O gateway.
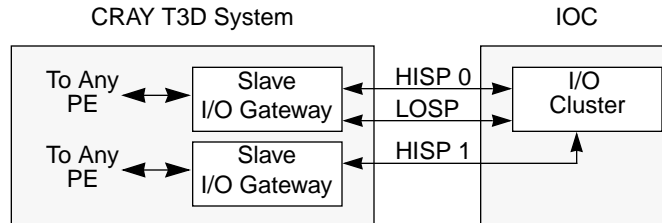

Figure 9. Phase 3 Slave I/O Gateway Channel Configuration

When the CRAY T3D system connects to an IOC using a phase 3 channel configuration, the CRAY T3D system controls disk device input and output. In addition, a HISP data path connects the CRAY T3D system to the IOC.