Glossary

HMM-144-A
CRAY T3E™ Series
Last Modified: August 1997

Record of Revision

February 1996

Original printing.

Revision A: August 1997

Adds and updates terminology.

A

Acknowledge (Ack) packet

The Ack packet informs the PE that initiated a message that the destination PE accepted the message.

Actor

A resource allocation entity in the UNICOS/mk™ operating system. Within UNICOS/mk, all user processes, servers, and daemons are actors.

Adaptive routing

A method of routing information through an interconnect network; the information may complete hops in any direction as long as it travels toward the destination node via a minimal-hops path.

Address space

A set of contiguous addresses that reference data or control information.

Application

A program that runs on multiple PEs.

Application PE

A PE that is configured to run user applications.

Argument

Information that is used in addition to an address. For example, the more operands (MO) pointer is an address argument.

Atomic E-register commands

Commands that read, modify, and/or write system data in one indivisible operation.

Atomic memory mover

The atomic memory mover moves data from an old physical page to a new physical page in local memory.

Auxiliary registers

A set of registers in a PE that the microprocessor accesses directly without using the E registers.

B

Backmap table

The backmap table stores partial address tags of the microprocessor's secondary cache and is used to maintain cache coherency.

Barrier/eureka (B/E) circuit

See Barrier/eureka synchronization unit (BESU).

Barrier/eureka (B/E) context

A set of B/E circuits. In the CRAY T3E system, each B/E context contains two B/E circuits.

Barrier/eureka synchronization unit (BESU)

BESUs provide a a low-latency method of synchronizing PEs in the CRAY T3E system. Each processing element node in the CRAY T3E system supports 32 separate BESUs. Each BESU functions as an independent synchronization resource.

Barrier synchronization

A method that programmers use to ensure that all PEs within a partition reach a specified point in the program before the PEs continue with the next task.

Base virtual address

The base virtual address is added to an index offset to create the final user virtual address or global virtual address. The base virtual address is stored in the more operand (MO) block of E registers.

Block

A unit of data, such as a 128-bit block of data or a 256-bit block of data. A block of data is not a fixed size.

Boot code

The boot code initializes the memory-mapped registers in the support circuitry.

Boot node

The boot node receives boot code, PAL code, and configuration information from the I/O controller.

C

Capability

A general class of structures that are generated to facilitate quick access during subsequent references to an object (for example, file structures).

Centrifuge

A component in the support circuitry that separates the global index into a PE number (virtual PE number or logical PE number) and an index offset.

Centrifuge mask

A software-defined value that indicates to the centrifuge which bits of the global index are the PE number and which bits are the index offset.

Command

1. Information generated by the microprocessor that signals the support circuitry to perform functions on system data. The microprocessor sends a command to the support circuitry by placing information on the microprocessor command pins and, in some cases, the microprocessor address pins. 2. A user process that runs in a single PE.

Command PE

A PE that is configured to run user commands and single-PE jobs.

Communication link

A component of the interconnect network that transfers data and control information between two network routers in the interconnect network. A communication link is actually two unidirectional channels in one dimension.

Comperand

A value that the support circuitry compares to data that is read from a memory location. The support circuitry uses the comperand while performing a compare-and-swap (CSWAP) E-register command.

Configuration information

The configuration information identifies the type of system (liquid cooled or air cooled), the size of the system (number of PEs), etc.

Content addressable memory (CAM)

Data registers in the support circuitry of a PE that store data and address tags. The support circuitry maintains memory coherency by comparing the address for a data request to the address tags in CAM to ensure that the requested data is not on its way to the microprocessor.

Context

A group of registers or a set of circuitry. For example, there are two contexts of E registers: E-register context 0 and E-register context 1. Software assigns the access permissions for each context.

Credentials

Security information (for example, user ID, permission bits, etc.) that indicates whether the requester has permission to access a file.

D

Data cache

A data cache is a small, high-speed random-access memory that stores frequently or recently accessed data.

Data translation buffer (DTB)

A component in the microprocessor that translates data addressing information.

Dateline physical node

One node in each direction of the interconnect network that software designates to determine which set of virtual channels information will use to travel through the interconnect network.

Deterministic routing

See Direction order routing.

Direction order routing

Direction order routing is a method of routing information through the interconnect network where the information completes hops in the following order:

1. All the hops in the +X direction
2. All the hops in the +Y direction
3. All the hops in the +Z direction
4. All the hops in the –X direction
5. All the hops in the –Y direction
6. All the hops in the –Z direction

Direction order routing may also be referred to as deterministic routing.

Direct local memory access

Direct local memory access is a method of accessing data in local memory where the microprocessor places the actual local memory physical address on the microprocessor address pins. Software uses the direct local memory access method to transfer data between an internal microprocessor register and a physical address in local memory. Direct local memory access may also be referred to as a private data read or write.

Direct memory access (DMA) transfer

The DMA transfer enables a GigaRing™ client to read from and write to the memory of another GigaRing client.

Dirty data

Data that was modified by the microprocessor but was not transferred to the support circuity to be updated in local memory.

E

Eureka synchronization

A method in which all PEs stop searching for requested data when one PE finds the data.

E-register command

A command that signals the support circuitry to transfer data between the microprocessor and the E registers or to transfer data between the E registers and global memory.

E-register global-memory access method

A method of referencing data; the microprocessor uses the E registers as the source or destination for data transfers with global memory.

E-register local-memory access method

A method of referencing data; the microprocessor uses the E registers as the source or destination for data transfers with local memory.

E registers

Latency-hiding registers in the support circuitry that are the source or destination for all global data transfers.

E-register state (ERS) code bits

A pair of bits that is associated with each E register and indicates the present state of the E register.

Execution units

Components in the microprocessor that perform all arithmetic and logical instructions.

F

Final hop

One hop in the –Z direction that information may make after completing all of the hops indicated in the routing tag.

FLUSH

A command that the support circuitry sends to the microprocessor. The FLUSH command indicates that a remote PE is writing new data to a local memory location and old data from that memory location may be located in the secondary cache of the microprocessor. After receiving the FLUSH command, the microprocessor ensures that old data in the secondary cache is marked invalid.

Flush

The operation in which the microprocessor empties the contents of the write buffer and sends the contents to the support circuitry.

Fullweight IPC

Communication between actors that reside on different processing element nodes.

G

GigaRing channel

An interconnect system that is based on scalable coherent interface (SCI) technology. The CRAY T3E system uses the GigaRing channel to communicate with external devices.

Global index

The global index consists of a PE number (virtual PE number or logical PE number) and an index offset in a software-defined format.

Global memory

Global memory consists of the local memory of all PEs in the CRAY T3E system. You can calculate the global memory size of a CRAY T3E system by adding the local memory of all PEs.

Global page

A subset of the global virtual address. Each global page contains 512 Kbytes of address space.

Global segment

A portion of global memory that can be one of four sizes: 256 Mbytes, 512 Mbytes, 1 Gbyte, or 2 Gbytes. The size of the global segment, which is usually equal to or greater than the size of local memory, is determined by the GTB_CTL register. A global segment is further divided into a variable number of 512-Kbyte global pages. The number of global pages is determined by the global segment size. Each global page (or contiguous global pages) maps to physical pages.

Global translation array (GTA)

A software-defined look-up table that is stored in the local memory of each PE and used to translate global virtual addresses into physical addresses. The GTB_CTL register bits <5 : 4> determine the size of the global segments. The GTB_EA_BASE register specifies the table's starting location in memory.

Global translation array (GTA) entry

Each GTA entry contains a page frame number, a page size mask, a small-page bit, and a valid bit.

Global translation buffer (GTB)

A buffer in the support circuitry that stores frequently or recently accessed global translation array entries.

Global virtual address

An intermediate address that the operating system uses to reference different types of data in the CRAY T3E memory.

H

History buffer

A component in the support circuitry that detects the presence of streams.

History table

The history table stores memory addresses of data that was recently accessed by direct local memory reads; it is used to detect streams.

Hop

A transfer of information over one communication link in the interconnect network.

I

Index offset

A value that is added to a base virtual address to generate the final user virtual address or global virtual address.

Initial hop

One hop in the +X, +Y, or +Z direction that information may make before completing the hops that the routing tag indicates.

Interconnect network

The interconnect network provides communication paths among the nodes in the CRAY T3E system. The interconnect network connects the nodes in a three-dimensional matrix that consists of the X, Y, and Z dimensions. The highest- and lowest-numbered nodes in each dimension connect to form a torus (or loop), which improves resiliency and performance.

Interprocess communication (IPC)

The method that actors use to send and receive messages within the UNICOS/mk operating system. Lightweight IPC refers to communication between two actors that reside on the same physical processing element. Fullweight IPC refers to communication between actors that reside on different processing element nodes (also referred to as messaging).

I/O controller

The I/O controller transfers system data and control information between the CRAY T3E system and the GigaRing channel. The I/O controller in the CRAY T3E system consists of the I option and the GigaRing option.

I/O node (ION)

A device that connects to the GigaRing channel. An ION can be either an MPN (multipurpose node), SPN (single-purpose node), or an I/O controller in a mainframe.

L

Large page

A physical page in local memory that contains more than 512 Kbytes of address space.

Leading-zero count

The number of bits that are set to 0 in a set amount of data, starting with the most significant bit and ending with the first bit that is set to 1.

Lightweight IPC

Communication between two actors that reside on the same physical processing element.

Local index

A user virtual address or a global virtual address. The microprocessor places the local index on the data bus when it requests an E-register local-memory command.

Local memory

With respect to the microprocessor in a processing element, memory that is physically located in the same PE as the microprocessor.

Logically shared memory

A memory system that enables any microprocessor in the system to access the memory in another PE without involving the microprocessor in that PE.

M

Memory-mapped command

A command that the microprocessor issues by placing command information on the microprocessor command pins and by setting bit 39 of the microprocessor-generated address to 1.

Memory region

A portion of memory in the actor's address space.

Memory space

A set of contiguous addresses that the microprocessor can reference by using the microprocessor address pins.

Merge

The process in which the microprocessor write buffer combines 32- or 64-bit store instructions for the same 32-byte line into a single 32-byte write operation.

Message packet

A message packet contains a data payload that ranges from 0 to 32 64-bit words. This type of packet is used during peer-to-peer message transfers.

Messaging

Communication between actors that reside on different processing element nodes (also referred to as *fullweight IPC*).

Microkernel

A component of the UNICOS/mk operating system that supports a minimal set of primitives and the specifics of the logic design. A copy of the microkernel resides in the local memory of each PE in a CRAY T3E system.

Microkernel privileged architecture library (mkPAL)

The mkPAL software resides below the microkernel and provides a software interface to the low-level hardware functions of the microprocessor.

Microprocessor

In the CRAY T3E system, a reduced instruction set computer (RISC) 64-bit microprocessor.

Microprocessor-generated address

Information that the microprocessor places on the microprocessor address pins and sends to the support circuitry.

Miss

The process in which the microprocessor attempts to read data from the data cache or secondary cache, but the data is not in either location. When the microprocessor encounters a miss, it signals the support circuitry to retrieve data from memory.

Missed address file (MAF)

A component in the microprocessor that attempts to combine load instructions from the same 32-byte line into one 32-byte load operation.

More operands (MO) block of E registers

Four contiguous E registers that contain additional information for an E-register command. For example, the MO block of E registers may contain the centrifuge mask, the base virtual address, a comperand value, and a swaperand value.

More operands (MO) pointer

A field on the microprocessor data bus that indicates the location of the MO block of E registers when the microprocessor requests a global data transfer.

Multipurpose node (MPN)

One of the I/O nodes that connect to the GigaRing channel. The MPN interfaces the GigaRing channel to an SBus controller. The SBus controller in turn connects to a maximum of eight SBus devices. These devices include SCSI (small computer system interface) disk or tape devices, and ATM (asynchronous transfer mode), FDDI (fiber distributed data interface), and Ethernet network connections. MPN devices reside in the peripheral cabinet.

N

Network port

A pair of unidirectional channels that connects a PE to the interconnect network.

Network router

Network routers transfer information between PEs and the interconnect network in a CRAY T3E system. The network routers transfer data and control information through communication links.

Node

A point in the interconnect network where information can transfer from one communication link to another communication link or from one communication link to a PE. In the CRAY T3E system, a node contains one network port, six communication links, one network router, and one PE. A node contains a processing element (PE) and a network router.

Node shape

The number of physical nodes in each dimension of the interconnect network.

O

Opcode

A field in the microprocessor-generated address that indicates to the support circuitry which E-register command to perform.

Options file

The `bootsys`, `haltsys`, and `dumpsys` commands use the options file to determine which part of the system will be acted upon. (The options file is in the /opt/config directory on the SWS.)

Origin

The node to which software assigns a physical number of X=0, Y=0, and Z=0.

P

Packet

All request and response information is transferred through the network in the form of packets. A packet contains a header and a body.

Page

A page is a portion of memory and can vary in size.

Page frame number

The page frame number is combined with the segment offset (which is generated on the source PE) to form the 31-bit physical memory address on the destination PE. The page frame number is used for the upper portion of the address and the segment offset is used for the lower portion of the address; the page size mask is used to determine how many bits of each file are used.

Page size mask

The page size mask indicates which bits of the page frame number and which bits of the segment offset make up bits <26 : 16> of the physical address.

Partial plane

A set of nodes that connect in the same plane of the interconnect network; the set contains fewer than the maximum number of nodes allowed in the plane.

Partition

A group of processing element nodes that is assigned to one application at application run time.

Peer-to-peer message transfer

The peer-to-peer message transfer enables GigaRing clients to communicate with each other without negotiating transfer rates; this transfer does not require a response from the destination client.

Peripheral cabinet (PC-10)

The PC-10 is the SIO cabinet, which has 33 standard units (SUs) of configurable space. The PC-10 can hold a combination of MPN-1, NSR-1, DSS-1, and DSF-1 subracks.

Physical address

The physical address is a byte-oriented address that references a location in local memory.

Physically distributed memory

A memory system in which each physical segment is located in a different PE.

Physical page

A portion of local memory that can vary in size from 64 Kbytes to 128 Mbytes. The page size is specified by the page size mask in the global translation array table entry. Small pages (64 Kbytes to 256 Kbytes) map to the upper or lower portion of the 512-Kbyte physical page. Only one small page can map to a 512-Kbyte page.

Privileged architecture library (PAL) code

A set of subroutines that high-level software uses to interface with the hardware. The PAL code insulates the diagnostic, system, and user code from the hardware.

Population count

The total number of bits that are set to 1 in a fixed amount of data.

Port

A unique reference address that facilitates location-independent communication between actors. Each port is assigned to a specific actor and is supported by an associated message queue.

Prefetch

The process in which the support circuitry retrieves data from local memory in preparation for a load request from the microprocessor for that data.

Private data

See Direct local-memory access.

Processing element (PE)

A component of the CRAY T3E system; a PE contains a microprocessor, local memory, and support circuitry.

R

Registration

In the CRAY T3E system, registration occurs during the initialization of a command, when software indicates the addresses of all functions that are associated with the servers.

Remote memory

With respect to the microprocessor in a PE, memory that is physically located in another PE.

Routing tag

An address that indicates the path a packet will follow through the interconnect network to get to a destination PE.

S

Secondary cache

A component in the microprocessor that stores frequently or recently accessed data or instructions from local memory.

Segment offset

A component of the user virtual address or global virtual address that references a byte of data in a user segment or a global segment.

Segment translation table (STT)

A component in the support circuitry that translates a user segment into a global segment.

Server

A component of the UNICOS/mk operating system. A server is a statically linked executable file that supports a set of operating system services.

Single-purpose node (SPN)

A device on the GigaRing channel that provides a connection to a single type of peripheral. The various types of SPNs are the fibre channel node (FCN-1), intelligent peripheral interface (IPI) node (IPN-1), Enterprise System Connection (ESCON®) node (ESN-1), and High Performance Parallel Interface (HIPPI) node (HPN-1 or HPN-2). SPNs reside in a node subrack (NSR-1) within the peripheral cabinet.

Small page

A physical page in local memory that contains less than 512 Kbytes of address space.

Small-page bit

The small-page bit indicates whether the small page is located in the upper or lower portion of the 512-Kbyte global page.

Stream

A set of load commands that reference contiguous addresses in local memory.

Support circuitry

A component of a PE that extends the control and addressing functions of the microprocessor in the PE. The support circuitry contains a portion of the barrier circuitry and passes barrier signals between the microprocessor and the network router option.

Support PE

A processing element node that runs operating system software.

Swaperand

A value that the support circuitry writes into a memory location during a compare-and-swap (CSWAP) or atomic swap (SWAP) E-register command.

System routing tags

System routing tags indicate the path of a packet from the source physical node to a destination physical node.

T

Thread

A software abstraction that represents a unit of sequential execution.

Topology file

The topology file is located in the /opt/config directory on the SWS and is used during system initialization to determine the configuration of the GigaRing channels and associated I/O nodes. It also specifies which node on the ring is the maintenance node, identifies disabled nodes, and indicates folded or masked components on the ring.

Torus

A series of communication links that form a ring in which information can travel from one node, through all of the nodes in the same dimension, and back to the original node.

U

User PE

A processing element node that is configured to run user software. User PEs include *command PEs* and *application PEs*.

User process

An instance of a program. When a user process runs in one PE, it is referred to as a command; when it runs in multiple PEs, it is referred to as an application.

User segment

The user virtual address consists of eight equal address spaces called user segments.

User virtual address

A byte-oriented address that the program compiler generates and that references different types of data in an application’s address space.

V

Valid bit

A valid bit indicates that the current entry in a cache line or table contains valid information. (Hardware or software functions that modify data in system memory can cause an entry to be invalidated.)

Victim

A valid, dirty secondary cache entry that the microprocessor removes to make room for new data.

Virtual address space

Software maps the local memory of each PE into virtual address space. The virtual address space of a PE consists of 64 global segments.

Virtual channel

A virtual channel is created when different types of information travel over the same physical channel but are stored in separate channel buffers.

Virtual PE number

A number that an application uses to reference each PE in a partition.

W

Wrapper

Code that is added to existing UNICOS® code and allows it to function as a server in the UNICOS/mk environment.

Write buffer

A component of the microprocessor that temporarily stores write data before it transfers to the secondary cache or system memory.