Chapter 15

Architecture Overview

A Connection Machine CM-5 system contains thousands of computational processing nodes, one or more control processors, and I/O units that support mass storage, graphic display devices, and VME and HIPPI peripherals. These are connected by the Control Network and the Data Network. (For a high-level sketch of these components, see Figure 27.)

15.1 Processors

Every processing node is a general-purpose computer that can fetch and interpret its own instruction stream, execute arithmetic and logical instructions, calculate memory addresses, and perform interprocessor communication. The processing nodes in a CM-5 system can perform independent tasks or collaborate on a single problem. Each processing node has 8, 16, or 32 Mbytes of memory; with the high-performance arithmetic accelerator, it has 32 or 128 Mbytes of memory and delivers up to 160 Mops or 160 Mflops.

The control processors are responsible for administrative actions such as scheduling user tasks, allocating resources, servicing I/O requests, accounting, enforcing security, and diagnosing component failures. In addition, they may also execute some of the code for a user program. Control processors have the same general capabilities as processing nodes but are specialized for performing managerial functions rather than computational functions. For example, control processors have additional I/O connections and lack the high-performance arithmetic accelerator. (See Figure 28.)

In a small system, one control processor may play a number of roles. In larger systems, individual control processors are often dedicated to particular tasks and

Copyright © 1993 Thinking Machines Corporation

referred to by names that reflect those tasks. Thus, a control processor that manages a partition and initiates execution of applications on that partition is referred to as a partition manager (PM), while a processor that controls an I/O device is called an I/O control processor (IOCP).

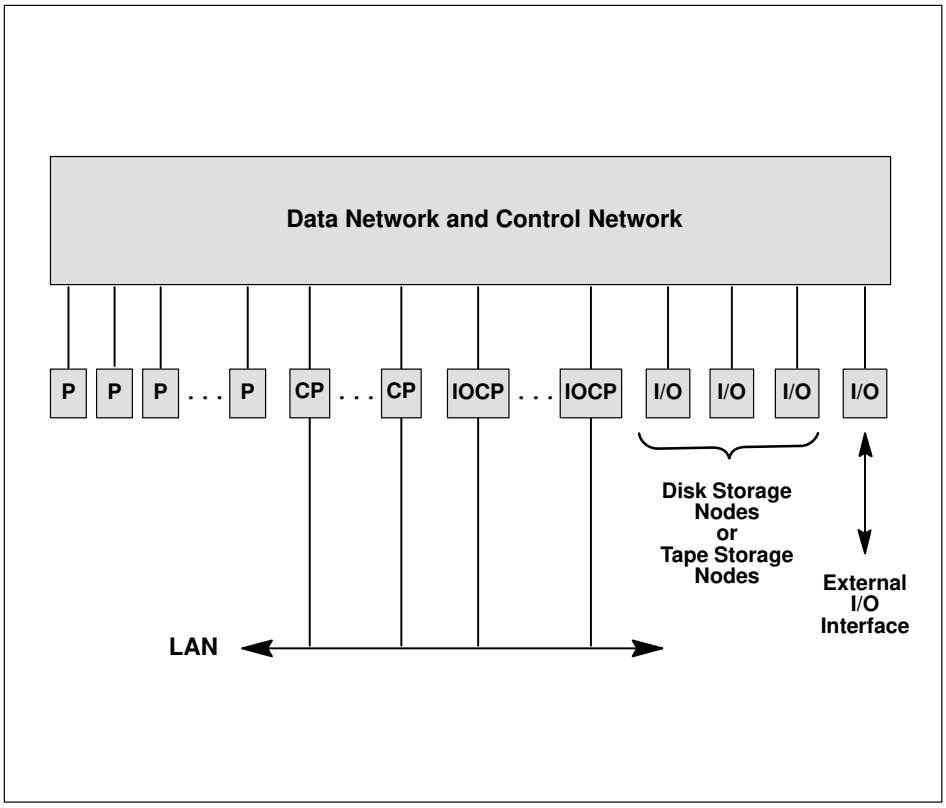

Figure 27. System components.

A CM-5 system contains tens, hundreds, or thousands of processing nodes, each with up to 160 Mflops of 64-bit floating-point performance. It also contains a number of I/O devices (disk storage nodes or tape storage nodes) and external connections (such as FDDI or HIPPI). The number of I/O devices and external connections is independent of the number of processing nodes. Both processing and I/O resources are managed by a relatively small set of control processors. All these components are uniformly integrated into the system by two internal communications networks, the Control Network and the Data Network. The Control Network provides multiway operations that can coordinate thousands of participants, while the Data Network supports high-bandwidth bulk data transfers. The capacity of each network scales up with the size of the system; every processing node or I/O device gets the network capacity it needs.

Copyright © 1993 Thinking Machines Corporation

15.2 Networks and I/O

The Control Network provides tightly coupled communications services. It is optimized for fast response (low latency). Its functions include synchronizing the processing nodes, broadcasting a single value to every node, combining a value from every node to produce a single result, and computing certain parallel prefix operations.

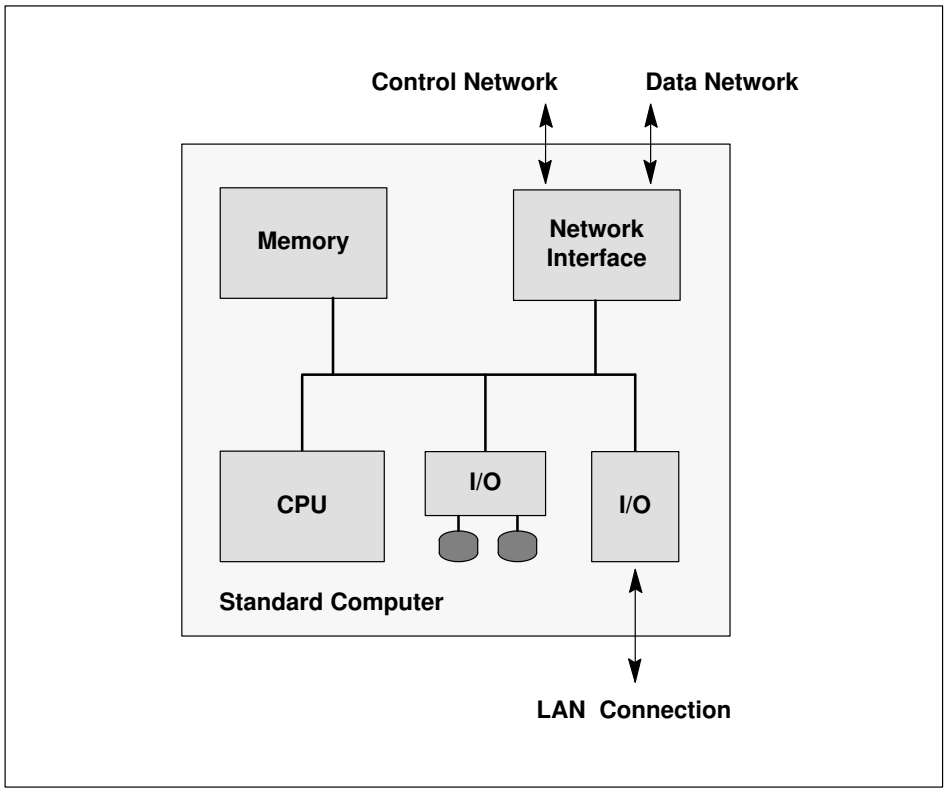

Figure 28. Control processor.

The basic CM-5 control processor consists of a RISC microprocessor, memory subsystem, I/O (including local disks and Ethernet connections), and a CM-5 Network Interface, all connected to a standard 64-bit bus. Except for the Network Interface, this is a standard off-the-shelf workstation-class computer system. The Network Interface connects the control processor to the rest of the system through the Control Network and Data Network. Each control processor runs CMOST, a UNIX-based operating system with extensions for managing the parallel-processing resources of the CM-5. Some control processors are used to manage computational resources and some are used to manage I/O resources.

The Data Network provides loosely coupled communications services. It is optimized for high bandwidth and excellent price/performance at any machine size.

Copyright © 1993 Thinking Machines Corporation

Its basic function is to provide point-to-point data delivery for tens of thousands of items simultaneously. Special cases of this functionality include nearest-neighbor communication and FFT butterflies. Communications requests and data delivery need not be synchronized. Once the Data Network has accepted a message, it takes on all responsibility for its eventual delivery; the sending processor can then perform other computations while the message is in transit. Recipients may poll for messages or be notified by interrupt on arrival. The Data Network also transmits data between the processing nodes and I/O units.

A standard Network Interface (NI) connects each node or control processor to the Control Network and Data Network. This is a memory-mapped control unit; reading or writing particular memory addresses will access network control registers or trigger communication operations.

The I/O units are connected to the Control Network and Data Network in exactly the same way as the processors, using the same Network Interface. Many I/O devices require more data bandwidth than a single NI can provide; in such cases multiple NI units are ganged. For example, a CM5-HIPPI channel interface contains 6 NI units, which provide access to 6 Data Network ports, covering 24 network addresses. (At 20 Mbytes/sec apiece, 6 NI units provide enough bandwidth for a 100 Mbyte/sec HIPPI interface with some to spare.)

Individual I/O devices are controlled by dedicated I/O control processors (IOCP). Some I/O devices are interfaces to external buses or networks; these include interfaces to VME buses and HIPPI channels. Noteworthy features of the I/O architecture are that I/O and computation can proceed independently and in parallel, that data may be transferred between I/O devices without involving the processing nodes, and that the number of I/O devices may be increased completely independently of the number of processing nodes.

Hiding in the background is a third network, the Diagnostic Network. It can be used to isolate any hardware component and to test both the component itself and all connections to other components. The Diagnostic Network pervades the hardware system but is completely invisible to the user; indeed, it is invisible to most of the control processors. A small number of the control processors include command interfaces for the Diagnostic Network; at any given time, one of these control processors provides the System Console function.

Copyright © 1993 Thinking Machines Corporation

15.3 Further Information

The following chapters discuss the CM-5 architecture in more detail.

Chapter 16 contains a sketch of the user-level *virtual machine*, the programming model that is visible to a single user job. This virtual machine is supported by a combination of hardware, operating system, and run-time libraries.

In Chapter 17, *local architecture* is considered: the structure of individual processors and associated memory. This is the view seen from any single processor in the system; it is the level of architecture where program code is executed.

Chapter 18 discusses *global architecture*. This specifies how various components of the system operate together to solve a single problem. This level of architectural specification provides a framework for understanding the flow of control and the management of data in a massively parallel application.

Chapter 19 describes the *system architecture*, which addresses support of multiple user jobs, communication between jobs, I/O transfers, fault diagnosis and repair, and system administration.

Chapter 20 presents the *I/O architecture*, including the design of individual I/O devices and how they fit into the system structure.

Copyright © 1993 Thinking Machines Corporation