3 MEMORY

The memory portion of the CRAY EL series system consists of four modules. These modules currently are provided in two types: a fully populated module and a half-populated module. Both module types are constructed on 16 x 22 x 0.093 inch printed circuit (PC) boards consisting of 16 circuit layers. The 16 layers are:

- 1 top pad
- 1 bottom pad
- 4 ground (Vss) layers
- 4 power layers containing 6-volt application-specific integrated circuits (ASICs) and 6-volt dynamic random access memory (DRAM) chips
- 6 signal layers

The logic portion of a memory PC board comprises two types of application-specific integrated circuits (ASICs) and a group of DRAM chips. There are nine ASICs on each memory board: two memory array control (MAC) ASICs and seven memory array data (MAD) ASICs. The number of DRAM chips on each memory PC board is determined by the memory size selected by the customer. Currently, the CRAY EL series system can be supplied with either a 32-Mword memory, a 64-Mword memory, a 128-Mword memory, a 256-Mword memory, or a 512-Mword memory.

Other important characteristics of CRAY EL series system memory modules are as follows:

- Each memory module contains 8 Mwords, 16 Mwords, or 64 Mwords of memory if half-populated and 32 Mwords or 128 Mwords if fully populated
- Each module operates with a 30-ns clock period.
- All memory modules are air cooled.

- The memory modules are connected to the backplane by two types of connectors:

  120 pins x 7 connectors = 840 pins
  90 pins x 1 connector = 90 pins
  930 pins

- The memory modules are connected to the individual processor modules using four ports per processor module; all ports are read/write ports.

- There is no hardware error logger. Single- and double-bit errors are reported to the operating system and stored in error files.

- Each module requires 270 watts of power, which equals 1080 watts per system.

- Each module requires 6 Vac.

Memory Configurations

The 32-Mword, 64-Mword, and 256-Mword memory configurations consist of half-populated memory boards that contain 288 DRAM chips. The 32-Mword and the 64-Mword options are configured by shorting address bit 2^{25}, effectively changing a standard 64-Mword module into a 32-Mword module. Each of the DRAM chips on the 32-Mword, 64-Mword, and 128-Mword memory is a 1 meg x 4 memory chip, and the 256-Mword and 512-Mword memory configurations is a 4 meg x 4 memory chip with a 70-nanosecond (ns) access time. The 128-Mword and 512-Mword memory configurations use 576 memory chips mounted on the fully populated memory board. Refer to Figure 3-1 for a diagram of a fully populated memory module.

Memory is divided into 16 banks for addressing in both half-populated and fully populated modules. The fully populated module contains an upper 16 banks and a lower 16 banks of addressable memory; the half-populated module uses only the lower 16 banks.

Because there are 16 banks of memory on each memory module, the mainframe has a total of 64 banks of memory. The memory is also separated into 4 sections; each memory module is 1 section. This arrangement provides the mainframe with 4 memory sections, each consisting of 16 banks.

† 1 meg x 4 means 1,048,576 (2^{20}) 4-bit address locations.

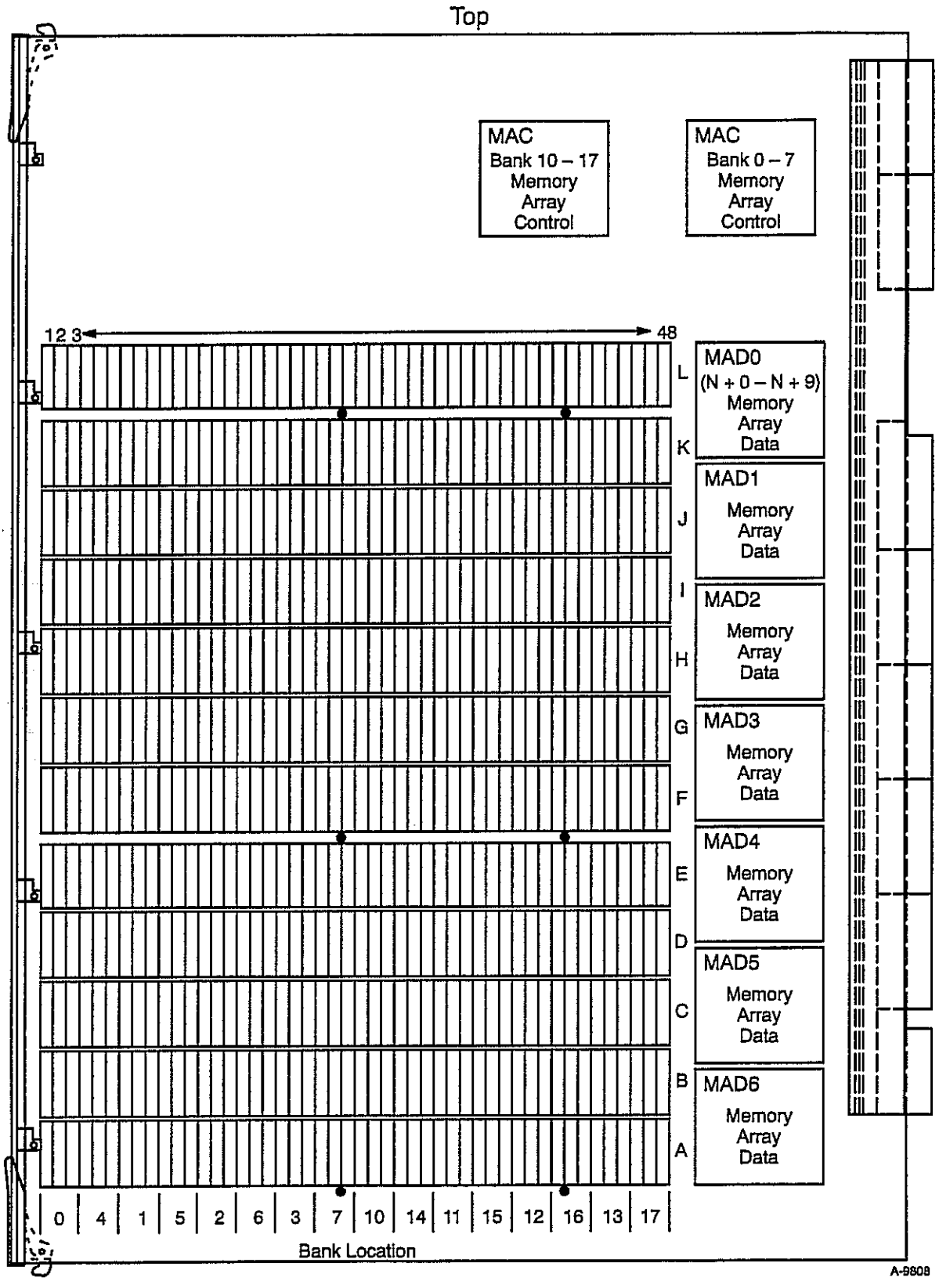

Figure 3-1. Fully Populated Memory Module

Memory Addressing

The CRAY EL series system uses a 32-bit address scheme, of which 29 bits (0 through 28) are used. The bits are assigned as shown in Figure 3-2. Bits 2^0 and 2^1 are used to select the appropriate section and bits 2^2 through 2^5 are used for bank selection.

The remainder of the address bits are used to select the actual memory location. The internal addressing scheme, represented by bits 2^6 through 2^{27}, uses a row and column scheme. The odd-numbered bits are used to count the rows of memory; the even-numbered bits are used to locate the column. To identify the failing memory module, decode bits 2^0 and 2^1 to determine which section is in error. This corresponds to a specific memory module. The locations of the memory sections are shown in Figure 3-3, a representation of the 8-slot mainframe card cage. All of the memory modules are interchangeable, so swapping the suspected failing module with a good module can help isolate the failure.

DRAM Specifications

NOTE: Because the DRAM chips used in the CRAY EL series system memory are 1 meg x 4 chips or 4 meg x 4 chips, it is expected that a single-bit error will rapidly escalate into a multiple-bit error. For this reason, it is important to repair single-bit errors as soon as possible.

Other specifications of the DRAM chip are:

- 0.4 x 1.03-inch ZIP package
- 5-clock-period bank access time
- Standby power = 5 milliwatts at 5 volts
- Active power = 550 milliwatts at 5 volts

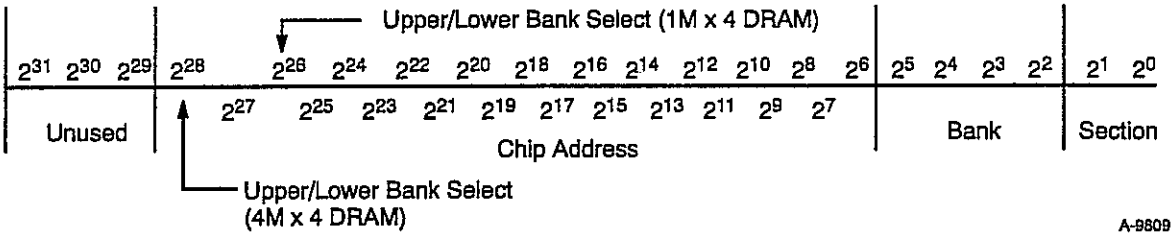

Figure 3-2. Address Bit Assignments

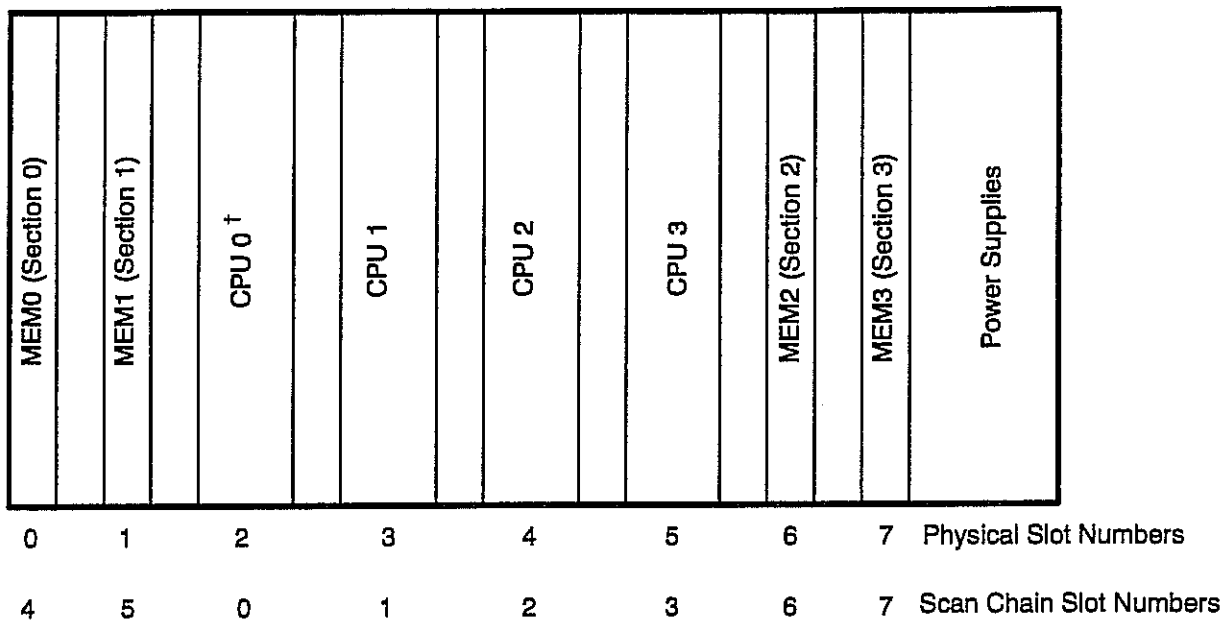

Figure 3-3. Processor Slot Configuration, Front View

Memory ASIC Description

The rest of the memory PC board is made up of two types of MAC and MAD ASICs. The MAC ASIC chip supports four memory operations:

- Read, a normal DRAM read operation, which lasts 5 clock periods.
- Write, a normal DRAM write operation, which lasts 5 clock periods.
- Refresh, which uses row address strobe (RAS) to refresh data, which lasts 5 clock periods.
- Read/Modify/Write (RMW), which is used during an exchange and uses normal DRAM read operations followed by a normal DRAM write operation to the same address. RMW lasts 10 clock periods.

The refresh control for the DRAM chips is located on the CPU and memory PC boards.

The MAC ASIC also performs the following functions:

- Connects any of the 4 processor modules to any of the 16 banks through an address crossbar
- Controls both address and control signals
- Holds a refresh address counter
- Controls all of the MAD ASICs

The MAD ASICs perform the following functions:

- Each handles a specific portion of the 72-bit memory data word
- Each contains a data crossbar that connects any of the 4 processor modules to any of the 16 banks

The interconnection between the MAD, MAC, and DRAM chips is represented in Figure 3-4. Note that MAC 0 controls operations for banks 0 through 7, and MAC 1 controls banks 10 through 17. The content of the entire data bus is presented to the seven MAD ASICs and each receives its assigned bits. The bits assigned to each MAD ASIC are in no specific order. If a random group of bits of data are reported as failing, it could, in fact, be a single MAD that is failing.

Memory Organization

Figure 3-5 shows the major architectural features of the CRAY EL series system memory. As shown, memory is divided into 4 sections, each containing 16 banks. This enables simultaneous memory references (two or more memory references that begin in the same clock period) and overlapping memory references (one or more memory references that begin while another reference is in progress).

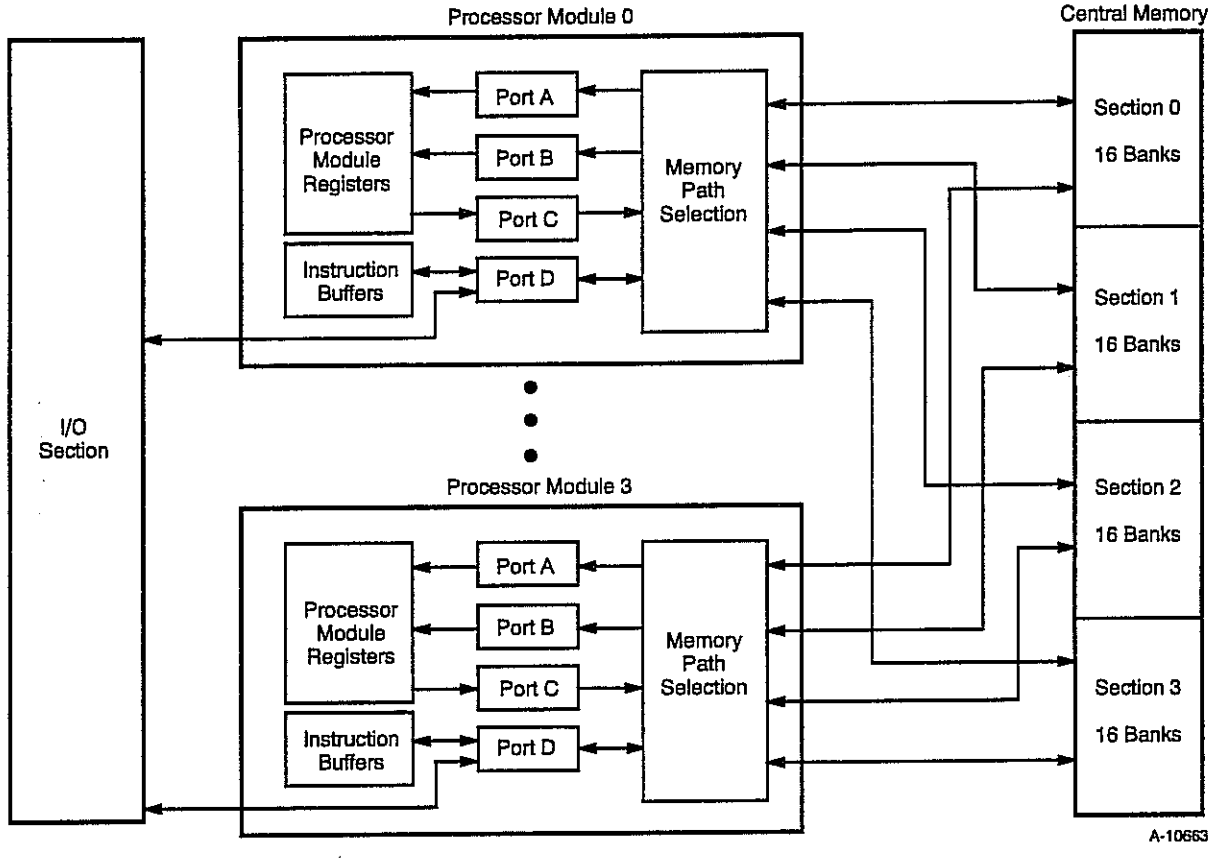

Figure 3-5. Central Memory Architecture

Memory Paths

Each processor module in the CRAY EL series system has an independent path into each of the memory sections. This allows a processor module to make four simultaneous memory references, one to each section. Overlapping memory references to different sections are allowed without restrictions. Likewise, overlapping memory references to the same memory section are allowed as long as each reference uses a different bank of memory. Because each processor module has only one

path into each section of memory, simultaneous memory references to the same section are not allowed, and one reference is forced to hold until the other reference is completed.

Simultaneous and overlapping memory references from two or more processor modules are subject to fewer restrictions. These memory references are allowed to one section of memory. The only restriction is that each reference must use a different bank. If simultaneous or overlapping memory references require the same bank, a priority definition is made between the requesting processor modules. Two priority schemes for simultaneous bank conflict are available. A scan-activated control bit determines which scheme is used. When the scan bit is a 0, a least recently used scheme is used to determine which processor module has highest priority. This means that whichever processor had access to memory least recently will have highest priority. The processor that had access to memory last (that is, most recently) will be given the lowest priority. When the scan control bit is a 1, a hard-coded scheme, shown in Table 3-1, is used.

Table 3-1. Priority Scheme

Processor Module	Section 0	Section 1	Section 2	Section 3
0	A	B	C	D
1	B	C	D	A
2	C	D	A	B
3	D	A	B	C

NOTE: A = Highest Priority
D = Lowest Priority

Memory Ports

In the CRAY EL series system, each processor module contains four memory ports for access to memory. Each of these ports has a specific function, which is defined jointly by the read mode bits and the port bits in the exchange package. Refer to Table 3-2. Ports A, B, and C are used for memory reference instructions (for details pertaining to instructions, refer to Appendix A). Port D is shared by the instruction buffers in the processor module and by the I/O section. All four ports are used for exchange operations and for refresh functions.

Table 3-2. Port and Read Mode Field Translation

Port Bits (P)								Read Mode Bits (RM)				Port Usage
61	60	61	60	61	60	61	60	59	58	57	56	
A		B		C		D						Exchange
0	0	0	1	1	0	1	1	0	0	0	1	
0	0	0	1	1	0	–	–	0	0	1	0	A or S
–	–	–	–	–	–	D		0	1	0	1	I/O single
						x	x					
–	–	–	–	–	–	D		0	1	1	1	I/O block
						x	x					
0	0	0	1	1	0	–	–	1	0	0	0	B or T
–	–	–	–	–	–	D		1	0	1	1	Fetch
						x	x					
0	0	0	1	1	0	–	–	1	1	0	0	Vector stride
0	0	0	1	1	0	–	–	1	1	1	0	Vector gather/scatter

Port D normally has a lower priority than ports A, B, and C when it is used for I/O transfers. However, if a memory reference is forced to wait from 0 to 255 clock periods during an I/O transfer, it is given highest priority for one memory reference. When the reference begins, port D is returned to lowest priority. The actual number of clock periods used as a hold figure can be determined using the scan chain capability of the CRAY EL series system.

All of the memory ports are bidirectional. The reservation system makes use of any memory port as it becomes available, but the ports are subject to a priority system. This priority system provides a fetch sequence, with the highest priority on port D; I/O operations are given the lowest port D priority.

Ports A, B, and C operate differently during block or vector transfers than when a scalar transfer is performed. When a port becomes available, the instruction issues and reserves the port. That port remains reserved until the instruction completes all of its memory references. After completion of the memory reference, the port reservation is cleared, and the port is available for the next instruction's reservation. Under normal circumstances, a block or vector transfer is capable of reading or writing 1 word of data each clock period. However, if the instruction encounters a memory conflict, it is suspended, or held, until the conflict is resolved. Because of this, the time needed to complete an instruction block or vector reference is unpredictable.

Block and vector transfers that use different ports are allowed to operate simultaneously, which may cause problems under certain circumstances. For example, if a 035*ijk* instruction precedes a 176*i*0*k* instruction that uses one or more of the same memory addresses, it is possible that some memory addresses could be read before they are written to. This is an out-of-sequence reference.

There are two ways to prevent out-of-sequence references. The first is to use the 002700 instruction, which forces completion of the memory reference. When inserted between the write instruction (035*ijk*) and the read instruction (176*i*0*k*), it forces the selected port to hold issue until the write instruction completes. Secondly, if the bidirectional mode (BDM) bit in the exchange package is clear (BDM = 0), out-of-sequence memory references are prevented. In this case, instructions using a port for a write operation prevent any other port from issuing a read instruction.

A scalar transfer instruction requires that ports A, B, and C be available before it can issue so that block transfers and scalar references within a processor module are sequential. A scalar reference uses one of the available ports, and consecutive scalar references are issued as long as a port is available. A scalar reference always completes in the order in which it was issued within the processor module.

The CRAY EL series system also has the added features of the concurrent block write (CBW) function and the scalar block overlap (SBO) function. The CBW and SBO bits are part of the exchange package. The SBO allows scalar and block references to intermix, and CBW allows more than one block write to occur at the same time, as long as different ports are used.

On port D, an instruction fetch sequence has priority over an I/O transfer. If a memory section conflict occurs, port D has priority over all other ports during a fetch operation.

Memory Conflicts

Several conditions can cause memory conflicts. For example, conflicts occur when a memory port attempts to access a portion of memory that is already busy, or when two or more ports try to access the same portion of memory at the same time. Intraprocessor module conflicts involve ports within the same processor module. Interprocessor module conflicts involve ports in different processor modules. In both cases, conflict resolution logic is used to impose a predetermined priority scheme to resolve the conflict.

There are four types of memory conflicts: path/bank busy, section, primary/secondary section and simultaneous bank. Each of these conflicts is resolved differently and the arbitration is done serially. Figure 3-6 shows the flow for the memory port arbitration.

First, bank busy conflicts and path conflicts are checked. If a port is held on a bank busy or a path conflict, it is not allowed to enter port section conflict or simultaneous bank conflict arbitration.

Secondly, a port section conflict occurs when more than one port in the same processor module attempts to access the same memory section simultaneously. Because each processor module has only one path to each memory section, simultaneous multiple references create conflicts. These conflicts are resolved by a priority system among the four ports. The port having the highest priority is granted access to memory first. All other ports that attempt a memory reference from that processor module at that time are subjected to a 1-clock-period hold issue condition. Port priority is determined using the following rules:

- Port D has priority over ports A, B, and C during a fetch sequence.
- If a conflict occurs between ports A, B, or C, any port with an odd address increment has priority over an even address increment. An address has an odd or an even increment, depending on the following conditions:
  - A port used for a block reference instruction has an address increment of 1, which is odd.
  - A port used by a stride reference instruction can have any constant increment (even or odd).
  - A port used by a gather/scatter instruction can have an increment that changes after each reference. For conflict resolution, a gather/scatter instruction always has an odd increment.
- Among ports A, B, and C, if all have the same type of memory increment, priority is determined by the relative time of instruction issue. The port used by the instruction issued first has the highest priority.
- Port D has the lowest priority when used for I/O transfers.

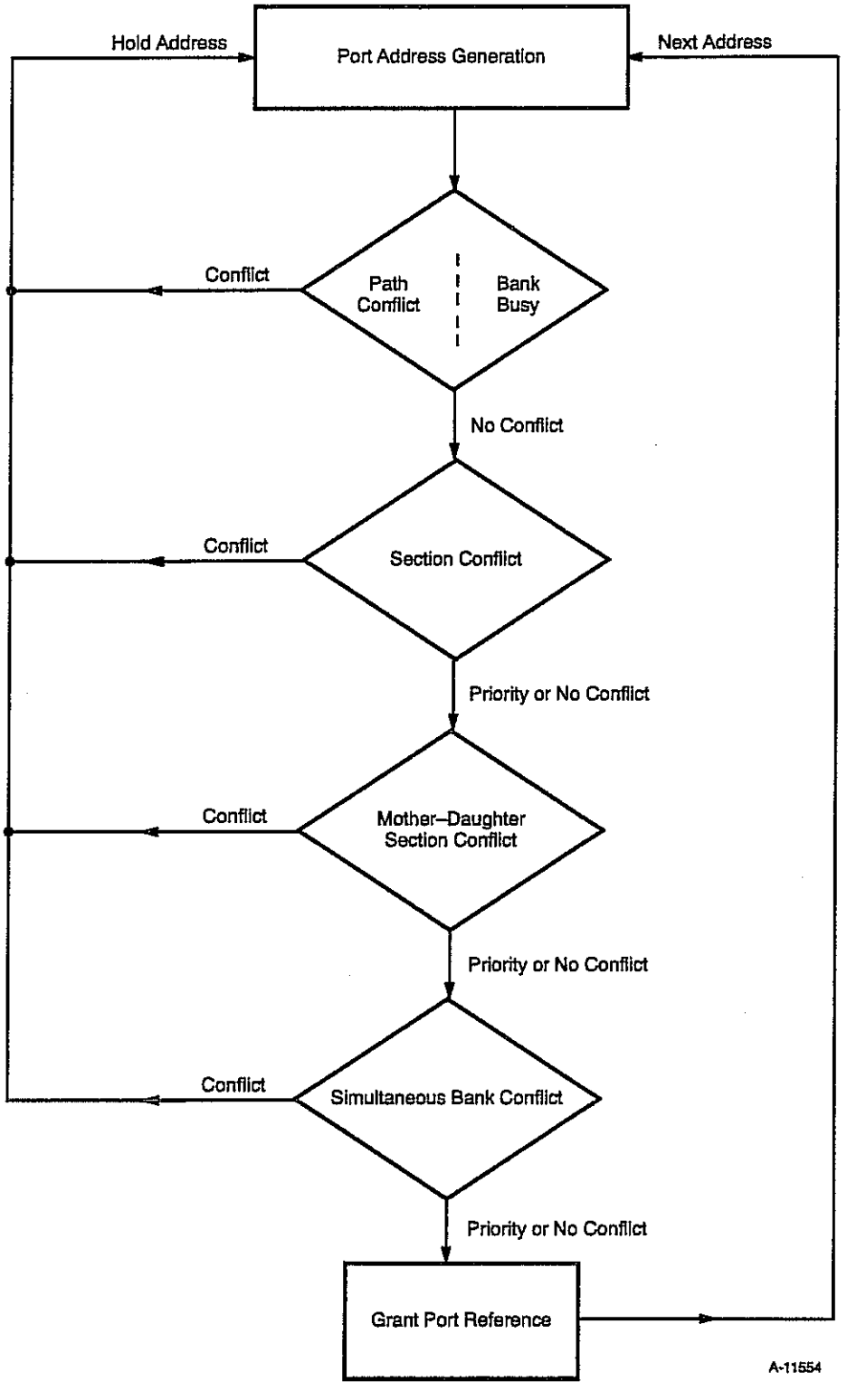

Figure 3-6. Flow for Memory Port Arbitration

The bank busy conflict occurs because each memory reference by a processor module holds that memory bank busy to all ports in all processor modules for 5 clock periods. If any port in another processor module tries to make a memory reference to the same bank during this 5 clock periods, a bank busy conflict occurs. The new memory reference holds for 1 to 4 clock periods until the old reference is no longer holding the bank busy. If other ports try to access a busy memory bank and are forced to hold, they experience a simultaneous bank conflict when the bank busy conflict is resolved.

Third, the section conflict between primary and secondary modules is resolved. This conflict is caused by the sharing of the memory address, commands, and data busses between primary and secondary modules. The priority for resolving a primary/secondary section conflict is based on a least recently used (LRU) scheme.

Finally, simultaneous bank conflicts are resolved. The AR10 ASIC contains two priority schemes for resolving bank conflicts. A scan activated control bit determines which scheme is used. When this scan bit is low, an LRU scheme is used to determine which processor module has highest priority. When the scan control bit is a 1, a hard-coded scheme identical to the one used in the CRAY Y-MP EL system is followed. Note that bank conflicts do not occur between a primary processor and its secondary processor. These conflicts are resolved by the primary/secondary section conflict. The LRU scheme was included to prevent a processor module from being locked out of a bank for an indefinite amount of time. The LRU scheme puts an upper limit on the amount of time a processor has to wait for access to a bank. The wait time is determined by which processor is used least, no matter which section of memory is accessed.

Memory Access Time

Access time is the time required for an instruction to transfer one or more operands from memory to an operating register. Access time depends on the type of register receiving the operand and the number of operands transferred. If no memory conflicts are encountered, the access times for each register are:

- 16 clock periods for A registers
- 16 clock periods for S registers
- 16 clock periods plus block length for B and T registers
- 16 clock periods plus vector length for vector register stride references

- 18 clock periods plus vector length for vector register gather references