2. PROGRAMMING INTERFACE

In this chapter, the internal operations of the IBM 6x86MX CPU are described mainly from an application programmer's point of view. Included in this chapter are descriptions of processor initialization, the register set, memory addressing, various types of interrupts and the shutdown and halt process. An overview of real, virtual 8086, and protected operating modes is also included in this chapter. The FPU operations are described separately at the end of the chapter.

This manual does not -and is not intended to- describe the IBM 6x86MX microprocessor or its operations at the circuit level.

2.1 Processor Initialization

The IBM 6x86MX CPU is initialized when the RESET signal is asserted. The processor is placed in real mode and the registers listed in Table 2-1 (Page 2-2) are set to their initialized values. RESET invalidates and disables the cache and turns off paging. When RESET is asserted, the IBM 6x86MX CPU terminates all local bus activity and all internal execution. During the entire time that RESET is asserted, the internal pipelines are flushed and no instruction execution or bus activity occurs.

Approximately 150 to 250 external clock cycles after RESET is negated, the processor begins executing instructions at the top of physical memory (address location FFFF FFF0h). Typically, an intersegment JUMP is placed at FFFF FFF0h. This instruction will force the processor to begin execution in the lowest 1 MByte of address space.

Note: The actual time depends on the clock scaling in use. Also an additional 2^{20} clock cycles are needed if self-test is requested.

Table 2-1. Initialized Register Controls

REGISTER	REGISTER NAME	INITIALIZED CONTENTS	COMMENTS
EAX	Accumulator	xxxx xxxxh	0000 0000h indicates self-test passed.
EBX	Base	xxxx xxxxh	
ECX	Count	xxxx xxxxh	
EDX	Data	06 + Device ID	Device ID = 51h or 59h (2X clock) Device ID = 55h or 5Ah (2.5X clock) Device ID = 53h or 5Bh (3X clock) Device ID = 54h or 5Ch (3.5X clock)
EBP	Base Pointer	xxxx xxxxh	
ESI	Source Index	xxxx xxxxh	
EDI	Destination Index	xxxx xxxxh	
ESP	Stack Pointer	xxxx xxxxh	
EFLAGS	Flag Word	0000 0002h	
EIP	Instruction Pointer	0000 FFF0h	
ES	Extra Segment	0000h	Base address set to 0000 0000h. Limit set to FFFFh.
CS	Code Segment	F000h	Base address set to FFFF 0000h. Limit set to FFFFh.
SS	Stack Segment	0000h	Base address set to 0000 0000h. Limit set to FFFFh.
DS	Data Segment	0000h	Base address set to 0000 0000h. Limit set to FFFFh.
FS	Extra Segment	0000h	Base address set to 0000 0000h. Limit set to FFFFh.
GS	Extra Segment	0000h	Base address set to 0000 0000h. Limit set to FFFFh.
IDTR	Interrupt Descriptor Table Register	Base = 0, Limit = 3FFh	
GDTR	Global Descriptor Table Register	xxxx xxxxh, xxxxh	
LDTR	Local Descriptor Table Register	xxxx xxxxh, xxxxh	
TR	Task Register	xxxxh	
CR0	Machine Status Word	6000 0010h	
CR2	Control Register 2	xxxx xxxxh	
CR3	Control Register 3	xxxx xxxxh	
CR4	Control Register 4	0000 0000h	
CCR (0-6)	Configuration Control (0-6)	00h CCR(0-3, 5-6) 80h CCR4	
ARR (0-7)	Address Region Registers (0-7)	00h	
RCR (0-7)	Region Control Registers (0-7)	00h	
DR7	Debug Register 7	0000 0400h	

Note: x = Undefined value

2.2 Instruction Set Overview

The IBM 6x86MX CPU instruction set performs ten types of general operations:

- Arithmetic
- Bit Manipulation
- Control Transfer
- Data Transfer
- Floating Point
- High-Level Language Support
- Operating System Support
- Shift/Rotate
- String Manipulation
- MMX Instructions

All IBM 6x86MX CPU instructions operate on as few as zero operands and as many as three operands. An NOP instruction (no operation) is an example of a zero operand instruction. Two operand instructions allow the specification of an explicit source and destination pair as part of the instruction. These two operand instructions can be divided into eight groups according to operand types:

- Register to Register
- Register to Memory
- Memory to Register
- Memory to Memory
- Register to I/O
- I/O to Register
- Immediate Data to Register
- Immediate Data to Memory

An operand can be held in the instruction itself (as in the case of an immediate operand), in one of the processor's registers or I/O ports, or in memory. An immediate operand is prefetched as part of the opcode for the instruction.

Operand lengths of 8, 16, or 32 bits are supported as well as 64-or 80-bit associated with floating point instructions. Operand lengths of 8 or 32 bits are generally used when executing code written for 386- or 486-class (32-bit code) processors. Operand lengths of 8 or 16 bits are generally used when executing existing 8086 or 80286 code (16-bit code). The default length of an operand can be overridden by placing one or more instruction prefixes in front of the opcode. For example, by using prefixes, a 32-bit operand can be used with 16-bit code, or a 16-bit operand can be used with 32-bit code.

Chapter 6 of this manual lists each instruction in the IBM 6x86MX CPU instruction set along with the associated opcodes, execution clock counts, and effects on the FLAGS register.

2.2.1 Lock Prefix

The LOCK prefix may be placed before certain instructions that read, modify, then write back to memory. The prefix asserts the LOCK# signal to indicate to the external hardware that the CPU is in the process of running multiple indivisible memory accesses. The LOCK prefix can be used with the following instructions:

Bit Test Instructions (BTS, BTR, BTC)
Exchange Instructions (XADD, XCHG, CMPXCHG)
One-operand Arithmetic and Logical Instructions (DEC, INC, NEG, NOT)
Two-operand Arithmetic and Logical Instructions (ADC, ADD, AND, OR, SBB, SUB, XOR).

An invalid opcode exception is generated if the LOCK prefix is used with any other instruction, or with the above instructions when no write operation to memory occurs (i.e., the destination is a register). The LOCK# signal can be negated to allow weak-locking for all of memory or on a regional basis. Refer to the descriptions of the NO-LOCK bit (within CCR1) and the WL bit (within RCRx) later in this chapter.

2.3 Register Sets

From the programmer's point of view there are 58 accessible registers in the IBM 6x86MX CPU. These registers are grouped into two sets. The application register set contains the registers frequently used by application programmers, and the system register set contains the registers typically reserved for use by operating system programmers.

The application register set is made up of general purpose registers, segment registers, a flag register, and an instruction pointer register.

The system register set is made up of the remaining registers which include control registers, system address registers, debug registers, configuration registers, and test registers.

Each of the registers is discussed in detail in the following sections.

2.3.1 Application Register Set

The application register set, (Figure 2-1, Page 2-5) consists of the registers most often used by the applications programmer. These registers are generally accessible and are not protected from read or write access.

The **General Purpose Register** contents are frequently modified by assembly language instructions and typically contain arithmetic and logical instruction operands.

Segment Registers in real mode contain the base address for each segment. In protected mode the segment registers contain segment selectors. The segment selectors provide indexing for tables (located in memory) that contain the base address and limit for each segment, as well as access control information.

The **Flag Register** contains control bits used to reflect the status of previously executed instructions. This register also contains control bits that affect the operation of some instructions.

The **Instruction Pointer** register points to the next instruction that the processor will execute. This register is automatically incremented by the processor as execution progresses.

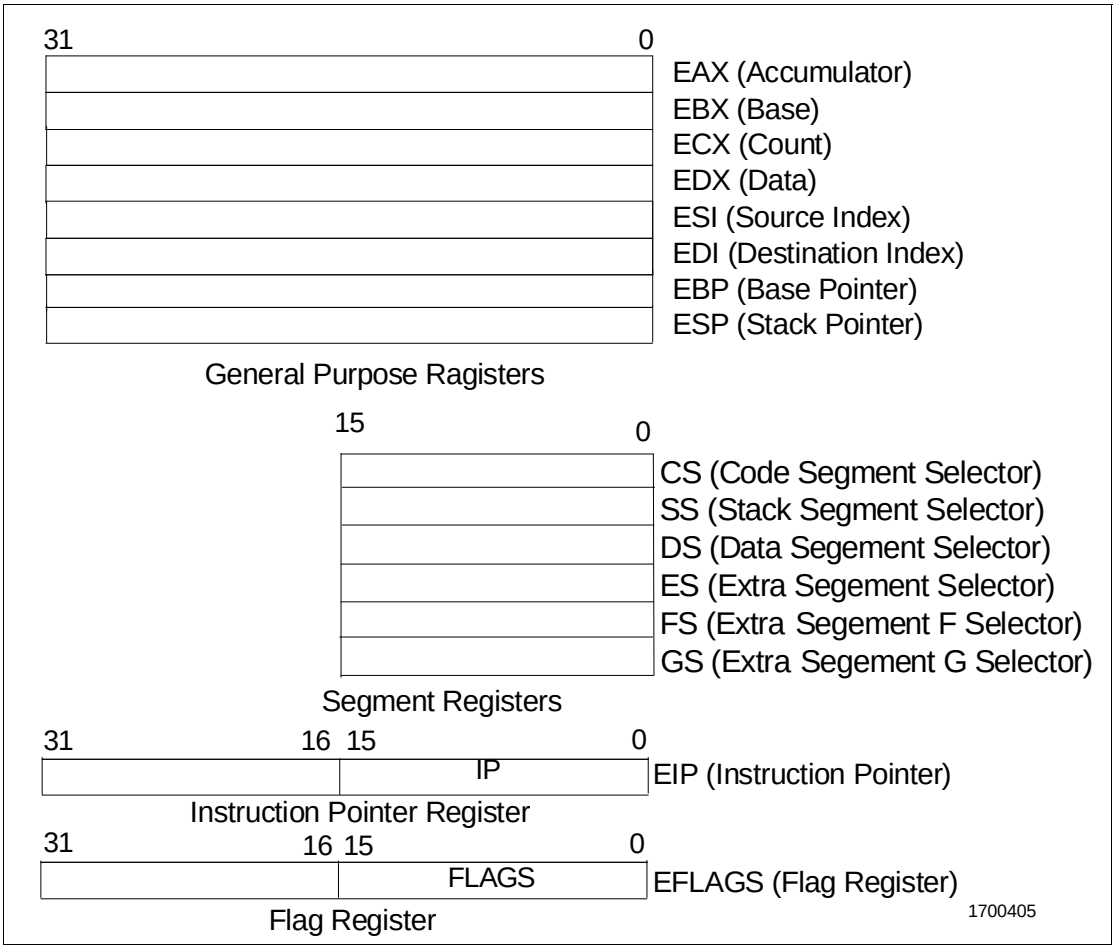


Figure 2-1. Application Register Set

2.3.2 General Purpose Registers

The general purpose registers are divided into four data registers, two pointer registers, and two index registers as shown in Figure 2-2 (Page 2-6).

The **Data Registers** are used by the applications programmer to manipulate data structures and to hold the results of logical and arithmetic operations. Different portions of the general data registers can be addressed by using different names.

An “E” prefix identifies the complete 32-bit register. An “X” suffix without the “E” prefix identifies the lower 16 bits of the register.

The lower two bytes of a data register can be addressed with an “H” suffix (identifies the upper byte) or an “L” suffix (identifies the lower byte). The _L and _H portions of a data registers act as independent registers. For example, if the AH register is written to by an instruction, the AL register bits remain unchanged.

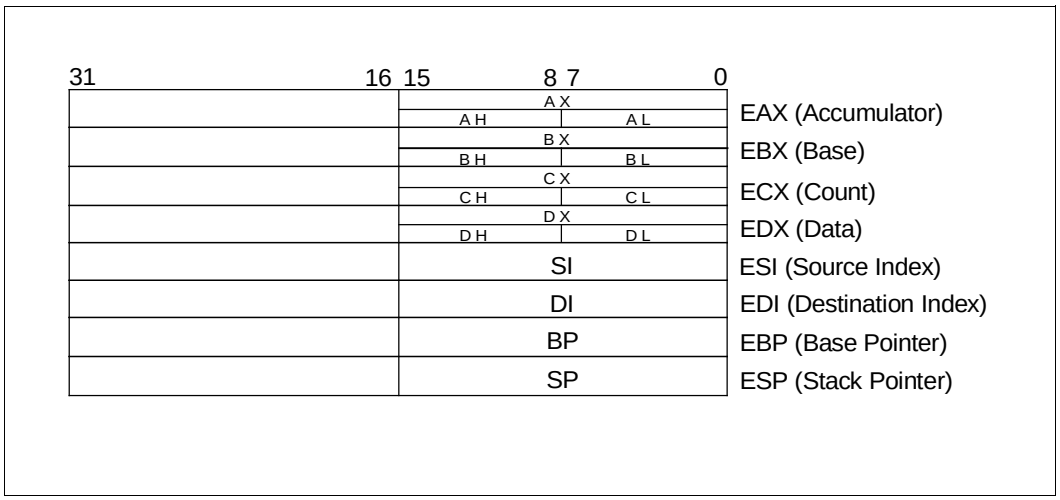


Figure 2-2. General Purpose Registers

The **Pointer and Index Registers** are listed below.

SI or ESI	Source Index
DI or EDI	Destination Index
SP or ESP	Stack Pointer
BP or EBP	Base Pointer

These registers can be addressed as 16- or 32-bit registers, with the “E” prefix indicating 32 bits. The pointer and index registers can be used as general purpose registers, however, some instructions use a fixed assignment of these registers. For example, repeated string operations always use ESI as the source pointer, EDI as the destination pointer, and ECX as the counter. The instructions using fixed registers include multiply and divide, I/O access, string operations, translate, loop, variable shift and rotate, and stack operations.

The IBM 6x86MX CPU processor implements a stack using the ESP register. This stack is accessed during the PUSH and POP instructions, procedure calls, procedure returns, interrupts, exceptions, and interrupt/exception returns.

The microprocessor automatically adjusts the value of the ESP during operation of these instructions.The EBP register may be used to reference data passed on the stack during procedure calls. Local data may also be placed on the stack and referenced relative to BP. This register provides a mechanism to access stack data in high-level languages.

2.3.3 Segment Registers and Selectors

Segmentation provides a means of defining data structures inside the memory space of the microprocessor. There are three basic types of segments: code, data, and stack. Segments are used automatically by the processor to determine the location in memory of code, data, and stack references.

There are six 16-bit segment registers:

CS	Code Segment
DS	Data Segment
ES	Extra Segment
SS	Stack Segment
FS	Additional Data Segment
GS	Additional Data Segment.

In real and virtual 8086 operating modes, a segment register holds a 16-bit segment base. The 16-bit segment is multiplied by 16 and a 16-bit or 32-bit offset is then added to it to create a linear address. The offset size is dependent on the current address size. In real mode and in virtual 8086 mode with paging disabled, the linear address is also the physical address. In virtual 8086 mode with paging enabled, the linear address is translated to the physical address using the current page tables. Paging is described in Section 2.12.4 (Page 2-52).

In protected mode a segment register holds a **Segment Selector** containing a 13-bit index, a Table Indicator (TI) bit, and a two-bit Requested Privilege Level (RPL) field as shown in Figure 2-3.

The **Index** points into a descriptor table in memory and selects one of 8192 (2^{13}) segment descriptors contained in the descriptor table.

A segment descriptor is an eight-byte value used to describe a memory segment by defining the segment base, the segment limit, and access control information. To address data within a segment, a 16-bit or 32-bit offset is added to the segment's base address. Once a segment selector has been loaded into a segment register, an instruction needs only to specify the segment register and the offset.

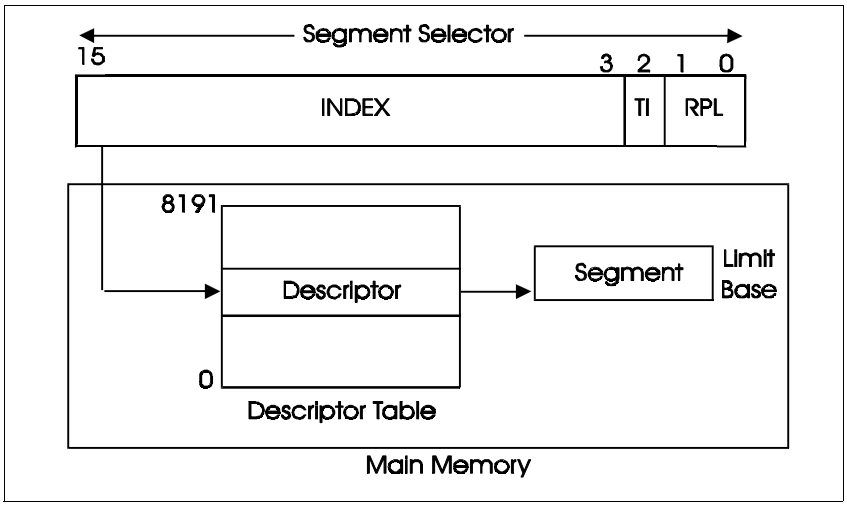


Figure 2-3. Segment Selector in Protected Mode

The **Table Indicator** (TI) bit of the selector defines which descriptor table the index points into. If TI=0, the index references the Global Descriptor Table (GDT). If TI=1, the index references the Local Descriptor Table (LDT). The GDT and LDT are described in more detail in Section 2.4.2 (Page 2-16). Protected mode addressing is discussed further in Sections 2.6.2 (Page 2-52).

The **Requested Privilege Level** (RPL) field in a segment selector is used to determine the Effective Privilege Level of an instruction (where RPL=0 indicates the most privileged level, and RPL=3 indicates the least privileged level).

If the level requested by RPL is less than the Current Program Level (CPL), the RPL level is accepted and the Effective Privilege Level is changed to the RPL value. If the level requested by RPL is greater than CPL, the CPL overrides the requested RPL and Effective Privilege Level remains unchanged.

When a segment register is loaded with a segment selector, the segment base, segment limit and access rights are loaded from the descriptor table entry into a user-invisible or hidden portion of the segment register (i.e., cached on-chip). The CPU does not access the descriptor table entry again until another segment register load occurs. If the descriptor tables are modified in memory, the segment registers must be reloaded with the new selector values by the software.

The processor automatically selects an implied (default) segment register for memory references. Table 2-2 describes the selection rules. In general, data references use the selector contained in the DS register, stack references use the SS register and instruction fetches use the CS register. While some of these selections may be overridden, instruction fetches, stack operations, and the destination write of string operations cannot be overridden. Special segment override instruction prefixes allow the use of alternate segment registers including the use of the ES, FS, and GS segment registers.

Table 2-2. Segment Register Selection Rules

TYPE OF MEMORY REFERENCE	IMPLIED (DEFAULT) SEGMENT	SEGMENT OVERRIDE PREFIX
Code Fetch	CS	None
Destination of PUSH, PUSHF, INT, CALL, PUSHA instructions	SS	None
Source of POP, POPA, POPF, IRET, RET instructions	SS	None
Destination of STOS, MOVS, REP STOS, REP MOVS instructions	ES	None
Other data references with effective address using base registers of:		
EAX, EBX, ECX, EDX, ESI, EDI	DS	CS, ES, FS, GS, SS
EBP, ESP	SS	CS, DS, ES, FS, GS

2.3.4 Instruction Pointer Register

The **Instruction Pointer** (EIP) register contains the offset into the current code segment of the next instruction to be executed. The register is normally incremented with each instruction execution unless implicitly modified through an interrupt, exception or an instruction that changes the sequential execution flow (e.g., JMP, CALL).

2.3.5 Flags Register

The **Flags Register**, EFLAGS, contains status information and controls certain operations on the IBM 6x86MX CPU microprocessor. The lower 16 bits of this register are referred to as the FLAGS register that is used when executing 8086 or 80286 code. The flag bits are shown in Figure 2-4 and defined in Table 2-3 (Page 2-10).

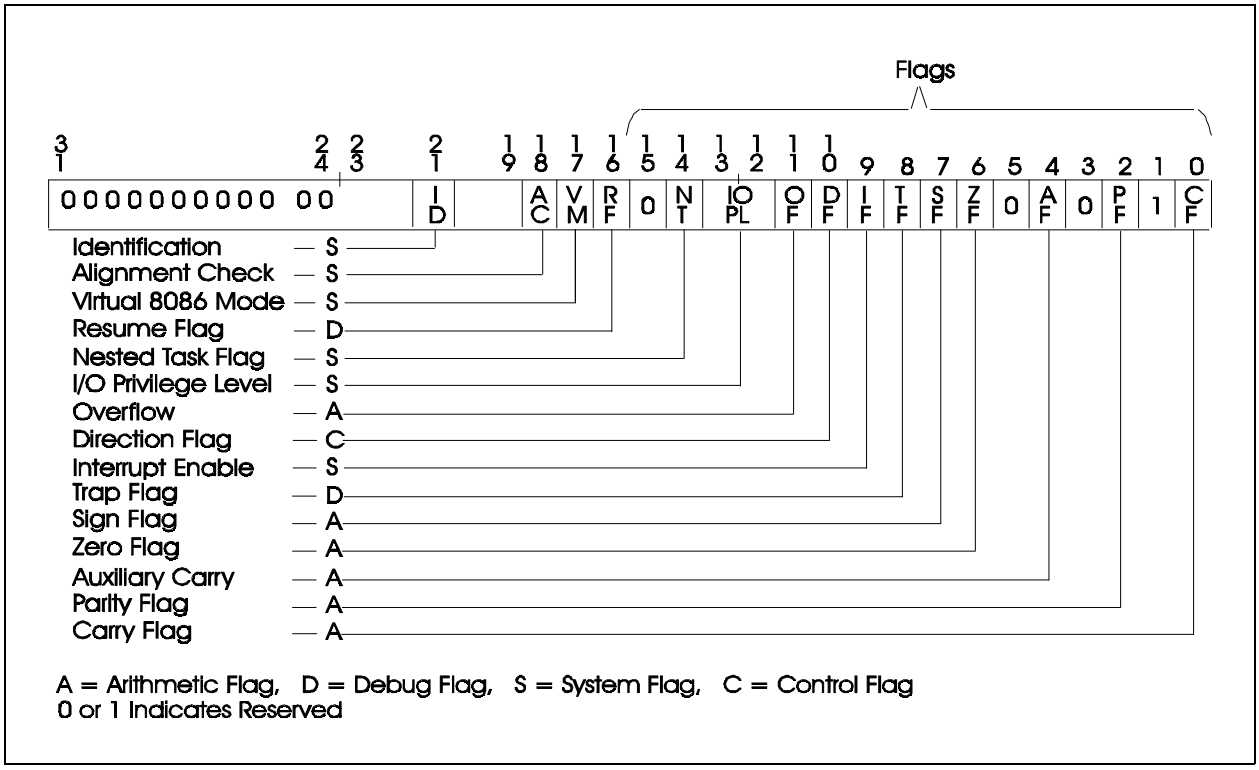


Figure 2-4. EFLAGS Register

Table 2-3. EFLAGS Bit Definitions

BIT POSITION	NAME	FUNCTION
0	CF	Carry Flag: Set when a carry out of (addition) or borrow into (subtraction) the most significant bit of the result occurs; cleared otherwise.
2	PF	Parity Flag: Set when the low-order 8 bits of the result contain an even number of ones; cleared otherwise.
4	AF	Auxiliary Carry Flag: Set when a carry out of (addition) or borrow into (subtraction) bit position 3 of the result occurs; cleared otherwise.
6	ZF	Zero Flag: Set if result is zero; cleared otherwise.
7	SF	Sign Flag: Set equal to high-order bit of result (0 indicates positive, 1 indicates negative).
8	TF	Trap Enable Flag: Once set, a single-step interrupt occurs after the next instruction completes execution. TF is cleared by the single-step interrupt.
9	IF	Interrupt Enable Flag: When set, maskable interrupts (INTR input pin) are acknowledged and serviced by the CPU.
10	DF	Direction Flag: If DF=0, string instructions auto-*in*crement (default) the appropriate index registers (ESI and/or EDI). If DF=1, string instructions auto-*de*crement the appropriate index registers.
11	OF	Overflow Flag: Set if the operation resulted in a carry or borrow into the sign bit of the result but did not result in a carry or borrow out of the high-order bit. Also set if the operation resulted in a carry or borrow out of the high-order bit but did not result in a carry or borrow into the sign bit of the result.
12, 13	IOPL	I/O Privilege Level: While executing in protected mode, IOPL indicates the maximum current privilege level (CPL) permitted to execute I/O instructions without generating an exception 13 fault or consulting the I/O permission bit map. IOPL also indicates the maximum CPL allowing alteration of the IF bit when new values are popped into the EFLAGS register.
14	NT	Nested Task: While executing in protected mode, NT indicates that the execution of the current task is nested within another task.
16	RF	Resume Flag: Used in conjunction with debug register breakpoints. RF is checked at instruction boundaries before breakpoint exception processing. If set, any debug fault is ignored on the next instruction.
17	VM	Virtual 8086 Mode: If set while in protected mode, the microprocessor switches to virtual 8086 operation handling segment loads as the 8086 does, but generating exception 13 faults on privileged opcodes. The VM bit can be set by the IRET instruction (if current privilege level=0) or by task switches at any privilege level.
18	AC	Alignment Check Enable: In conjunction with the AM flag in CR0, the AC flag determines whether or not misaligned accesses to memory cause a fault. If AC is set, alignment faults are enabled.
21	ID	Identification Bit: The ability to set and clear this bit indicates that the CPUID instruction is supported. The ID can be modified only if the CPUID bit in CCR4 is set.

2.4 System Register Set

The system register set, shown in Figure 2-5 (Page 2-12), consists of registers not generally used by application programmers. These registers are typically employed by system level programmers who generate operating systems and memory management programs.

The **Control Registers** control certain aspects of the IBM 6x86MX microprocessor such as paging, coprocessor functions, and segment protection. When a paging exception occurs while paging is enabled, some control registers retain the linear address of the access that caused the exception.

The **Descriptor Table Registers** and the **Task Register** can also be referred to as system address or memory management registers. These registers consist of two 48-bit and two 16-bit registers. These registers specify the location of the data structures that control the segmentation used by the IBM 6x86MX microprocessor. Segmentation is one available method of memory management.

The **Configuration Registers** are used to configure the IBM 6x86MX CPU on-chip cache operation, power management features and System Management Mode. The configuration registers also provide information on the CPU device type and revision.

The **Debug Registers** provide debugging facilities to enable the use of data access breakpoints and code execution breakpoints.

The **Test Registers** provide a mechanism to test the contents of both the on-chip 16 KByte cache and the Translation Lookaside Buffer (TLB). In the following sections, the system register set is described in greater detail.

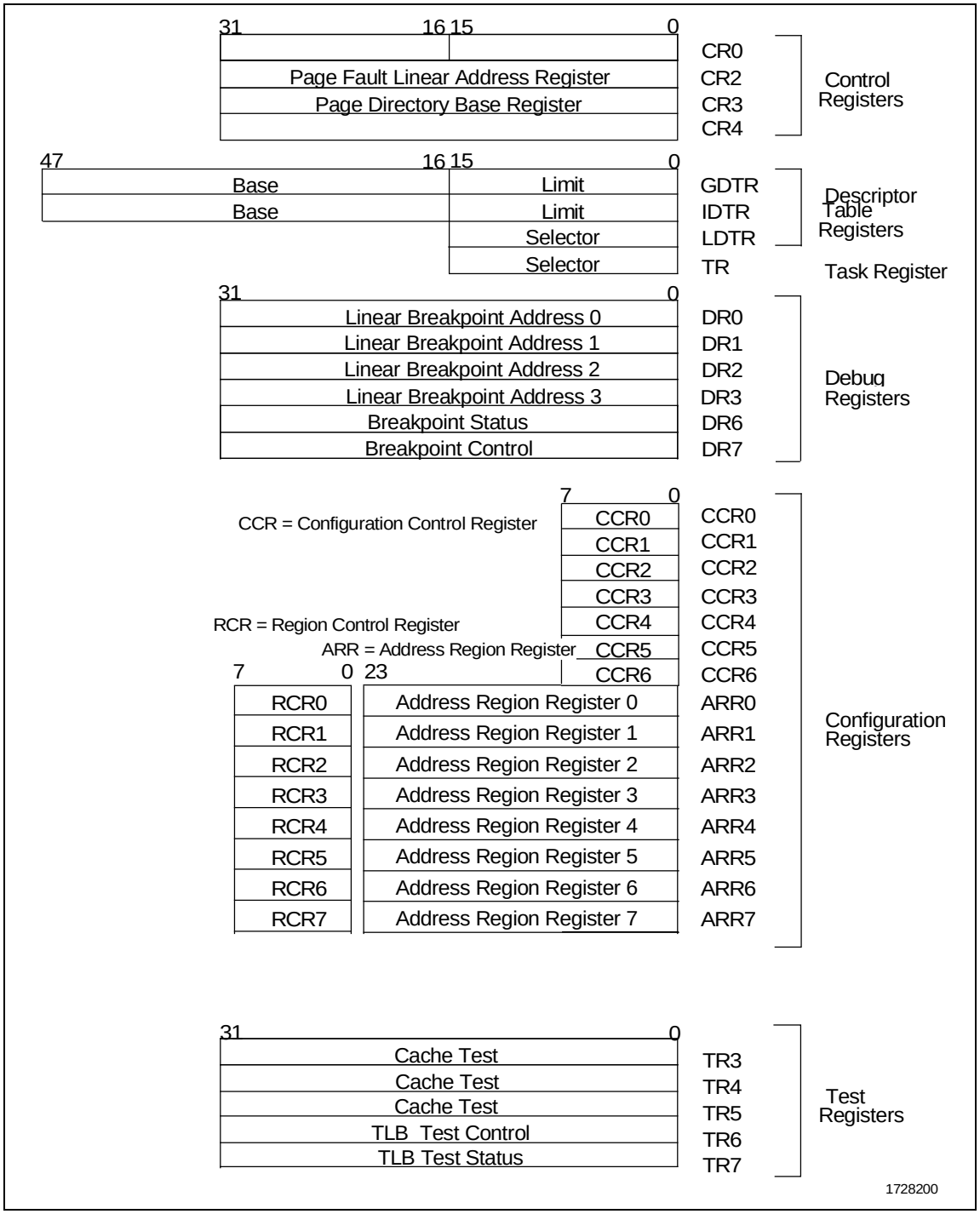


Figure 2-5. System Register Set

2.4.1 Control Registers

The Control Registers (CR0, CR2, CR3 and CR4), are shown in Figure 2-6. (These registers should not be confused with the CCRn registers.) The CR0 register contains system control bits which configure operating modes and indicate the general state of the CPU. The lower 16 bits of CR0 are referred to as the Machine Status Word (MSW). The CR0 bit definitions are described in Table 2-4 and Table 2-5 (Page 2-14). The reserved bits in CR0 should not be modified.

When paging is enabled and a page fault is generated, the CR2 register retains the 32-bit linear address of the address that caused the fault. When a double page fault occurs, CR2 contains the address for the second fault. Register CR3 contains the 20 most significant bits of the physical base address of the page directory. The page directory must always be aligned to a 4-KByte page boundary, therefore, the lower 12 bits of CR3 are not required to specify the base address.

CR3 contains the Page Cache Disable (PCD) and Page Write Through (PWT) bits. During bus cycles that are not paged, the state of the PCD bit is reflected on the PCD pin and the PWT bit is driven on the PWT pin. These bus cycles include interrupt acknowledge cycles and all bus cycles, when paging is not enabled. The PCD pin should be used to control caching in an external cache. The PWT pin should be used to control write policy in an external cache.

Control register CR4 (Table 2-6, Page 2-15) controls usage of the Time Stamp Counter Instruction, Debugging Extensions, Page Global Enable and the RDPMC instruction.

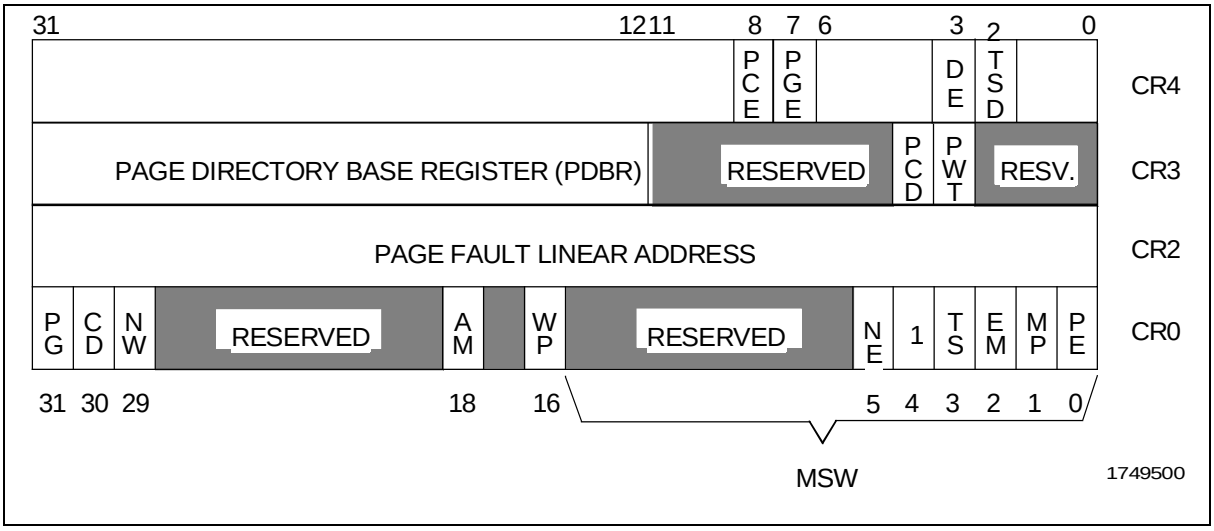


Figure 2-6. Control Registers

Table 2-4. CR0 Bit Definitions

BIT POSITION	NAME	FUNCTION
0	PE	Protected Mode Enable: Enables the segment based protection mechanism. If PE=1, protected mode is enabled. If PE=0, the CPU operates in real mode and addresses are formed as in an 8086-style CPU.
1	MP	Monitor Processor Extension: If MP=1 and TS=1, a WAIT instruction causes Device Not Available (DNA) fault 7. The TS bit is set to 1 on task switches by the CPU. Floating point instructions are not affected by the state of the MP bit. The MP bit should be set to one during normal operations.
2	EM	Emulate Processor Extension: If EM=1, all floating point instructions cause a DNA fault 7.
3	TS	Task Switched: Set whenever a task switch operation is performed. Execution of a floating point instruction with TS=1 causes a DNA fault. If MP=1 and TS=1, a WAIT instruction also causes a DNA fault.
4	1	Reserved: Do not attempt to modify.
5	NE	Numerics Exception. NE=1 to allow FPU exceptions to be handled by interrupt 16. NE=0 if FPU exceptions are to be handled by external interrupts.
16	WP	Write Protect: Protects read-only pages from supervisor write access. WP=0 allows a read-only page to be written from privilege level 0-2. WP=1 forces a fault on a write to a read-only page from any privilege level.
18	AM	Alignment Check Mask: If AM=1, the AC bit in the EFLAGS register is unmasked and allowed to enable alignment check faults. Setting AM=0 prevents AC faults from occurring.
29	NW	Not Write-Back: If NW=1, the on-chip cache operates in write-through mode. In write-through mode, all writes (including cache hits) are issued to the external bus. If NW=0, the on-chip cache operates in write-back mode. In write-back mode, writes are issued to the external bus only for a cache miss, a line replacement of a modified line, or as the result of a cache inquiry cycle.
30	CD	Cache Disable: If CD=1, no further cache line fills occur. However, data already present in the cache continues to be used if the requested address hits in the cache. Writes continue to update the cache and cache invalidations due to inquiry cycles occur normally. The cache must also be invalidated to completely disable any cache activity.
31	PG	Paging Enable Bit: If PG=1 and protected mode is enabled (PE=1), paging is enabled. After changing the state of PG, software must execute an unconditional branch instruction (e.g., JMP, CALL) to have the change take effect.

Table 2-5. Effects of Various Combinations of EM, TS, and MP Bits

CR0 BIT			INSTRUCTION TYPE	
EM	TS	MP	WAIT	ESC
0	0	0	Execute	Execute
0	0	1	Execute	Execute
0	1	0	Execute	Fault 7
0	1	1	Fault 7	Fault 7
1	0	0	Execute	Fault 7
1	0	1	Execute	Fault 7
1	1	0	Execute	Fault 7
1	1	1	Fault 7	Fault 7

Table 2-6. CR4 Bit Definitions

BIT POSITION	NAME	FUNCTION
2	TSD	Time Stamp Counter Instruction If = 1 RDTSC instruction enabled for CPL=0 only; Reset State If = 0 RDTSC instruction enabled for all CPL states
3	DE	Debugging Extensions If = 1 enables I/O breakpoints and R/W bits for each debug register are defined as: 00 -Break on instruction execution only. 01 -Break on data writes only. 10 -Break on I/O reads or writes. 11 -Break on data reads or writes but not instruction fetches. If = 0 I/O breakpoints and R/W bits for each debug register are not enabled.
7	PGE	Page Global Enable If = 1 global page feature is enabled. If = 0 global page feature is disabled. Global pages are not flushed from TLB on a task switch or write to CR3
8	PCE	Performance Monitoring Counter Enable If = 1 enables execution of RDPMC instruction at any protection level. If = 0 RDPMC instruction can only be executed at protection level 0.

2.4.2 Descriptor Table Registers and Descriptors

Descriptor Table Registers

The Global, Interrupt, and Local Descriptor Table Registers (GDTR, IDTR and LDTR), shown in Figure 2-7, are used to specify the location of the data structures that control segmented memory management. The GDTR, IDTR and LDTR are loaded using the LGDT, LIDT and LLDT instructions, respectively. The values of these registers are stored using the corresponding store instructions. The GDTR and IDTR load instructions are privileged instructions when operating in protected mode. The LDTR can only be accessed in protected mode.

The **Global Descriptor Table Register** (GDTR) holds a 32-bit linear base address and 16-bit limit for the Global Descriptor Table (GDT). The GDT is an array of up to 8192 8-byte descriptors. When a segment register is loaded from memory, the TI bit in the segment selector chooses either the GDT or the Local Descriptor Table (LDT) to locate a descriptor. If TI = 0, the index portion of the selector is used to locate the descriptor within the GDT table. The contents of the GDTR are completely visible to the programmer by using a SGDT instruction. The first descriptor in the GDT (location 0) is not used by the CPU and is referred to as the “null descriptor”. The GDTR is initialized using a LGDT instruction.

The **Interrupt Descriptor Table Register** (IDTR) holds a 32-bit linear base address and 16-bit limit for the Interrupt Descriptor Table (IDT). The IDT is an array of 256 interrupt descriptors, each of which is used to point to an interrupt service routine. Every interrupt that may occur in the system must have an associated entry in the IDT. The contents of the IDTR are completely visible to the programmer by using a SIDT instruction. The IDTR is initialized using the LIDT instruction.

The **Local Descriptor Table Register** (LDTR) holds a 16-bit selector for the Local Descriptor Table (LDT). The LDT is an array of up to 8192 8-byte descriptors. When the LDTR is loaded, the LDTR selector indexes an LDT descriptor that resides in the Global Descriptor Table (GDT). The base address and limit are loaded automatically and cached from the LDT descriptor within the GDT.

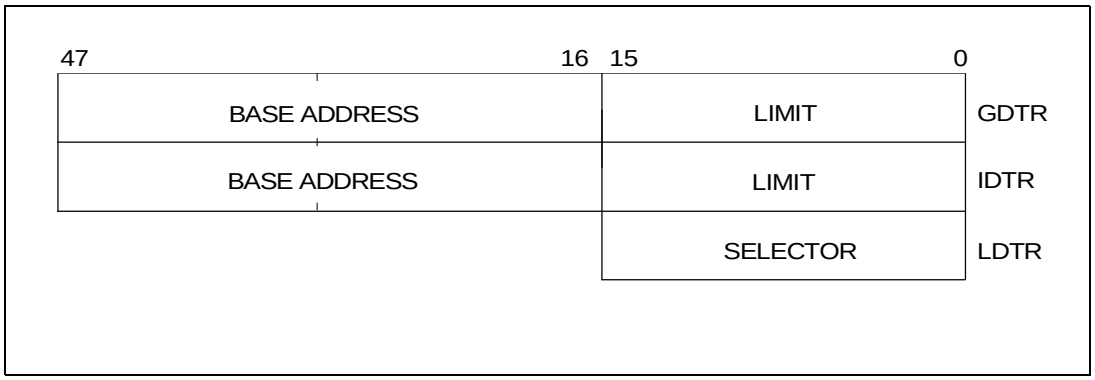


Figure 2-7. Descriptor Table Registers

Subsequent access to entries in the LDT use the hidden LDTR cache to obtain linear addresses. If the LDT descriptor is modified in the GDT, the LDTR must be reloaded to update the hidden portion of the LDTR.

When a segment register is loaded from memory, the TI bit in the segment selector chooses either the GDT or the LDT to locate a segment descriptor. If TI = 1, the index portion of the selector is used to locate a given descriptor within the LDT. Each task in the system may be given its own LDT, managed by the operating system. The LDTs provide a method of isolating a given task's segments from other tasks in the system.

The LDTR can be read or written by the LLDT and SLDT instructions.

Descriptors

There are three types of descriptors:

- Application Segment Descriptors that define code, data and stack segments.
- System Segment Descriptors that define an LDT segment or a Task State Segment (TSS) table described later in this text.
- Gate Descriptors that define task gates, interrupt gates, trap gates and call gates.

Application Segment Descriptors can be located in either the LDT or GDT. System Segment Descriptors can only be located in the GDT. Dependent on the gate type, gate descriptors may be located in either the GDT, LDT or IDT. Figure 2-8 illustrates the descriptor format for both Application Segment Descriptors and System Segment Descriptors. Table 2-7 (Page 2-18) lists the corresponding bit definitions.

Table 2-8. (Page 2-18) and Table 2-9. (Page 2-19) defines the DT field within the segment descriptor.

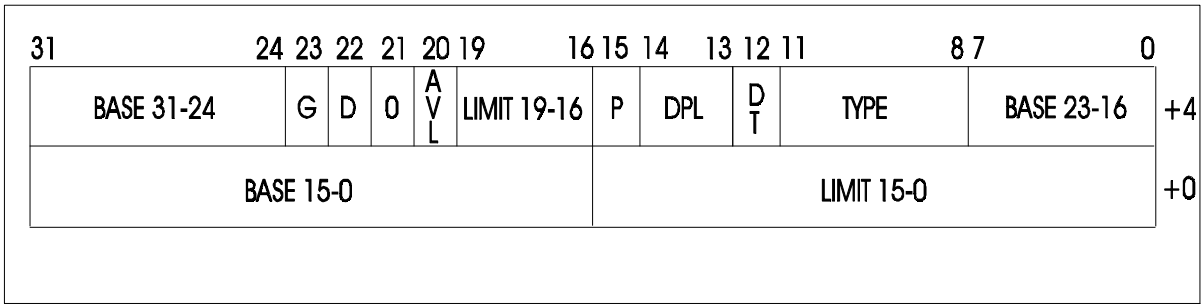


Figure 2-8. Application and System Segment Descriptors

Table 2-7. Segment Descriptor Bit Definitions

BIT POSITION	MEMORY OFFSET	NAME	DESCRIPTION
31-24 7-0 31-16	+4 +4 +0	BASE	Segment base address. 32-bit linear address that points to the beginning of the segment.
19-16 15-0	+4 +0	LIMIT	Segment limit.
23	+4	G	Limit granularity bit: 0 = byte granularity, 1 = 4 KBytes (page) granularity.
22	+4	D	Default length for operands and effective addresses. Valid for code and stack segments only: 0 = 16 bit, 1 = 32-bit.
20	+4	AVL	Segment available.
15	+4	P	Segment present.
14-13	+4	DPL	Descriptor privilege level.
12	+4	DT	Descriptor type: 0 = system, 1 = application.
11-8	+4	TYPE	Segment type. See Tables 2-7 and 2-8.

Table 2-8. TYPE Field Definitions with DT = 0

TYPE (BITS 11-8)	DESCRIPTION
0001	TSS-16 descriptor, task not busy.
0010	LDT descriptor.
0011	TSS-16 descriptor, task busy.
1001	TSS-32 descriptor, task not busy
1011	TSS-32 descriptor, task busy.

Table 2-9. TYPE Field Definitions with DT = 1

TYPE				APPLICATION DECRIPTOR INFORMATION
E	C/D	R/W	A	
0	0	x	x	data, expand up, limit is upper bound of segment
0	1	x	x	data, expand down, limit is lower bound of segment
1	0	x	x	executable, non-conforming
1	1	x	x	executable, conforming (runs at privilege level of calling procedure)
0	x	0	x	data, non-writable
0	x	1	x	data, writable
1	x	0	x	executable, non-readable
1	x	1	x	executable, readable
x	x	x	0	not-accessed
x	x	x	1	accessed

Gate Descriptors provide protection for executable segments operating at different privilege levels. Figure 2-9 illustrates the format for Gate Descriptors and Table 2-10 lists the corresponding bit definitions.

Task Gate Descriptors are used to switch the CPU's context during a task switch. The selector portion of the task gate descriptor locates a Task State Segment. These descriptors can be located in the GDT, LDT or IDT tables.

Interrupt Gate Descriptors are used to enter a hardware interrupt service routine. Trap Gate Descriptors are used to enter exceptions or software interrupt service routines. Trap Gate and Interrupt Gate Descriptors can only be located in the IDT.

Call Gate Descriptors are used to enter a procedure (subroutine) that executes at the same or a more privileged level. A Call Gate Descriptor primarily defines the procedure entry point and the procedure's privilege level.

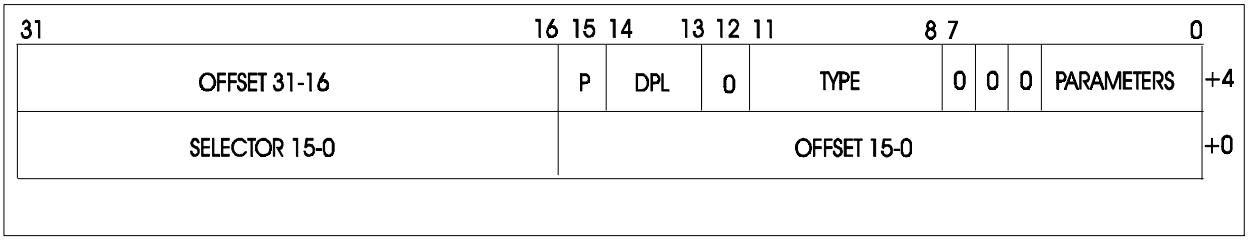


Figure 2-9. Gate Descriptor

Table 2-10. Gate Descriptor Bit Definitions

BIT POSITION	MEMORY OFFSET	NAME	DESCRIPTION
31-16 15-0	+4 +0	OFFSET	Offset used during a call gate to calculate the branch target.
31-16	+0	SELECTOR	Segment selector used during a call gate to calculate the branch target.
15	+4	P	Segment present.
14-13	+4	DPL	Descriptor privilege level.
11-8	+4	TYPE	Segment type: 0100 = 16-bit call gate 0101 = task gate 0110 = 16-bit interrupt gate 0111 = 16-bit trap gate 1100 = 32-bit call gate 1110 = 32-bit interrupt gate 1111 = 32-bit trap gate.
4-0	+4	PARAME-TERS	Number of 32-bit parameters to copy from the caller's stack to the called procedure's stack (valid for calls).

2.4.3 Task Register

The **Task Register** (TR) holds a 16-bit selector for the current Task State Segment (TSS) table as shown in Figure 2-10. The TR is loaded and stored via the LTR and STR instructions, respectively. The TR can only be accessed during protected mode and can only be loaded when the privilege level is 0 (most privileged). When the TR is loaded, the TR selector field indexes a TSS descriptor that must reside in the Global Descriptor Table (GDT). The contents of the selected descriptor are cached on-chip in the hidden portion of the TR.

During task switching, the processor saves the current CPU state in the TSS before starting a new task. The TR points to the current TSS. The TSS can be either a 386/486-style 32-bit TSS (Figure 2-11, Page 2-22) or a 286-style 16-bit TSS type (Figure 2-12, Page 2-23). An I/O permission bit map is referenced in the 32-bit TSS by the I/O Map Base Address.

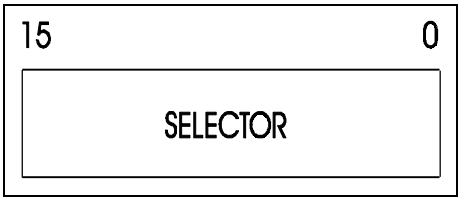


Figure 2-10. Task Register

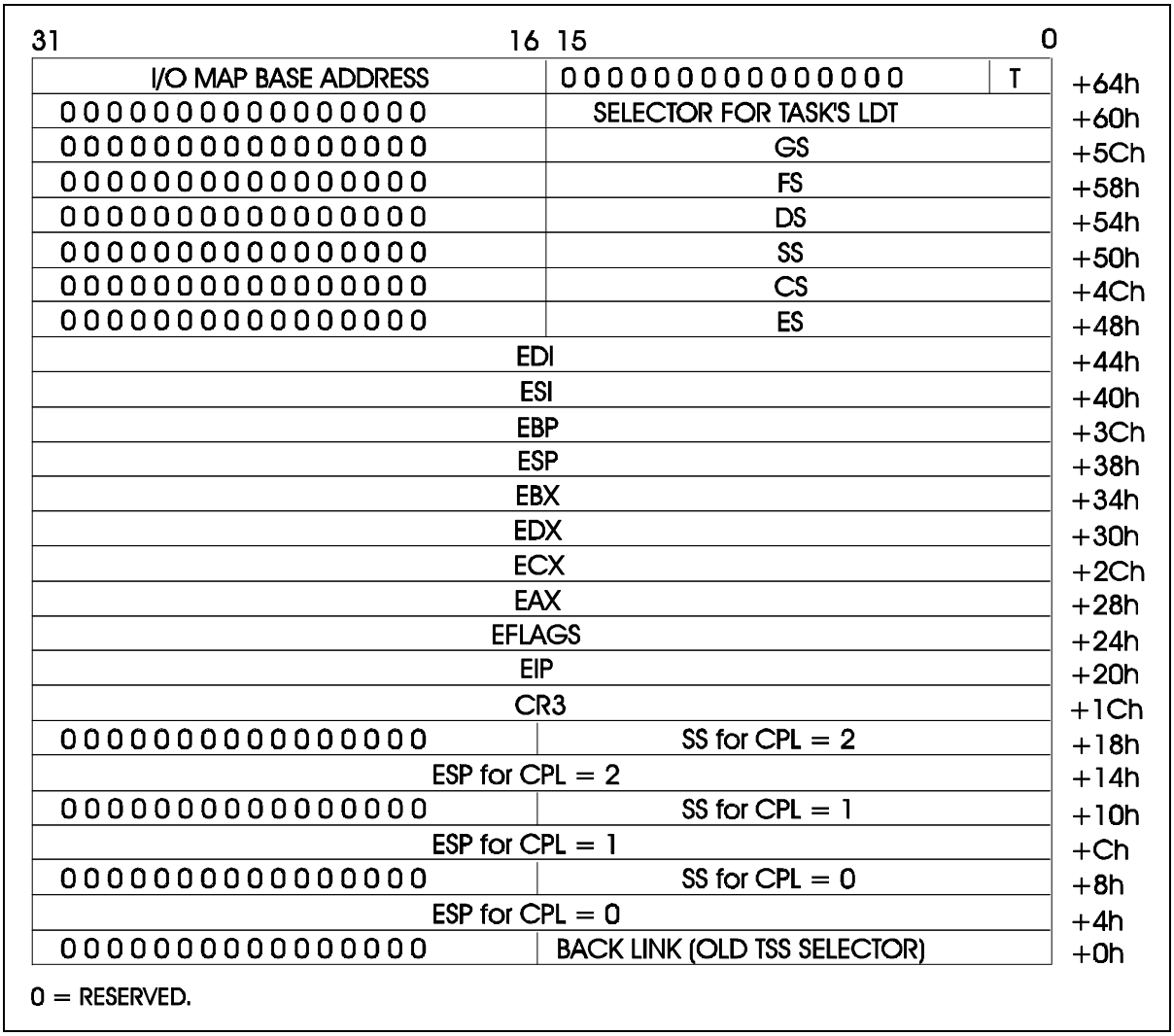


Figure 2-11. 32-Bit Task State Segment (TSS) Table

SELECTOR FOR TASK'S LDT	+2Ah
DS	+28h
SS	+26h
CS	+24h
ES	+22h
DI	+20h
SI	+1Eh
BP	+16h
SP	+1Ah
BX	+18h
DX	+16h
CX	+14h
AX	+12h
FLAGS	+10h
IP	+Eh
SS FOR PRIVILEGE LEVEL 2	+Ch
SP FOR PRIVILEGE LEVEL 2	+Ah
SS FOR PRIVILEGE LEVEL 1	+8h
SP FOR PRIVILEGE LEVEL 1	+6h
SS FOR PRIVILEGE LEVEL 0	+4h
SP FOR PRIVILEGE LEVEL 0	+2h
BACK LINK (OLD TSS SELECTOR)	+0h

Figure 2-12. 16-Bit Task State Segment (TSS) Table

2.4.4 IBM 6x86MX CPU Configuration Registers

The IBM 6x86MX CPU configuration registers are used to enable features in the IBM 6x86MX CPU. These registers assign non-cached memory areas, set up SMM, provide CPU identification information and control various features such as cache write policy, and bus locking control. There are four groups of registers within the IBM 6x86MX CPU configuration register set:

- 7 Configuration Control Registers (CCRx)
- 8 Address Region Registers (ARRx)
- 8 Region Control Registers (RCRx)

Access to the configuration registers is achieved by writing the register index number for the configuration register to I/O port 22h. I/O port 23h is then used for data transfer.

Each I/O port 23h data transfer must be preceded by a valid I/O port 22h register index selection. Otherwise, the current 22h, and the second and later I/O port 23h operations communicate through the I/O port to produce external I/O cycles. All reads from I/O port 22h produce external I/O cycles. Accesses that hit within the on-chip configuration registers do not generate external I/O cycles.

After reset, configuration registers with indexes C0-CFh and FC-FFh are accessible. To prevent potential conflicts with other devices which may use ports 22 and 23h to access their registers, the remaining registers (indexes D0-FBh) are accessible only if the MAPEN(3-0) bits in CCR3 are set to 1h. See Figure 2-16 (Page 2-29) for more information on the MAPEN(3-0) bit locations.

If MAPEN[3-0] = 1h, any access to indexes in the range 00-FFh will not create external I/O bus cycles. Registers with indexes C0-CFh, FC- FFh are accessible regardless of the state of MAPEN[3-0]. If the register index number is outside the C0-CFh or FC-FFh ranges, and MAPEN[3-0] are set to 0h, external I/O bus cycles occur. Table 2-11 (Page 2-25) lists the MAPEN[3-0] values required to access each IBM 6x86MX CPU configuration register. All bits in the configuration registers are initialized to zero following reset unless specified otherwise.

2.4.4.1 Configuration Control Registers

(CCR0 - CCR6) control several functions, including non-cacheable memory, write-back regions, and SMM features. A list of the configuration registers is listed in Table 2-11 (Page 2-25). The configuration registers are described in greater detail in the following pages.

Table 2-11. IBM 6x86MX CPU Configuration Registers

REGISTER NAME	ACRONYM	REGISTER INDEX	WIDTH (Bits)	MAPEN VALUE NEEDED FOR ACCESS
Configuration Control 0	CCR0	C0h	8	x
Configuration Control 1	CCR1	C1h	8	x
Configuration Control 2	CCR2	C2h	8	x
Configuration Control 3	CCR3	C3h	8	x
Configuration Control 4	CCR4	E8h	8	1
Configuration Control 5	CCR5	E9h	8	1
Configuration Control 6	CCR6	EAh	8	1
Address Region 0	ARR0	C4h - C6h	24	x
Address Region 1	ARR1	C7h - C9h	24	x
Address Region 2	ARR2	CAh - CCh	24	x
Address Region 3	ARR3	CDh - CFh	24	x
Address Region 4	ARR4	D0h - D2h	24	1
Address Region 5	ARR5	D3h - D5h	24	1
Address Region 6	ARR6	D6h - D8h	24	1
Address Region 7	ARR7	D9h - DBh	24	1
Region Control 0	RCR0	DCh	8	1
Region Control 1	RCR1	DDh	8	1
Region Control 2	RCR2	DEh	8	1
Region Control 3	RCR3	DFh	8	1
Region Control 4	RCR4	E0h	8	1
Region Control 5	RCR5	E1h	8	1
Region Control 6	RCR6	E2h	8	1
Region Control 7	RCR7	E3h	8	1

Note: x = Don’t Care

7	6	5	4	3	2	1	0
Reserved	*Reserved*	*Reserved*	*Reserved*	*Reserved*	*Reserved*	NC1	*Reserved*

Figure 2-13. IBM 6x86MX CPU Configuration Control Register 0 (CCR0)

Table 2-12. CCR0 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
1	NC1	No Cache 640 KByte - 1 MByte If = 1: Address region 640 KByte to 1 MByte is non-cacheable. If = 0: Address region 640 KByte to 1 MByte is cacheable.

Note: Bits 0, 2 through 7 are reserved.

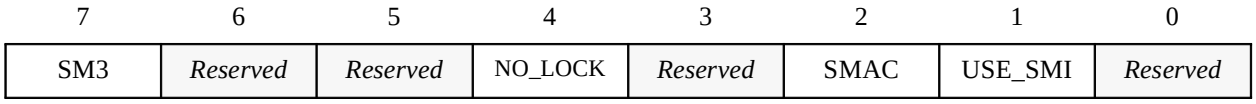


Figure 2-14. IBM 6x86MX CPU Configuration Control Register 1 (CCR1)

Table 2-13. CCR1 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
7	SM3	SMM Address Space Address Region 3 If = 1: Address Region 3 is designated as SMM address space.
4	NO_LOCK	Negate LOCK# If = 1: All bus cycles are issued with LOCK# pin negated except page table accesses and interrupt acknowledge cycles. Interrupt acknowledge cycles are executed as locked cycles even though LOCK# is negated. With NO_LOCK set, previously noncacheable locked cycles are executed as unlocked cycles and therefore, may be cached. This results in higher performance. Refer to Region Control Registers for information on eliminating locked CPU bus cycles only in specific address regions.
2	SMAC	System Management Memory Access If = 1: Any access to addresses within the SMM address space, access system management memory instead of main memory. SMI# input is ignored. Used when initializing or testing SMM memory. If = 0: No effect on access.
1	USE_SMI	Enable SMM and SMIACT# Pins If = 1: SMI# and SMIACT# pins are enabled. If = 0: SMI# pin ignored and SMIACT# pin is driven inactive.

Note: Bits 0, 3, 5 and 6 are reserved.

7	6	5	4	3	2	1	0
USE_SUSP	*Reserved*	*Reserved*	WPR1	SUSP_HLT	LOCK_NW	SADS	*Reserved*

Figure 2-15. IBM 6x86MX CPU Configuration Control Register 2 (CCR2)

Table 2-14. CCR2 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
7	USE_SUSP	Use Suspend Mode (Enable Suspend Pins) If = 1: SUSP# and SUSPA# pins are enabled. If = 0: SUSP# pin is ignored and SUSPA# pin floats.
4	WPR1	Write-Protect Region 1 If = 1: Designates any cacheable accesses in 640 KByte to 1 MByte address region are write protected.
3	SUSP_HLT	Suspend on Halt If = 1: Execution of the HLT instruction causes the CPU to enter low power suspend mode.
2	LOCK_NW	Lock NW If = 1: NW bit in CR0 becomes read only and the CPU ignores any writes to the NW bit. If = 0: NW bit in CR0 can be modified.
1	SADS	If = 1: CPU inserts an idle cycle following sampling of BRDY# and inserts an idle cycle prior to asserting ADS#

Note: Bits 0, 5 and 6 are reserved.

7	6	5	4	3	2	1	0
MAPEN3	MAPEN2	MAPEN1	MAPEN0	*Reserved*	LINBRST	NMI_EN	SMI_LOCK

Figure 2-16. IBM 6x86MX CPU Configuration Control Register 3 (CCR3)

Table 2-15. CCR3 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
7 - 4	MAPEN(3-0)	MAP Enable If = 1h: All configuration registers are accessible. If = 0h: Only configuration registers with indexes C0-CFh, FEh and FFh are accessible.
2	LINBRST	If = 1: Use linear address sequence during burst cycles. If = 0: Use “1 + 4” address sequence during burst cycles. The “1 + 4” address sequence is compatible with Pentium’s burst address sequence.
1	NMI_EN	NMI Enable If = 1: NMI interrupt is recognized while servicing an SMI interrupt. NMI_EN should be set only while in SMM, after the appropriate SMI interrupt service routine has been setup.
0	SMI_LOCK	SMI Lock If = 1: The following SMM configuration bits can only be modified while in an SMI service routine: CCR1: USE_SMI, SMAC, SM3 CCR3: NMI_EN CCR6: N, SMM_MODE ARR3: Starting address and block size. Once set, the features locked by SMI_LOCK cannot be unlocked until the RESET pin is asserted.

Note: Bit 3 is reserved.

7	6	5	4	3	2	1	0
CPUID	*Reserved*	*Reserved*	*Reserved*	*Reserved*	IORT2	IORT1	IORT

Figure 2-17. IBM 6x86MX CPU Configuration Control Register 4 (CCR4)

Table 2-16. CCR4 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
7	CPUID	Enable CPUID instruction. If = 1: the ID bit in the EFLAGS register can be modified and execution of the CPUID instruction occurs as documented in section 6.3. If = 0: the ID bit in the EFLAGS register can not be modified and execution of the CPUID instruction causes an invalid opcode exception.
2 - 0	IORT(2-0)	I/O Recovery Time Specifies the minimum number of bus clocks between I/O accesses: 0h = 1 clock delay 1h = 2 clock delay 2h = 4 clock delay 3h = 8 clock delay 4h = 16 clock delay 5h = 32 clock delay (default value after RESET) 6h = 64 clock delay 7h = no delay

Note: Bits 3 - 6 are reserved.

7	6	5	4	3	2	1	0
Reserved	*Reserved*	ARREN	*Reserved*	*Reserved*	*Reserved*	*Reserved*	WT_ALLOC

Figure 2-18. IBM 6x86MX CPU Configuration Control Register 5 (CCR5)

Table 2-17. CCR5 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
5	ARREN	Enable ARR Registers If = 1: Enables all ARR registers. If = 0: Disables the ARR registers. If SM3 is set, ARR3 is enabled regardless of the setting of ARREN.
0	WT_ALLOC	Write-Through Allocate If = 1: New cache lines are allocated for read and write misses. If = 0: New cache lines are allocated only for read misses.

Note: Bits 1 - 3 and 6 - 7 are reserved.

7	6	5	4	3	2	1	0
Reserved	N	*Reserved*	*Reserved*	*Reserved*	*Reserved*	WP_ARR3	SMM_MODE

Figure 2-19. IBM 6x86MX CPU Configuration Control Register 6 (CCR6)

Table 2-18. CCR6 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
6	N	Nested SMI Enable bit: If operating in Cyrix enhanced SMM mode and: If = 1: Enables nesting of SMI's If = 0: Disable nesting of SMI's. This bit is automatically CLEARED upon entry to every SMM routine and is SET upon every RSM. Therefore enabling/disabling of nested SMI can only be done while operating in SMM mode.
1	WP_ARR3	If = 1: Memory region defined by ARR3 is write protected when operating outside of SMM mode. If = 0: Disable write protection for memory region defined by ARR3. Reset State = 0.
0	SMM_MODE	If = 1: Enables Cyrix Enhanced SMM mode. If = 0: Disables Cyrix Enhanced SMM mode.

Note: Bit 1 is reserved.

2.4.4.2 Address Region Registers

The Address Region Registers (ARR0 - ARR7) (Figure 2-20) are used to specify the location and size for the eight address regions.

Attributes for each address region are specified in the Region Control Registers (RCR0-RCR7). ARR7 and RCR7 are used to define system main memory and differ from ARR0-6 and RCR0-6.

With non-cacheable regions defined on-chip, the IBM 6x86MX CPU delivers optimum performance by using advanced techniques to eliminate data dependencies and resource conflicts in its execution pipelines. If KEN# is active for accesses to regions defined as non-cacheable by the RCRs, the region is not cached. The RCRs take precedence in this case.

A register index, shown in Table 2-19 (Page 2-34) is used to select one of three bytes in each ARR.

The starting address of the ARR address region, selected by the START ADDRESS field, must be on a block size boundary. For example, a 128 KByte block is allowed to have a starting address of 0 KBytes, 128 KBytes, 256 KBytes, and so on.

The SIZE field bit definition is listed in (Page 2-34). If the SIZE field is zero, the address region is of zero size and thus disabled.

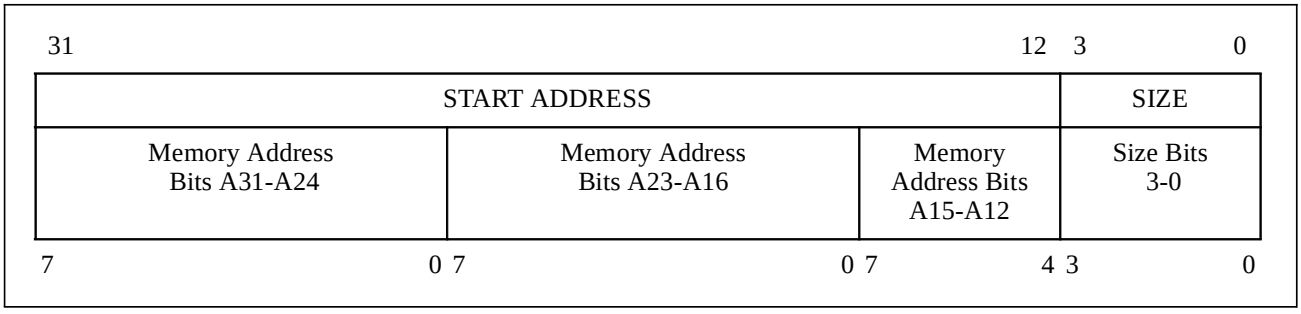


Figure 2-20. Address Region Registers (ARR0 - ARR7)

Table 2-19. ARR0 - ARR7 Register Index Assignments

ARR Register	Memory Address (A31 - A24)	Memory Address (A23 - A16)	Memory Address (A15 - A12)	Address Region Size (3 - 0)
ARR0	C4h	C5h	C6h	C6h
ARR1	C7h	C8h	C9h	C9h
ARR2	CAh	CBh	CCh	CCh
ARR3	CDh	CEh	CFh	CFh
ARR4	D0h	D1h	D2h	D2h
ARR5	D3h	D4h	D5h	D5h
ARR6	D6h	D7h	D8h	D8h
ARR7	D9h	DAh	DBh	DBh

Table 2-20. Bit Definitions for SIZE Field

SIZE (3-0)	BLOCK SIZE ARR0-6	BLOCK SIZE ARR7
0h	Disabled	Disabled
1h	4 KBytes	256 KBytes
2h	8 KBytes	512 KBytes
3h	16 KBytes	1 MBytes
4h	32 KBytes	2 MBytes
5h	64 KBytes	4 MBytes
6h	128 KBytes	8 MBytes
7h	256 KBytes	16 MBytes

SIZE (3-0)	BLOCK SIZE ARR0-6	BLOCK SIZE ARR7
8h	512 KBytes	32 MBytes
9h	1 MBytes	64 MBytes
Ah	2 MBytes	128 MBytes
Bh	4 MBytes	256 MBytes
Ch	8 MBytes	512 MBytes
Dh	16 MBytes	1 GBytes
Eh	32 MBytes	2 GBytes
Fh	4 GBytes	4 GBytes

2.4.4.3 Region Control Registers

The Region Control Registers (RCR0 - RCR7) specify the attributes associated with the ARRx address regions. The bit definitions for the region control registers are shown in Figure 2-21 (Page 2-36) and in Table 2-21 (Page 2-36). Cacheability, weak locking, write gathering, and cache write through policies can be activated or deactivated using the attribute bits.

If an address is accessed that is not in a memory region defined by the ARRx registers, the following conditions will apply:

- If the memory address is cached, write-back is enabled if WB/WT# is returned high.
- Writes are not gathered
- Strong locking takes place
- The memory access is cached, if KEN# is returned asserted.

Overlapping Conditions Defined. If two regions specified by ARRx registers overlap and conflicting attributes are specified, the following attributes take precedence:

- Write-back is disabled
- Writes are not gathered
- Strong locking takes place
- The overlapping regions are non-cacheable.

7	6	5	4	3	2	1	0
Reserved	INV_RGN	*Reserved*	WT	WG	WL	*Reserved*	CD

*Note: RCD is defined for RCR0-RCR6. RCE is defined for RCR7.

Figure 2-21. Region Control Registers (RCR0-RCR7)

Table 2-21. RCR0-RCR7 Bit Definitions

BIT POSITION	NAME	DESCRIPTION
6	INV_RGN	Applicable to RCR(0-6) only. If set, apply controls specified in RCRx to all memory addresses outside the region specified in corresponding ARR.
4	WT	Write-through- If set, defines the address region as write through instead of write back. This bit works in conjunction with the CR0_NW and PWT bits and the WB/WT# pin to determine write-through or write-back cacheability. See the Data Cache document for a complete description of how these various bits work in combination to affect cache write policy.
3	WG	Write Gathering - If set, enables write gathering for the associated address region. With WG enabled, multiple byte, word or dword writes to sequential addresses that would normally occur as individual cycles on the bus are collapsed, or “gathered” within the processor and then completed as a single write cycle. WG improves bus utilization and should be used on memory regions that are not sensitive to gathering.
2	WL	Weak Locking - If set, enables weak locking for that address region. With WL enabled, all bus cycles are issued with the LOCK# pin negated except for page table accesses. Interrupt acknowledge cycles are executed as locked cycles even though LOCK# is negated. With WL=1, previously non-cacheable locked cycles are executed as unlocked cycles and therefore, may be cached, resulting in higher CPU performance. Note that the NO_LOCK bit globally performs the same function that the WL bit performs on a single address region.
0	CD	Cache Disable - If set, defines the address region as non-cacheable. This bit works in conjunction with the CR0_CD and PCD bits and the KEN# pin to determine line cacheability. Whenever possible, the ARR/RCR combination should be used to define non-cacheable regions rather than using external address decoding and driving the KEN# pin as the IBM 6x86MX CPU can better utilize its advanced techniques for eliminating data dependencies and resource conflicts with non-cacheable regions defined on-chip.

Note: Bits 1, 5 and 7 are reserved.

Region Cache Disable (RCD). Setting RCD to a one defines the address region as non-cacheable. Whenever possible, the RCRs should be used to define non-cacheable regions rather than using external address decoding and driving the KEN# pin.

Region Cache Enable (RCE). Setting RCE to a one defines the address region as cacheable. RCE is used to define the system main memory as cacheable memory. It is implied that memory outside the region is non-cacheable.

Weak Locking (WL). Setting WL=1 enables weak locking for that address region. With WL enabled, all bus cycles are issued with the LOCK# pin negated except for page table accesses and interrupt acknowledge cycles. Interrupt acknowledge cycles are executed as locked cycles even though LOCK# is negated. With WL=1, previously non-cacheable locked cycles are executed as unlocked cycles and therefore, may be cached, resulting in higher performance. The NO_LOCK bit of CCR1 enables weak locking for the entire address space. The WL bit allows weak locking only for specific address regions. WL is independent of the cacheability of the address region.

Write Gathering (WG). Setting WG=1 enables write gathering for the associated address region. Write gathering allows multiple byte, word, or dword sequential address writes to accumulate in the on-chip write buffer. (As instructions are executed, the results are placed in a series of output buffers. These buffers are gathered into the final output buffer).

When access is made to a non-sequential memory location or when the 8-byte buffer becomes full, the contents of the buffer are written on the external 64-bit data bus. Performance is enhanced by avoiding as many as seven memory write cycles.

WG should not be used on memory regions that are sensitive to write cycle gathering. WG can be enabled for both cacheable and non-cacheable regions.

Write Through (WT). Setting WT=1 defines the address region as write-through instead of write-back, assuming the region is cacheable. Regions where system ROM are loaded (shadowed or not) should be defined as write-through.

2.5 Model Specific Registers

The IBM 6x86MX CPU contains four model specific registers (MSR0 - MSR3). These 64-bit registers are listed in Table 2-22.

Table 2-22. Machine Specific Register

REGISTER DESCRIPTION	MSR ADDRESS	REGISTER
Time Stamp Counter (TSC)	10h	MSR10
Counter Event Selection and Control Register	11h	MSR11
Performance Counter #0	12h	MSR12
Performance Counter #1	13h	MSR13

The MSR registers can be read using the RDMSR instruction, opcode 0F32h. During an MSR register read, the contents of the particular MSR register, specified by the ECX register, is loaded into the EDX:EAX registers.

The MSR registers can be written using the WRMSR instruction, opcode 0F30h. During a MSR register write the contents of EDX:EAX are loaded into the MSR register specified in the ECX register.

The RDMSR and WRMSR instructions are privileged instructions.

2.6 Time Stamp Counter

The Time Stamp Counter (TSC) Register (MSR10) is a 64-bit counter that counts the internal CPU clock cycles since the last reset. The TSC uses a continuous CPU core clock and will continue to count clock cycles even when the IBM 6x86MX CPU is suspend mode or shutdown.

The TSC can be accessed using the RDMSR and WRMSR instructions. In addition, the TSC can be read using the RDTSC instruction, opcode 0F31h. The RDTSC instruction loads the contents of the TSC into EDX:EAX. The use of the RDTSC instruction is restricted by the Time Stamp Disable, (TSD) flag in CR4. When the TSD flag is 0, the RDTSC instruction can be executed at any privilege level. When the TSD flag is 1, the RDTSC instruction can only be executed at privilege level 0.

2.7 Performance Monitoring

Performance monitoring allows counting of over a hundred different event occurrences and durations. Two 48-bit counters are used: Performance Monitor Counter 0 and Performance Monitor Counter 1. These two performance monitor counters are controlled by the Counter Event Control Register (MSR11). The performance monitor counters use a continuous CPU core clock and will continue to count clock cycles even when the IBM 6x86MX CPU is in suspend mode or shutdown.

2.8 Performance Monitoring Counters 1 and 2

The 48-bit Performance Monitoring Counters (PMC) Registers (MSR12, MSR13) count events as specified by the counter event control register.

The PMCs can be accessed by the RDMSR and WRMSR instructions. In addition, the PMCs can be read by the RDPMC instruction, opcode 0F33h. The RDPMC instruction loads the contents of the PMC register specified in the ECX register into EDX:EAX. The use of RDPMC instructions is restricted by the Performance Monitoring Counter Enable, (PCE) flag in C4.

When the PCE flag is set to 1, the RDPMC instruction can be executed at any privilege level. When the PCE flag is 0, the RDPMC instruction can only be executed at privilege level 0.

2.8.1 Counter Event Control Register

Register MSR 11h controls the two internal counters, #0 and #1. The events to be counted have been chosen based on the micro-architecture of the IBM 6x86MX processor. The control register for the two event counters is described in Figure 2-21 (Page 2-36) and Table 2-23 (Page 2-40).

2.8.1.1 PM Pin Control

The Counter Event Control register (MSR11) contains PM control fields that define the PM0 and PM1 pins as counter overflow indicators or counter event indicators. When defined as event counters, the PM pins indicate that one or more events occurred during a particular clock cycle and do not count the actual events.

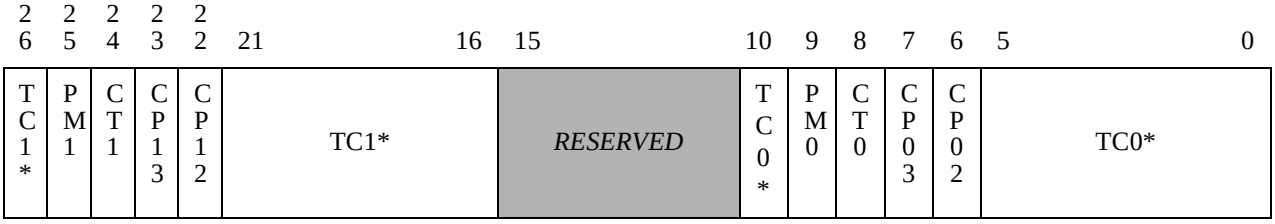
When defined as overflow indicators, the event counters can be preset with a value less the 2^{48}-1 and allowed to increment as events occur. When the counter overflows the PM pin becomes asserted.

2.8.1.2 Counter Type Control

The Counter Type bit determines whether the counter will count clocks or events. When counting clocks the counter operates as a timer.

2.8.1.3 CPL Control

The Current Privilege Level (CPL) can be used to determine if the counters are enabled. The CP02 bit in the MSR 11 register enables counting when the CPL is less than three, and the CP03 bit enables counting when CPL is equal to three. If both bits are set, counting is not dependent on the CPL level; if neither bit is set, counting is disabled.



*Note: Split Fields

Figure 2-22. Counter Event Control Register

Table 2-23. Counter Event Control Register Bit Definitions

BIT POSITION	NAME	DESCRIPTION
25	PM1	Define External PM1 Pin If = 1: PM1 pin indicates counter overflows If = 0: PM1 pin indicates counter events
24	CT1	Counter #1 Counter Type If = 1: Count clock cycles If = 0: Count events (reset state).
23	CP13	Counter #1 CPL 3 Enable If = 1: Enable counting when CPL=3. If = 0: Disable counting when CPL=3. (reset state)
22	CP12	Counter #1 CPL Less Than 3 Enable If = 1: Enable counting when CPL < 3. If = 0: Disable counting when CPL < 3. (reset state)
26, 21 - 16	TC1(5-0)	Counter #1 Event Type Reset state = 0
9	PM0	Define External PM0 Pin If = 1: PM0 pin indicates counter overflows If = 0: PM0 pin indicates counter events
8	CT0	Counter #0 Counter Type If = 1: Count clock cycles If = 0: Count events (reset state).
7	CP03	Counter #0 CPL 3 Enable If = 1: Enable counting when CPL=3. If = 0: Disable counting when CPL=3. (reset state)
6	CP02	Counter #0 CPL Less Than 3 Enable If = 1: Enable counting when CPL < 3. If = 0: Disable counting when CPL < 3. (reset state)
10, 5 - 0	TC0(5-0)	Counter #0 Event Type Reset state = 0

Note: Bits 10 - 15 are reserved.

2.8.2 Event Type and Description

The events that can be counted by the performance monitoring counters are listed in Table 2-24. Each of the 127 event types is assigned an event number.

A particular event number to be counted is placed in one of the MSR 11 Event Type fields. There is a separate field for counter #0 and #1.

The events are divided into two groups. The occurrence type events and duration type events. The occurrence type events, such as hardware interrupts, are counted as single events. The duration type events such as “clock while bus cycles are in progress” count the number of clock cycles that occur during the event.

During occurrence type events, the PM pins are configured to indicate the counter has incremented The PM pins will then assert every time the counter increments in regards to an occurrence event. Under the same PM control, for a duration event the PM pin will stay asserted for the duration of the event.

Table 2-24. Event Type Register

NUMBER	COUNTER 0	COUNTER 1	DESCRIPTION	TYPE
00h	yes	yes	Data Reads	Occurrence
01h	yes	yes	Data Writes	Occurrence
02h	yes	yes	Data TLB Misses	Occurrence
03h	yes	yes	Cache Misses: Data Reads	Occurrence
04h	yes	yes	Cache Misses: Data Writes	Occurrence
05h	yes	yes	Data Writes that hit on Modified or Exclusive Liens	Occurence
06h	yes	yes	Data Cache Lines Written Back	Occurrence
07h	yes	yes	External Inquiries	Occurrence
08h	yes	yes	External Inquires that hit	Occurrence
09h	yes	yes	Memory Accesses in both pipes	Occurrence
0Ah	yes	yes	Cache Bank conflicts	Occurrence
0Bh	yes	yes	Misaligned data references	Occurrence
0Ch	yes	yes	Instruction Fetch Requests	Occurrence
0Dh	yes	yes	L2 TLB Code Misses	Occurrence
0Eh	yes	yes	Cache Misses: Instruction Fetch	Occurrence
0Fh	yes	yes	Any Segment Register Load	Occurrence
10h	yes	yes	Reserved	Occurrence
11h	yes	yes	Reserved	Occurrence
12h	yes	yes	Any Branch	Occurrence

Table 2-24. Event Type Register (Continued)

NUMBER	COUNTER 0	COUNTER 1	DESCRIPTION	TYPE
13h	yes	yes	BTB hits	Occurrence
14h	yes	yes	Taken Branches or BTB hits	Occurrence
15h	yes	yes	Pipeline Flushes	Occurrence
16h	yes	yes	Instructions executed in both pipes	Occurrence
17h	yes	yes	Instructions executed in Y pipe	Occurrence
18h	yes	yes	Clocks while bus cycles are in progress	Duration
19h	yes	yes	Pipe Stalled by full write buffers	Duration
1Ah	yes	yes	Pipe Stalled by waiting on data memory reads	Duration
1Bh	yes	yes	Pipe Stalled by writes to not-Modified or not-Exclusive cache lines.	Duration
1Ch	yes	yes	Locked Bus Cycles	Occurrence
1Dh	yes	yes	I/O Cycles	Occurrence
1Eh	yes	yes	Non-cacheable Memory Requests	Occurrence
1Fh	yes	yes	Pipe Stalled by Address Generation Interlock	Duration
20h	yes	yes	Reserved	
21h	yes	yes	Reserved	
22h	yes	yes	Floating Point Operations	Occurrence
23h	yes	yes	Breakpoint Matches on DR0 register	Occurrence
24h	yes	yes	Breakpoint Matches on DR1 register	Occurrence
25h	yes	yes	Breakpoint Matches on DR2 register	Occurrence
26h	yes	yes	Breakpoint Matches on DR3 register	Occurrence
27h	yes	yes	Hardware Interrupts	Occurrence
28h	yes	yes	Data Reads or Data Writes	Occurrence
29h	yes	yes	Data Read Misses or Data Write Misses	Occurrence
2Bh	yes	no	MMX Instruction Executed in X pipe	Occurrence
2Bh	no	yes	MMX Instruction Executed in Y pipe	Occurrence
2Dh	yes	no	EMMS Instruction Executed	Occurrence
2Dh	no	yes	Transition Between MMX Instruction and FP Instructions	Occurrence
2Eh	no	yes	Reserved	
2Fh	yes	no	Saturating MMX Instructions Executed	Occurrence
2Fh	no	yes	Saturations Performed	Occurrence
30h	yes	no	Reserved	
31h	yes	no	MMX Instruction Data Reads	Occurrence
32h	yes	no	Reserved	
32h	no	yes	Taken Branches	Occurrence
33h	no	yes	Reserved	
34h	yes	no	Reserved	
34h	no	yes	Reserved	
35h	yes	no	Reserved	

Table 2-24. Event Type Register (Continued)

NUMBER	COUNTER 0	COUNTER 1	DESCRIPTION	TYPE
35h	no	yes	Reserved	
36h	yes	no	Reserved	
36h	no	yes	Reserved	
37h	yes	no	Returns Predicted Incorrectly	Occurrence
37h	no	yes	Return Predicted (Correctly and Incorrectly)	Occurrence
38h	yes	no	MMX Instruction Multiply Unit Interlock	Duration
38h	no	yes	MODV/MOVQ Store Stall Due to Previous Operation	Duration
39h	yes	no	Returns	Occurrence
39h	no	yes	RSB Overflows	Occurrence
3A	yes	no	BTB False Entries	Occurrence
3A	no	yes	BTB Miss Prediction on a Not-Taken Back	Occurrence
3B	yes	no	Number of Clock Stalled Due to Full Write Buffers While Executing	Duration
3B	no	yes	Stall on MMX Instruction Write to E or M Line	Duration
3C - 3Fh	yes	yes	Reserved	Duration
40h	yes	yes	L2 TLB Misses (Code or Data)	Occurrence
41h	yes	yes	L1 TLB Data Miss	Occurrence
42h	yes	yes	L1 TLB Code Miss	Occurrence
43h	yes	yes	L1 TLB Miss (Code or Data)	Occurrence
44h	yes	yes	TLB Flushes	Occurrence
45h	yes	yes	TLB Page Invalidates	Occurrence
46h	yes	yes	TLB Page Invalidates that hit	Occurrence
47h	yes	yes	Reserved	
48h	yes	yes	Instructions Decoded	Occurrence
49h	yes	yes	Reserved	

2.9 Debug Registers

Six debug registers (DR0-DR3, DR6 and DR7), shown in Figure 2-23, support debugging on the IBM 6x86MX CPU. The bit definitions for the debug registers are listed in Table 2-25 (Page 2-45).

Memory addresses loaded in the debug registers, referred to as “breakpoints”, generate a debug exception when a memory access of the specified type occurs to the specified address. A data breakpoint can be specified for a particular kind of memory access such as a read or a write. Code breakpoints can also be set allowing debug exceptions to occur whenever a given code access (execution) occurs.

The size of the debug target can be set to 1, 2, or 4 bytes. The debug registers are accessed via MOV instructions which can be executed only at privilege level 0.

The Debug Address Registers (DR0-DR3) each contain the linear address for one of four possible breakpoints. Each breakpoint is further specified by bits in the Debug Control Register (DR7). For each breakpoint address in DR0-DR3, there are corresponding fields L, R/W, and LEN in DR7 that specify the type of memory access associated with the breakpoint.

The R/W field can be used to specify instruction execution as well as data access breakpoints. Instruction execution breakpoints are always taken before execution of the instruction that matches the breakpoint.

The Debug Status Register (DR6) reflects conditions that were in effect at the time the debug exception occurred. The contents of the DR6 register are not automatically cleared by the processor after a debug exception occurs and, therefore, should be cleared by software at the appropriate time.

31 30	29 28	27 26	25 24	23 22	21 20	19 18	17 16	15 14	13	12 11 10	9	8	7	6	5	4	3	2	1	0	
LEN 3	R/W 3	LEN 2	R/W 2	LEN 1	R/W 1	LEN 0	R/W 0	0 0	GD	0 0 1	GE	LE	G3	L3	G2	L2	G1	L1	G0	L0	DR7
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0								BT	BS	0	0 1 1 1 1 1 1 1 1						B3	B2	B1	B0	DR6
BREAKPOINT 3 LINEAR ADDRESS																					DR3
BREAKPOINT 2 LINEAR ADDRESS																					DR2
BREAKPOINT 1 LINEAR ADDRESS																					DR1
BREAKPOINT 0 LINEAR ADDRESS																					DR0

ALL BITS MARKED AS 0 OR 1 ARE RESERVED AND SHOULD NOT BE MODIFIED.

Figure 2-23. Debug Registers

Code execution breakpoints may also be generated by placing the breakpoint instruction (INT 3) at the location where control is to be regained. Additionally, the single-step feature may be enabled by setting the TF flag in the EFLAGS register. This causes the processor to perform a debug exception after the execution of every instruction.

Table 2-25. DR6 and DR7 Debug Register Field Definitions

REGISTER	FIELD	# OF BITS	DESCRIPTION
DR6	Bi	1	Bi is set by the processor if the conditions described by DRi, R/Wi, and LENi occurred when the debug exception occurred, even if the breakpoint is not enabled via the Gi or Li bits.
	BT	1	BT is set by the processor before entering the debug handler if a task switch has occurred to a task with the T bit in the TSS set.
	BS	1	BS is set by the processor if the debug exception was triggered by the single-step execution mode (TF flag in EFLAGS set).
DR7	R/Wi	2	Specifies type of break for the linear address in DR0, DR1, DR3, DR4: 00 - Break on instruction execution only 01 - Break on data writes only 10 - Not used 11 - Break on data reads or writes.
	LENi	2	Specifies length of the linear address in DR0, DR1, DR3, DR4: 00 - One byte length 01 - Two byte length 10 - Not used 11 - Four byte length.
	Gi	1	If set to a 1, breakpoint in DRi is globally enabled for all tasks and is not cleared by the processor as the result of a task switch.
	Li	1	If set to a 1, breakpoint in DRi is locally enabled for the current task and is cleared by the processor as the result of a task switch.
	GD	1	Global disable of debug register access. GD bit is cleared whenever a debug exception occurs.

2.10 Test Registers

The test registers can be used to test the on-chip unified cache and to test the main TLB.

Test registers TR3, TR4, and TR5 are used to test the unified cache. Use of these registers is described with the memory caches later in this chapter in section 2.13.1.1 on page 2-58.

Test registers TR6 and TR7 are used to test the TLB. Use of these test registers is described in section 2.12.4.1 on page 2-54.

2.11 Address Space

The IBM 6x86MX CPU can directly address 64 KBytes of I/O space and 4 GBytes of physical memory (Figure 2-24).

Memory Address Space. Access can be made to memory addresses between 0000 0000h and FFFF FFFFh. This 4 GByte memory space can be accessed using byte, word (16 bits), or doubleword (32 bits) format. Words and doublewords are stored in consecutive memory bytes with the low-order byte located in the lowest address. The physical address of a word or doubleword is the byte address of the low-order byte.

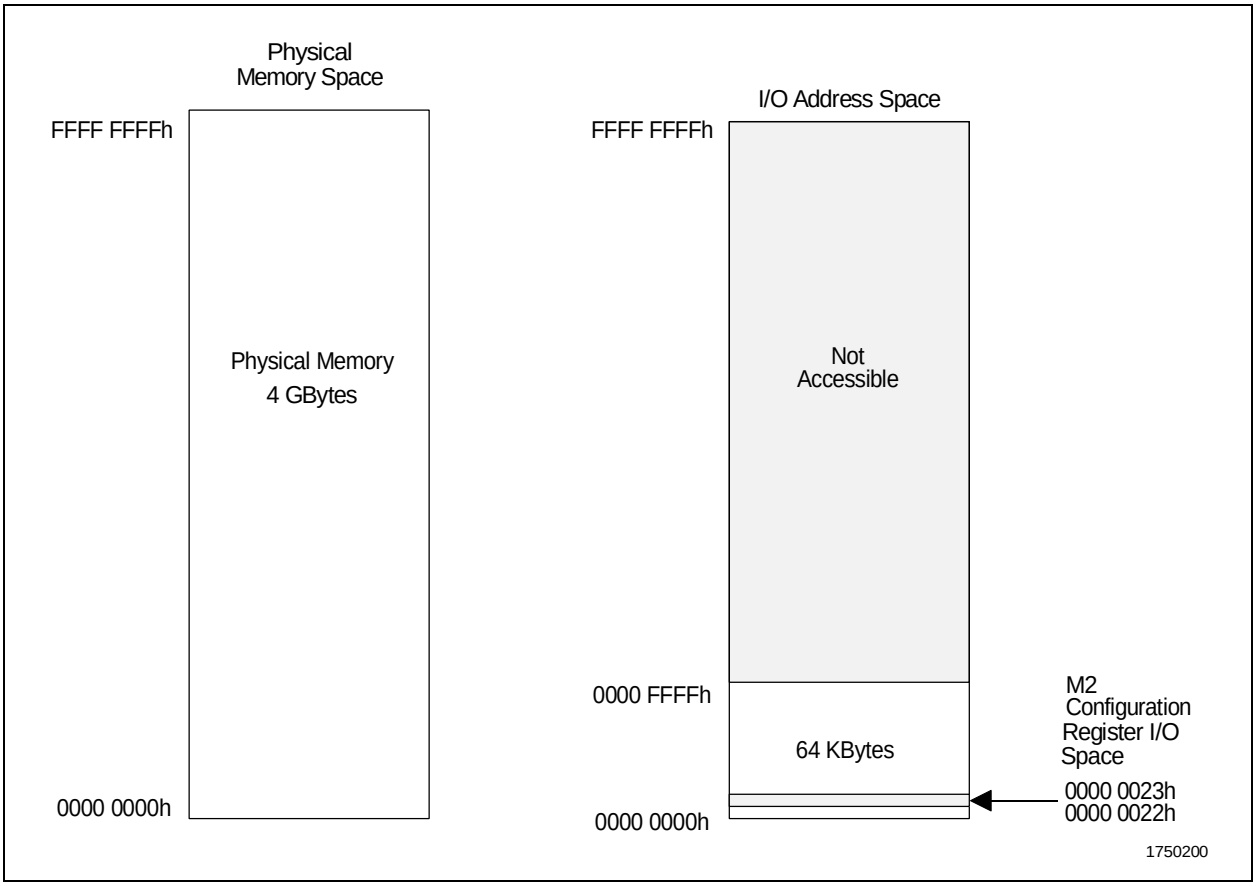


Figure 2-24 . Memory and I/O Address Spaces

I/O Address Space. The IBM 6x86MX I/O address space is accessed using IN and OUT instructions to addresses referred to as “ports”. The accessible I/O address space size is 64 KBytes and can be accessed through 8-bit, 16-bit or 32-bit ports. The execution of any IN or OUT instruction causes the M/IO# pin to be driven low, thereby selecting the I/O space instead of memory space.

The accessible I/O address space ranges between locations 0000 0000h and 0000 FFFFh (64 KBytes). The I/O locations (ports) 22h and 23h can be used to access the IBM 6x86MX configuration registers.

2.12 Memory Addressing Methods

With the IBM 6x86MX CPU, memory can be addressed using nine different addressing modes (Table 2-26, Page 2-49). These addressing modes are used to calculate an offset address often referred to as an effective address. Depending on the operating mode of the CPU, the offset is then combined using memory management mechanisms to create a physical address that actually addresses the physical memory devices.

Memory management mechanisms on the IBM 6x86MX CPU consist of segmentation and paging. Segmentation allows each program to use several independent, protected address spaces. Paging supports a memory subsystem that simulates a large address space using a small amount of RAM and disk storage for physical memory. Either or both of these mechanisms can be used for management of the IBM 6x86MX CPU memory address space.

2.12.1 Offset Mechanism

The offset mechanism computes an offset (effective) address by adding together one or more of three values: a base, an index and a displacement. When present, the base is the value of one of the eight 32-bit general registers. The index if present, like the base, is a value that is in one of the eight 32-bit general purpose registers (not including the ESP register). The index differs from the base in that the index is first multiplied by a scale factor of 1, 2, 4 or 8 before the summation is made. The third component added to the memory address calculation is the displacement. The displacement is a value of up to 32-bits in length supplied as part of the instruction. Figure 2-25 illustrates the calculation of the offset address.

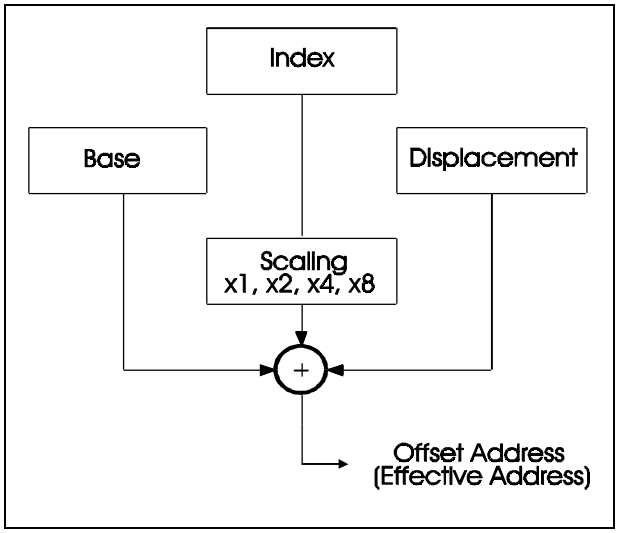


Figure 2-25. Offset Address Calculation

Nine valid combinations of the base, index, scale factor and displacement can be used with the IBM 6x86MX CPU instruction set. These combinations are listed in Table 2-26. The base and index both refer to contents of a register as indicated by [Base] and [Index].

Table 2-26. Memory Addressing Modes

ADDRESSING MODE	BASE	INDEX	SCALE FACTOR (SF)	DISPLACEMENT (DP)	OFFSET ADDRESS (OA) CALCULATION
Direct				x	OA = DP
Register Indirect	x				OA = [BASE]
Based	x			x	OA = [BASE] + DP
Index		x		x	OA = [INDEX] + DP
Scaled Index		x	x	x	OA = ([INDEX] * SF) + DP
Based Index	x	x			OA = [BASE] + [INDEX]
Based Scaled Index	x	x	x		OA = [BASE] + ([INDEX] * SF)
Based Index with Displacement	x	x		x	OA = [BASE] + [INDEX] + DP
Based Scaled Index with Displacement	x	x	x	x	OA = [BASE] + ([INDEX] * SF) + DP

2.12.2 Memory Addressing

Real Mode Memory Addressing

In real mode operation, the IBM 6x86MX CPU only addresses the lowest 1 MByte of memory. To calculate a physical memory address, the 16-bit segment base address located in the selected segment register is multiplied by 16 and then the 16-bit offset address is added. The resulting 20-bit address is then extended. Three hexadecimal zeros are added as upper address bits to create the 32-bit physical address. Figure 2-26 illustrates the real mode address calculation.

The addition of the base address and the offset address may result in a carry. Therefore, the resulting address may actually contain up to 21 significant address bits that can address memory in the first 64 KBytes above 1 MByte.

Protected Mode Memory Addressing

In protected mode three mechanisms calculate a physical memory address (Figure 2-27, Page 2-51).

- **Offset Mechanism** that produces the offset or effective address as in real mode.
- **Selector Mechanism** that produces the base address.
- **Optional Paging Mechanism** that translates a linear address to the physical memory address.

The offset and base address are added together to produce the linear address. If paging is not enabled, the linear address is used as the physical memory address. If paging is enabled, the paging mechanism is used to translate the linear address into the physical address. The offset mechanism is described earlier in this section and applies to both real and protected mode. The selector and paging mechanisms are described in the following paragraphs.

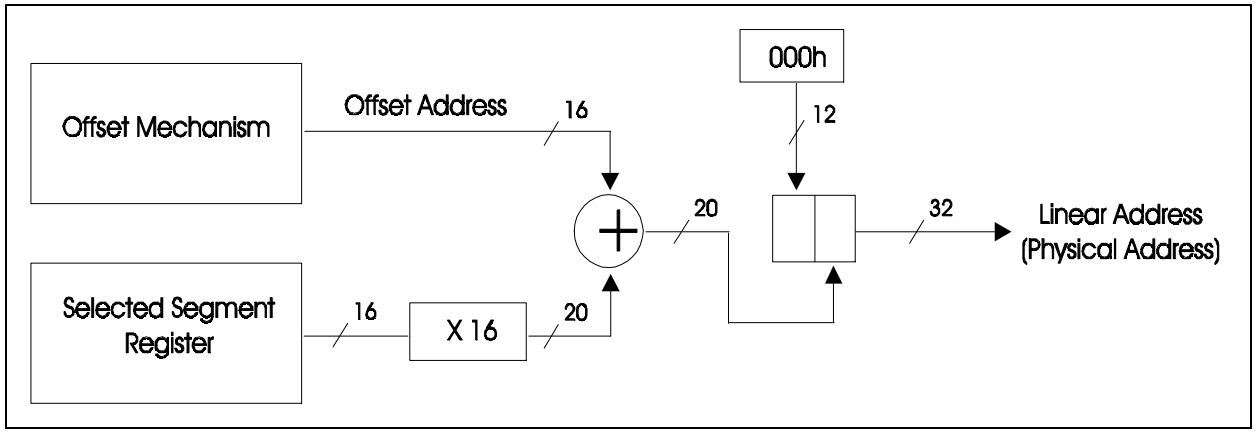


Figure 2-26. Real Mode Address Calculation

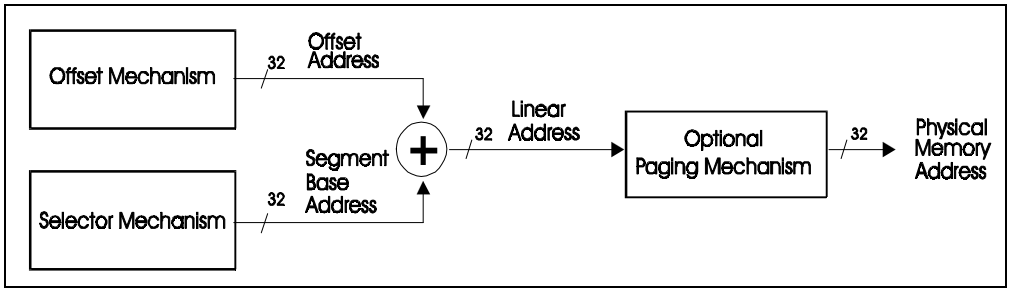


Figure 2-27. Protected Mode Address Calculation

2.12.3 Selector Mechanism

Using segmentation, memory is divided into an arbitrary number of segments, each containing usually much less than the 2^{32} byte (4 GByte) maximum.

The six segment registers (CS, DS, SS, ES, FS and GS) each contain a 16-bit selector that is used when the register is loaded to locate a segment descriptor in either the global descriptor table (GDT) or the local descriptor table (LDT). The segment descriptor defines the base address, limit, and attributes of the selected segment and is cached on the IBM 6x86MX CPU as a result of loading the selector. The cached descriptor contents are not visible to the programmer. When a memory reference occurs in protected mode, the linear address is generated by adding the segment base address in the hidden portion of the segment register to the offset address. If paging is not enabled, this linear address is used as the physical memory address. Figure 2-28 illustrates the operation of the selector mechanism.

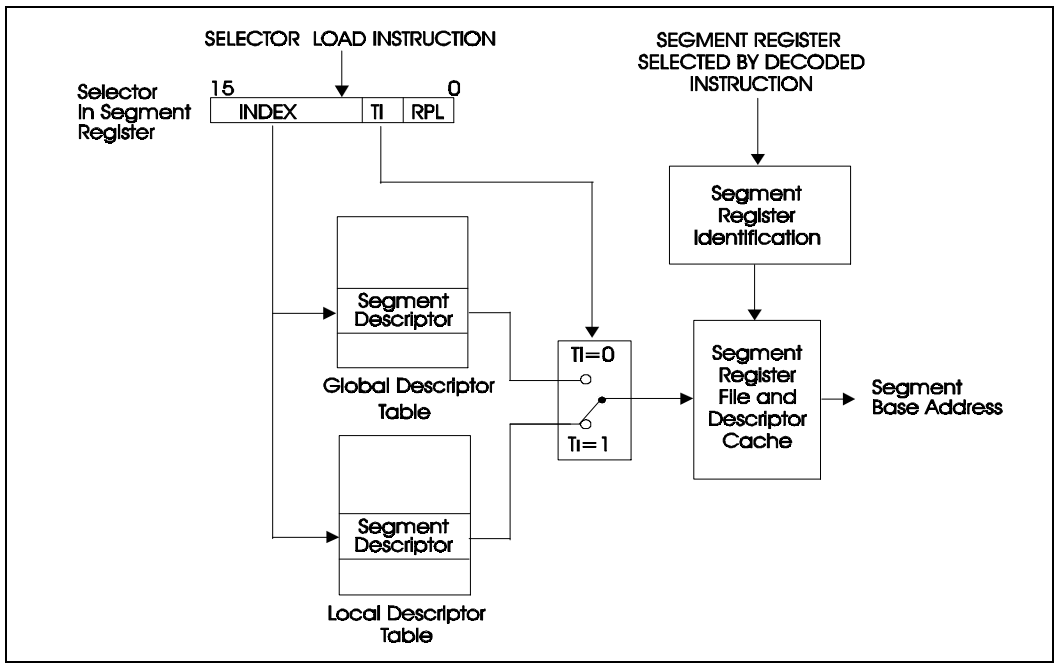


Figure 2-28. Selector Mechanism

2.12.4 Paging Mechanism

The paging mechanism translates linear addresses to their corresponding physical addresses. The page size is always 4 KBytes. Paging is activated when the PG and the PE bits within the CR0 register are set.

The paging mechanism translates the 20 most significant bits of a linear address to a physical address. The linear address is divided into three fields DTI, PTI, PFO (Figure 2-29, Page 2-53). These fields respectively select:

- an entry in the directory table,
- an entry in the page table selected by the directory table
- the offset in the physical page selected by the page table

The directory table and all the page tables can be considered as pages as they are 4-KBytes in size and are aligned on 4-KByte boundaries. Each entry in these tables is 32 bits in length. The fields within the entries are detailed in Figure 2-30 (Page 2-53) and Table 2-27 (Page 2-54).

A single page directory table can address up to 4 GBytes of virtual memory (1,024 page tables—each table can select 1,024 pages and each page contains 4 KBytes).

Translation Lookaside Buffer (TLB) is made up of two caches (Figure 2-29, Page 2-53).

- the L1 TLB caches page tables entries
- the L2 TLB stores PTEs that have been evicted from the L1 TLB

The L1 TLB is a 16-entry direct-mapped dual ported cache. The L2 TLB is a 384 entry, 6-way, dual ported cache.

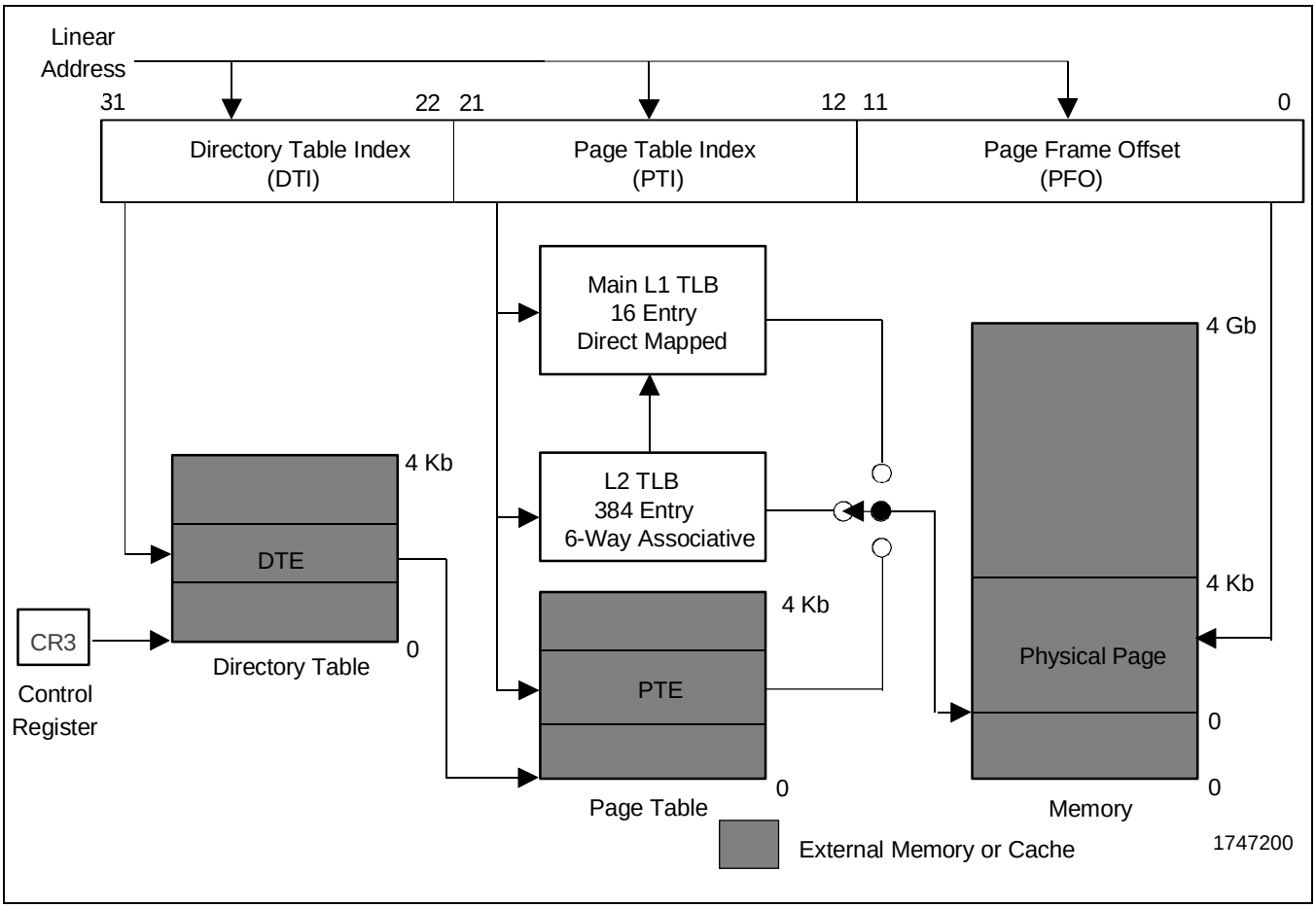


Figure 2-29. Paging Mechanism

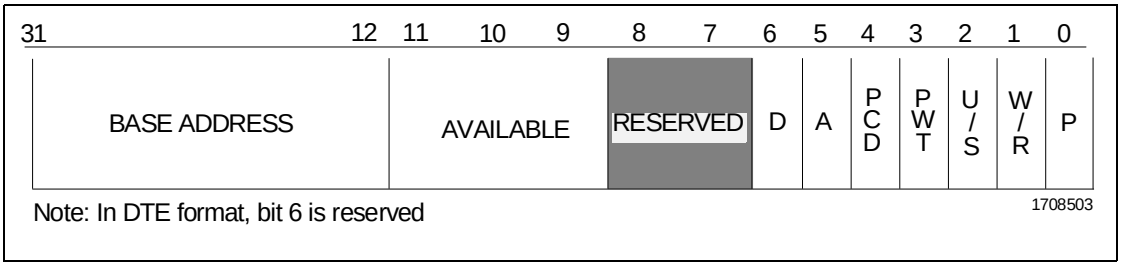


Figure 2-30. Directory and Page Table Entry (DTE and PTE) Format

Table 2-27. Directory and Page Table Entry (DTE and PTE) Bit Definitions

BIT POSITION	FIELD NAME	DESCRIPTION
31-12	BASE ADDRESS	Specifies the base address of the page or page table.
11-9	--	Undefined and available to the programmer.
8-7	--	Reserved and not available to the programmer.
6	D	Dirty Bit. If set, indicates that a write access has occurred to the page (PTE only, undefined in DTE).
5	A	Accessed Flag. If set, indicates that a read access or write access has occurred to the page.
4	PCD	Page Caching Disable Flag. If set, indicates that the page is not cacheable in the on-chip cache.
3	PWT	Page Write-Through Flag. If set, indicates that writes to the page or page tables that hit in the on-chip cache must update both the cache and external memory.
2	U/S	User/Supervisor Attribute. If set (user), page is accessible at privilege level 3. If clear (supervisor), page is accessible only when CPL ≤ 2.
1	W/R	Write/Read Attribute. If set (write), page is writable. If clear (read), page is read only.
0	P	Present Flag. If set, indicates that the page is present in RAM memory, and validates the remaining DTE/PTE bits. If clear, indicates that the page is not present in memory and the remaining DTE/PTE bits can be used by the programmer.

For a TLB hit, the TLB eliminates accesses to external directory and page tables.

The L1 TLB is a small cache optimized for speed whereas the L2 TLB is a much larger cache optimized for capacity. The L2 TLB is a proper superset of the L1 TLB.

The TLB must be flushed by the software when entries in the page tables are changed. Both the L1 and L2 TLBs are flushed whenever the CR3 register is loaded. A particular page can be flushed from the TLBs by using the INVLPG instruction.

2.12.4.1 Translation Lookaside Buffer Testing

The L1 and L2 Translation Lookaside Buffers (TLBs) can be tested by writing, then reading from the same TLB location. The operation to be performed is determined by the command (CMD) field (Table 2-28, Page 2-54) in the TR6 register.

Table 2-28. CMD Field

CMD	OPERATION	LINEAR ADDRESS BITS
x00	Write to L1	15 - 12
x01	Write to L2	17 - 12
010	Read from L1 X port	15 -12
011	Read from L2 X port	17 -12
110	Read from L1 Y port	15 - 12
110	Read from L2 Y port	17 - 12

TLB Write

To perform a write to the IBM 6x86MX TLBs, the TR7 register (Figure 2-31) is loaded with the desired physical address as well as the PCD and PWT bits. For a write to the L2 TLB, the SET field of TR7 must be also specified. The H1, H2, and HSET fields of TR7 are not used. The TR6 register is then loaded with the linear address, V, D, U, W and A fields and the appropriate CMD. For a L1 TLB write, the TLB entry is selected by bits 15-12 of the linear address. For a L2 TLB write, the TLB entry is selected by bits 17-12 of the linear address and the SET field of TR7.

TLB Read

For a L1 LTB read, the TR6 register is loaded with the linear address and the appropriate CMD. The L1 TLB entry selected by bits 15-12 of the linear address will then be accessed. The linear address, V, D, PG, U, W and A fields of TR6 and the physical address, PCD and PWT fields of TR7 are loaded from the specified L1 entry. The H1 bit of TR7 will indicate if the specified linear address hit in the L1 TLB.

For a L2 TLB read, the TR7 register is loaded with the desired SET. The TR6 register is then loaded with the linear address and the appropriate CMD. The L2 TLB entry selected by bits 17-12 of the linear address and the SET field in TR7 will then be accessed. The linear address, V,D, PG, V, W, and A fields of TR6 and the physical address, PCD and PWT fields of TR7 are loaded from the specified L2 entry. The H2 bit of TR7 will indicate if the specified linear address hit in the L2 TLB. If there was an L2 hit, the HSET field of TR7 will indicate which SET hit.

The TLB test register fields are defined in Table 2-29. (Page 2-56).

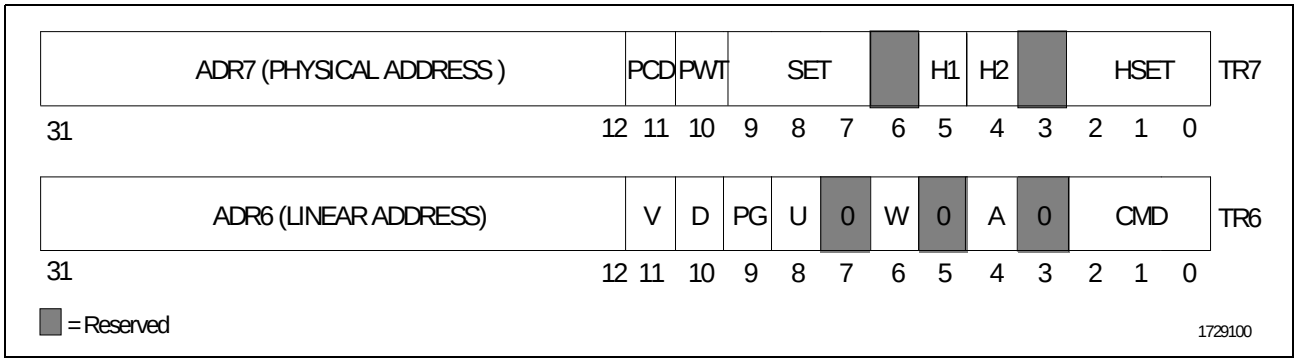


Figure 2-31. TLB Test Registers

Table 2-29. TLB Test Register Bit Definitions

REGISTER NAME	NAME	RANGE	DESCRIPTION
TR7	ADR7	31-12	Physical address or variable page size mechanism mask. TLB lookup: data field from the TLB. TLB write: data field written into the TLB.
	PCD	11	Page-level cache disable bit (PCD). Corresponds to the PCD bit of a page table entry.
	PWT	10	Page-level cache write-through bit (PWT). Corresponds to the PWT bit of a page table entry.
	SET	9-7	L2 TLB Set Selection (0h - 5h)
	H1	5	Hit in L1 TLB
	H2	4	Hit in L2 TLB
	HSET	2-0	L2 Set Selection when L2 TLB hit occurred (0h - 5h)
TR6	ADR6	31-12	Linear Address. TLB lookup: The TLB is interrogated per this address. If one and only one match occurs in the TLB, the rest of the fields in TR6 and TR7 are updated per the matching TLB entry. TLB write: A TLB entry is allocated to this linear address.
	V	11	PTE Valid. TLB write: If set, indicates that the TLB entry contains valid data. If clear, target entry is invalidated.
	D	10	Dirty Attribute Bit
	PG	9	Page Global
	U	8	User/Supervisor Attribute Bit
	W	6	Write Protect bit.
	CMD	2-0	Array Command Select. Determines TLB array command. Refer to Table 2-28, Page 2-54.

2.13 Memory Caches

The IBM 6x86MX CPU contains two memory caches as described in Chapter 1. The Unified Cache acts as the primary data cache, and secondary instruction cache. The Instruction Line Cache is the primary instruction cache and provides a high speed instruction stream for the Integer Unit.

The unified cache is dual-ported allowing simultaneous access to any two unique banks. Two different banks may be accessed at the same time permitting any two of the following operations to occur in parallel:

- Code fetch
- Data read (X pipe, Y pipe or FPU)
- Data write (X pipe, Y pipe or FPU).

2.13.1 Unified Cache MESI States

The unified cache lines are assigned one of four MESI states as determined by MESI bits stored in tag memory. Each 32-byte cache line is divided into two 16-byte sectors. Each sector contains its own MESI bits. The four MESI states are described below:

Modified MESI cache lines are those that have been updated by the CPU, but the corresponding main memory location has not yet been updated by an external write cycle. Modified cache lines are referred to as dirty cache lines.

Exclusive MESI lines are lines that are exclusive to the IBM 6x86MX CPU and are not duplicated within another caching agent's cache within the same system. A write to this cache line may be performed without issuing an external write cycle.

Shared MESI lines may be present in another caching agent's cache within the same system. A write to this cache line forces a corresponding external write cycle.

Invalid MESI lines are cache lines that do not contain any valid data.

2.13.1.1 Unified Cache Testing

The TR3, TR4, and TR5 on-chip test registers provide information so the unified cache can be tested. This information determines what particular area will be tested. Fields within these test registers identify which area of the cache will be selected for testing.

Cache Organization. The unified cache (Figure 2-32) is divided into 32-bytes lines. This cache is divided into four sets. Since a set (as well as the cache) is smaller than main memory, each line in the set corresponds to more than one line in main memory. When a cache line is allocated, bits A31-A14 of the main memory address are stored in the cache line tag. The remaining address bits are used to identify the specific 32-byte cache line (A13-A5), and the specific 4-byte entry within the cache line (A4-A2).

Test Initiation. A test register operation is initiated by writing to the TR5 register shown in Figure 2-33 (Page 2-59) using a special MOV instruction. The TR5 CTL field, detailed in Table 2-30 (Page 2-59), determines the function to be performed. For cache writes, the registers TR4 and TR3 must be initialized before a write is made to TR5. Eight 4-byte accesses are required to access a complete cache line.

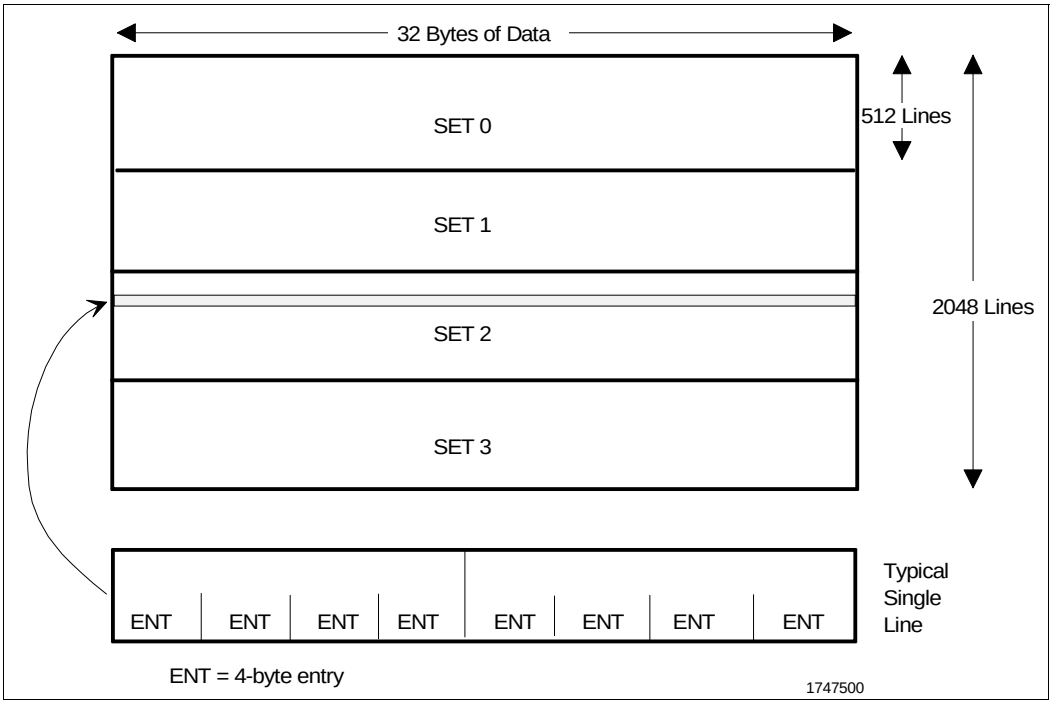


Figure 2-32. Unified Cache

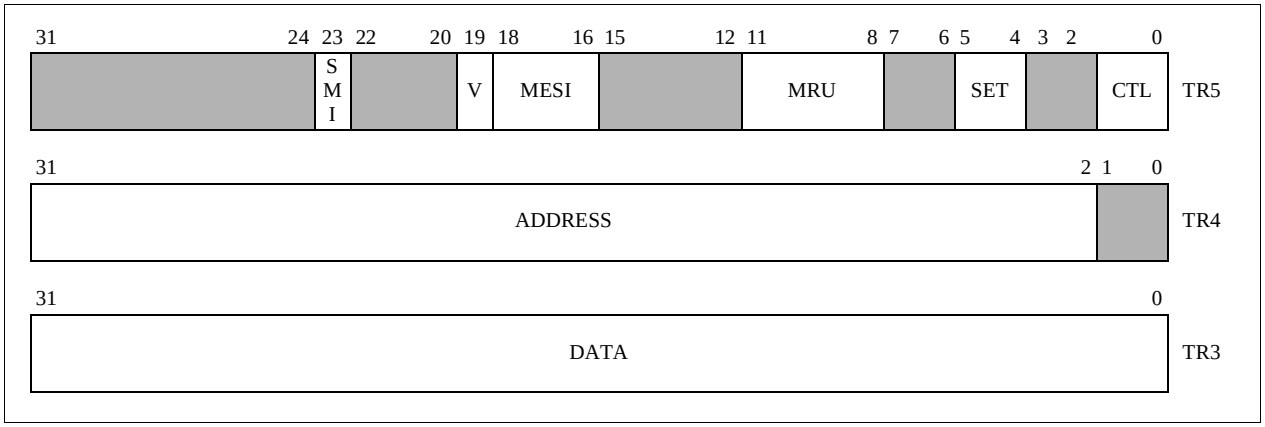


Figure 2-33. Cache Test Registers

Table 2-30. Cache Test Register Bit Definitions

REGISTER NAME	FIELD NAME	RANGE	DESCRIPTION
TR5	SMI	23	SMI Address Bit. Selects separate/cacheable SMI code/data space
	V, MESI	19 - 16	Valid, MESI Bits* If = 1000, Modified If = 1001, Shared If = 1010, Exclusive If = 0011, Invalid If = 1100, Locked Valid If = 0111, Locked Invalid Else = Undefined
	MRU	11 - 8	Used to determine the Least Recently Used (LRU) line.
	SET	5 - 4	Cache Set. Selects one of four cache sets to perform operation on.
	CTL	1 - 0	Control field If = 00: flush cache without invalidate If = 01: write cache If = 10: read cache If = 11: no cache or test register modification
TR4	ADDRESS	31 - 2	Physical Address
TR3	DATA	31 - 0	Data written or read during a cache test.

*Note: All 32 bytes should contain valid data before a line is marked as valid.

Write Operations. During a write, the TR3 DATA (32-bits) and TAG field information is written to the address selected by the ADDRESS field in TR4 and the SET field in TR5.

Read Operations. During a read, the cache address selected by the ADDRESS field in TR4 and the SET field in TR5. The TVB, MESI and MRU fields in TR5 are updated with the information from the selected line. TR3 holds the selected read data.

Cache Flushing. A cache flush occurs during a TR5 write if the CTL field is set to zero. During flushing, the CPU's cache controller reads through all the lines in the cache. "Modified" lines are redefined as "shared" by setting the shared MESI bit. Clean lines are left in their original state.

2.13.2 Scratch Pad RAM Locking

A Scratch Pad Ram is a private area of memory that can be assigned within the IBM 6x86MX unified cache. The Scratch Pad RAM is read/writable and is NOT kept coherent with the rest of the system.

Scratch Pad RAM may be implemented differently on different processors. On the IBM 6x86MX, the Scratch Pad RAM may be assigned on a cache line granularity.

RDMSR and WRMSR instructions with indices 03h to 05h are used to assign scratch pad memory. These instructions access the cache test registers. See section 2.13.1.1 (Page 2-58) for detailed description of cache test register operation. The cache line is assigned into Scratch Pad RAM by setting its MESI state to “locked valid.”

When locking physical addresses into the cache (Table 2-31), the programmer should be aware of several issues:

1) Locking all sets of the cache should not be done. It is required that one set always be available for general purpose caching. 2) Care must be taken by the programmer not to create synonyms. This is done by first checking to see if a particular address is locked before attempting to lock the address. If synonyms are created, IBM 6x86MX operation will be undefined.

When ever possible, it is recommended to flush the cache before assigning locked memory areas. Locked areas of the cache are cleared on reset, and are unaffected by warm reset and FLUSH#, or the INVD and WBINVD instructions.

Table 2-31. Cache Locking Operations

Read/Write	ECX	EDX	EAX	Operation
Read/Write	03h	----	Data to be read or written from/to the cache.	Loads or stores data to/from TR3.
Write	04h	----	32 bits of address	Address in EAX is loaded into TR4. This address is the cache line address that will be locked.
Read	04h	----	32 bits of address	Stores the contents of TR4 in EAX
Write	05h	----	Data to be written into TR5	Performs operation specified in CTL field of TR5.
Read	05h	----	Data in TR5 register	Reads data in TR5 and stores in EAX.

2.14 Interrupts and Exceptions

The processing of an interrupt or an exception changes the normal sequential flow of a program by transferring program control to a selected service routine. Except for SMM interrupts, the location of the selected service routine is determined by one of the interrupt vectors stored in the interrupt descriptor table.

Hardware interrupts are generated by signal sources external to the CPU. All exceptions (including so-called software interrupts) are produced internally by the CPU.

2.14.1 Interrupts

External events can interrupt normal program execution by using one of the three interrupt pins on the IBM 6x86MX CPU.

- Non-maskable Interrupt (NMI pin)
- Maskable Interrupt (INTR pin)
- SMM Interrupt (SMI# pin).

For most interrupts, program transfer to the interrupt routine occurs after the current instruction has been completed. When the execution returns to the original program, it begins immediately following the last completed instruction.

With the exception of string operations, interrupts are acknowledged between instructions. Long string operations have interrupt windows between memory moves that allow interrupts to be acknowledged.

The **NMI interrupt** cannot be masked by software and always uses interrupt vector 2 to locate its service routine. Since the interrupt vector is fixed and is supplied internally, no interrupt acknowledge bus cycles are performed. This interrupt is normally reserved for unusual situations such as parity errors and has priority over INTR interrupts.

Once NMI processing has started, no additional NMIs are processed until an IRET instruction is executed, typically at the end of the NMI service routine. If NMI is re-asserted prior to execution of the IRET instruction, one

and only one NMI rising edge is stored and processed after execution of the next IRET. During the NMI service routine, maskable interrupts may be enabled (unmasked). If an unmasked INTR occurs during the NMI service routine, the INTR is serviced and execution returns to the NMI service routine following the next IRET. If a HALT instruction is executed within the NMI service routine, the IBM 6x86MX CPU restarts execution only in response to RESET, an unmasked INTR or an SMM interrupt. NMI does not restart CPU execution under this condition.

The **INTR interrupt** is unmasked when the Interrupt Enable Flag (IF) in the EFLAGS register is set to 1. When an INTR interrupt occurs, the CPU performs two locked interrupt acknowledge bus cycles. During the second cycle, the CPU reads an 8-bit vector that is supplied by an external interrupt controller. This vector selects one of the 256 possible interrupt handlers which will be executed in response to the interrupt.

The **SMM interrupt** has higher priority than either INTR or NMI. After SMI# is asserted, program execution is passed to an SMI service routine that runs in SMM address space reserved for this purpose. The remainder of this section does not apply to the SMM interrupts. SMM interrupts are described in greater detail later in this chapter.

2.14.2 Exceptions

Exceptions are generated by an interrupt instruction or a program error. Exceptions are classified as traps, faults or aborts depending on the mechanism used to report them and the restartability of the instruction that first caused the exception.

A **Trap Exception** is reported immediately following the instruction that generated the trap exception. Trap exceptions are generated by execution of a software interrupt instruction (INTO, INT 3, INT n, BOUND), by a single-step operation or by a data breakpoint.

Software interrupts can be used to simulate hardware interrupts. For example, an INT n instruction causes the processor to execute the interrupt service routine pointed to by the nth vector in the interrupt table. Execution of the interrupt service routine occurs regardless of the state of the IF flag in the EFLAGS register.

The one byte INT 3, or breakpoint interrupt (vector 3), is a particular case of the INT n instruction. By inserting this one byte instruction in a program, the user can set breakpoints in the code that can be used during debug.

Single-step operation is enabled by setting the TF bit in the EFLAGS register. When TF is set, the CPU generates a debug exception (vector 1) after the execution of every instruction. Data breakpoints also generate a debug exception and are specified by loading the debug registers (DR0-DR7) with the appropriate values.

A Fault Exception is reported prior to completion of the instruction that generated the exception. By reporting the fault prior to instruction completion, the CPU is left in a state that allows the instruction to be restarted and the effects of the faulting instruction to be nullified. Fault exceptions include divide-by-zero errors, invalid opcodes, page faults and coprocessor errors. Instruction breakpoints (vector 1) are also handled as faults. After execution of the fault service routine, the instruction pointer points to the instruction that caused the fault.

An Abort Exception is a type of fault exception that is severe enough that the CPU cannot restart the program at the faulting instruction. The double fault (vector 8) is the only abort exception that occurs on the IBM 6x86MX CPU.

2.14.3 Interrupt Vectors

When the CPU services an interrupt or exception, the current program's FLAGS, code segment and instruction pointer are pushed onto the stack to allow resumption of execution of the interrupted program. In protected mode, the processor also saves an error code for some exceptions. Program control is then transferred to the interrupt handler (also called the interrupt service routine). Upon execution of an IRET at the end of the service routine, program execution resumes by popping from the stack, the instruction pointer, code segment, and FLAGS.

Interrupt Vector Assignments

Each interrupt (except SMI#) and exception is assigned one of 256 interrupt vector numbers Table 2-32, (Page 2-65). The first 32 interrupt vector assignments are defined or reserved. INT instructions acting as software interrupts may use any of the interrupt vectors, 0 through 255.

Table 2-32. Interrupt Vector Assignments

INTERRUPT VECTOR	FUNCTION	EXCEPTION TYPE
0	Divide error	FAULT
1	Debug exception	TRAP/FAULT*
2	NMI interrupt	
3	Breakpoint	TRAP
4	Interrupt on overflow	TRAP
5	BOUND range exceeded	FAULT
6	Invalid opcode	FAULT
7	Device not available	FAULT
8	Double fault	ABORT
9	Reserved	
10	Invalid TSS	FAULT
11	Segment not present	FAULT
12	Stack fault	FAULT
13	General protection fault	TRAP/FAULT
14	Page fault	FAULT
15	Reserved	
16	FPU error	FAULT
17	Alignment check exception	FAULT
18-31	Reserved	
32-255	Maskable hardware interrupts	TRAP
0-255	Programmed interrupt	TRAP

*Note: Data breakpoints and single-steps are traps. All other debug exceptions are faults.

In response to a maskable hardware interrupt (INTR), the IBM 6x86MX CPU issues interrupt acknowledge bus cycles to read the vector number from external hardware. These vectors should be in the range 32 - 255 as vectors 0 - 31 are reserved.

Interrupt Descriptor Table

The interrupt vector number is used by the IBM 6x86MX CPU to locate an entry in the interrupt descriptor table (IDT). In real mode, each IDT entry consists of a four-byte far pointer to the beginning of the corresponding interrupt service routine. In protected mode, each IDT entry is an eight-byte descriptor. The Interrupt Descriptor Table Register (IDTR) specifies the beginning address and limit of the IDT. Following reset, the IDTR contains a base address of 0h with a limit of 3FFh.

The IDT can be located anywhere in physical memory as determined by the IDTR register. The IDT may contain different types of descriptors: interrupt gates, trap gates and task gates. Interrupt gates are used primarily to enter a hardware interrupt handler. Trap gates are generally used to enter an exception handler or software interrupt handler. If an interrupt gate is used, the Interrupt Enable Flag (IF) in the EFLAGS register is cleared before the interrupt handler is entered. Task gates are used to make the transition to a new task.

2.14.4 Interrupt and Exception Priorities

As the IBM 6x86MX CPU executes instructions, it follows a consistent policy for prioritizing exceptions and hardware interrupts. The priorities for competing interrupts and exceptions are listed in Table 2-33 (Page 2-67). Debug traps for the previous instruction and the next instructions always take precedence. SMM interrupts are the next priority. When NMI and maskable INTR interrupts are both detected at the same instruction boundary, the IBM 6x86MX microprocessor services the NMI interrupt first.

The IBM 6x86MX CPU checks for exceptions in parallel with instruction decoding and execution. Several exceptions can result from a single instruction. However, only one exception is generated upon each attempt to execute the instruction. Each exception service routine should make the appropriate corrections to the instruction and then restart the instruction. In this way, exceptions can be serviced until the instruction executes properly.

The IBM 6x86MX CPU supports instruction restart after all faults, except when an instruction causes a task switch to a task whose task state segment (TSS) is partially not present. A TSS can be partially not present if the TSS is not page aligned and one of the pages where the TSS resides is not currently in memory.

Table 2-33. Interrupt and Exception Priorities

PRIORITY	DESCRIPTION	NOTES
0	Warm Reset	Caused by the assertion of WM_RST.
1	Debug traps and faults from previous instruction.	Includes single-step trap and data breakpoints specified in the debug registers.
2	Debug traps for next instruction.	Includes instruction execution breakpoints specified in the debug registers.
3	Hardware Cache Flush	Caused by the assertion of FLUSH#.
4	SMM hardware interrupt.	SMM interrupts are caused by SMI# asserted and always have highest priority.
5	Non-maskable hardware interrupt.	Caused by NMI asserted.
6	Maskable hardware interrupt.	Caused by INTR asserted and IF = 1.
7	Faults resulting from fetching the next instruction.	Includes segment not present, general protection fault and page fault.
8	Faults resulting from instruction decoding.	Includes illegal opcode, instruction too long, or privilege violation.
9	WAIT instruction and TS = 1 and MP = 1.	Device not available exception generated.
10	ESC instruction and EM = 1 or TS = 1.	Device not available exception generated.
11	Floating point error exception.	Caused by unmasked floating point exception with NE = 1.
12	Segmentation faults (for each memory reference required by the instruction) that prevent transferring the entire memory operand.	Includes segment not present, stack fault, and general protection fault.
13	Page Faults that prevent transferring the entire memory operand.	
14	Alignment check fault.	

2.14.5 Exceptions in Real Mode

Many of the exceptions described in Table 2-33 (Page 2-67) are not applicable in real mode. Exceptions 10, 11, and 14 do not occur in real mode. Other exceptions have slightly different meanings in real mode as listed in Table 2-34.

Table 2-34. Exception Changes in Real Mode

VECTOR NUMBER	PROTECTED MODE FUNCTION	REAL MODE FUNCTION
8	Double fault.	Interrupt table limit overrun.
10	Invalid TSS.	x
11	Segment not present.	x
12	Stack fault.	SS segment limit overrun.
13	General protection fault.	CS, DS, ES, FS, GS segment limit overrun.
14	Page fault.	x

Note: x = does not occur

2.14.6 Error Codes

When operating in protected mode, the following exceptions generate a 16-bit error code:

Double Fault	Invalid TSS
Alignment Check	Segment Not Present
Page Fault	Stack Fault
	General Protection Fault

The error code is pushed onto the stack prior to entering the exception handler. The error code format is shown in Figure 2-34 and the error code bit definitions are listed in Table 2-35. Bits 15-3 (selector index) are not meaningful if the error code was generated as the result of a page fault. The error code is always zero for double faults and alignment check exceptions.

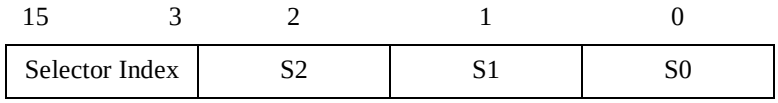


Figure 2-34. Error Code Format

Table 2-35. Error Code Bit Definitions

FAULT TYPE	SELECTOR INDEX (BITS 15-3)	S2 (BIT 2)	S1 (BIT 1)	S0 (BIT 0)
Double Fault or Alignment Check	0	0	0	0
Page Fault	Reserved.	Fault caused by: 0 = not present page 1 = page-level protection violation.	Fault occurred during: 0 = read access 1 = write access.	Fault occurred during: 0 = supervisor access. 1 = user access.
IDT Fault	Index of faulty IDT selector.	Reserved.	1	If = 1, exception occurred while trying to invoke exception or hardware interrupt handler.
Segment Fault	Index of faulty selector.	TI bit of faulty selector.	0	If =1, exception occurred while trying to invoke exception or hardware interrupt handler.

2.15 System Management Mode

System Management Mode (SMM) is a distinct CPU mode that differs from normal CPU x86 operating modes (real mode, V86 mode, and protected mode) and is most often used to perform power management.

The IBM 6x86MX is backward compatible with the SL-compatible SMM found on previous IBM and Cyrix microprocessors. On the IBM 6x86MX SMM has been enhanced to optimized software emulation of multimedia and I/O peripherals.

The Cyrix Enhanced SMM provides new features:

- Cacheability of SMM memory
- Support for nesting of multiple SMIs
- Improved SMM entry and exit time.

Overall Operation

The overall operation of a SMM operation is shown in (Figure 2-35). SMM is entered using the System Management Interrupt (SMI) pin. SMI interrupts have higher priority than any other interrupt, including NMI interrupts. SMM can also be entered using software by using an SMINT instruction.

Upon entering SMM mode, portions of the CPU state are automatically saved in the SMM address memory space header. The CPU enters real mode and begins executing the SMI service routine in SMM address space.

Execution of a SMM routine starts at the base address in SMM memory address space. Since the SMM routines reside in SMM memory space, SMM routines can be made totally transparent to all software, including protected-mode operating systems.

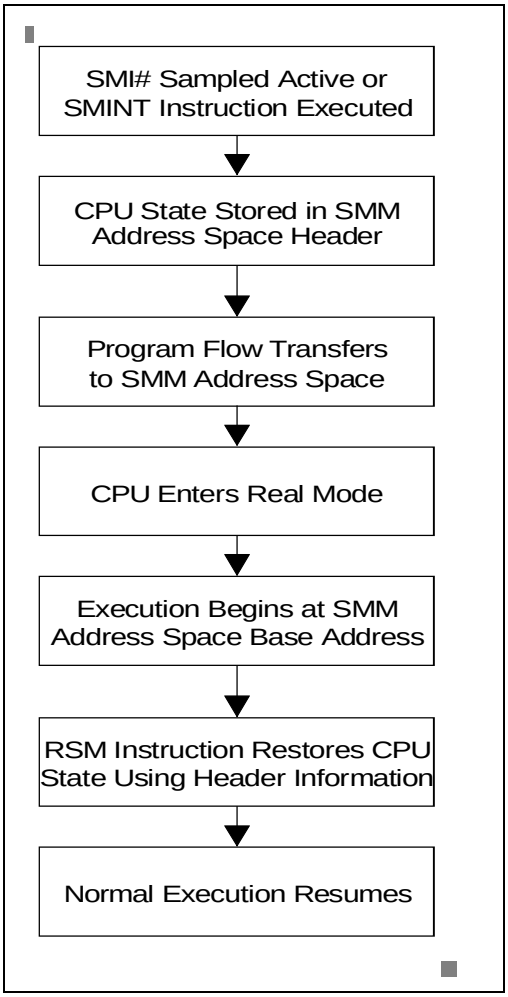


Figure 2-35. SMI Execution Flow Diagram

2.15.1 SMM Memory Space

SMM memory must reside within the bounds of physical memory and not overlap with system memory. SMM memory space (Figure 2-36) is defined by setting the SM3 bit in CCR1 and specifying the base address and size of the SMM memory space in the ARR3 register.

The base address must be a multiple of the SMM memory space size. For example, a 32 KByte SMM memory space must be located on a 32 KByte address boundary. The memory space size can range from 4 KBytes to 4 GBytes SMM accesses ignore the state of the A20M# input pin and drive the A20 address bit to the unmasked value.

SMM memory space can be accessed while in normal mode by setting the SMAC bit in the CCR1 register. This feature may be used to initialize SMM memory space.

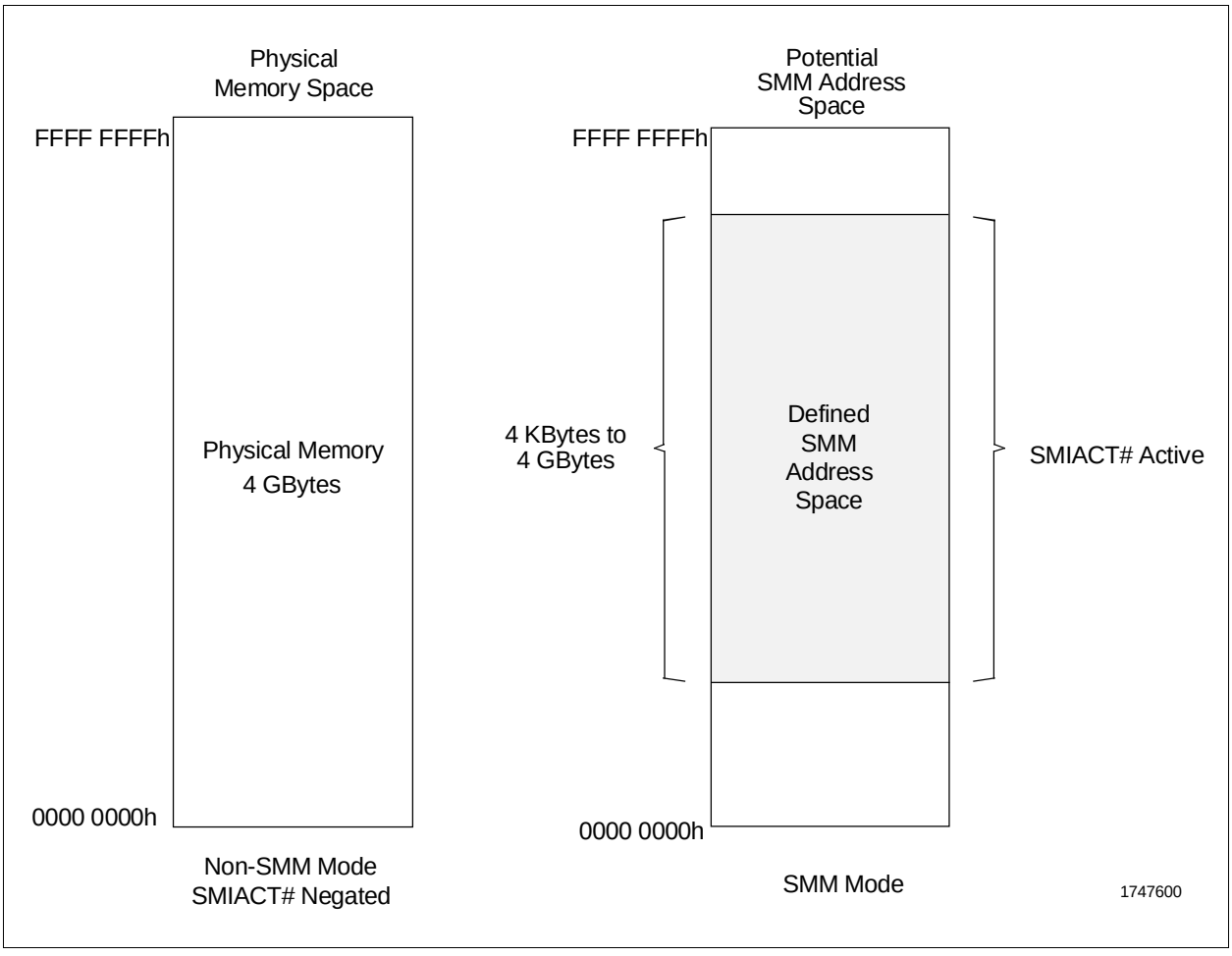


Figure 2-36. System Management Memory Space

2.15.2 SMM Memory Space Header

The SMM Memory Space Header (Figure 2-37) is used to store the CPU state prior to starting an SMM routine. The fields in this header are described in Table 2-36 (Page 2-73). After the SMM routine has completed, the header information is used to restore the original CPU state. The location of the SMM header is determined by the SMM Header Address Register (SMHR).

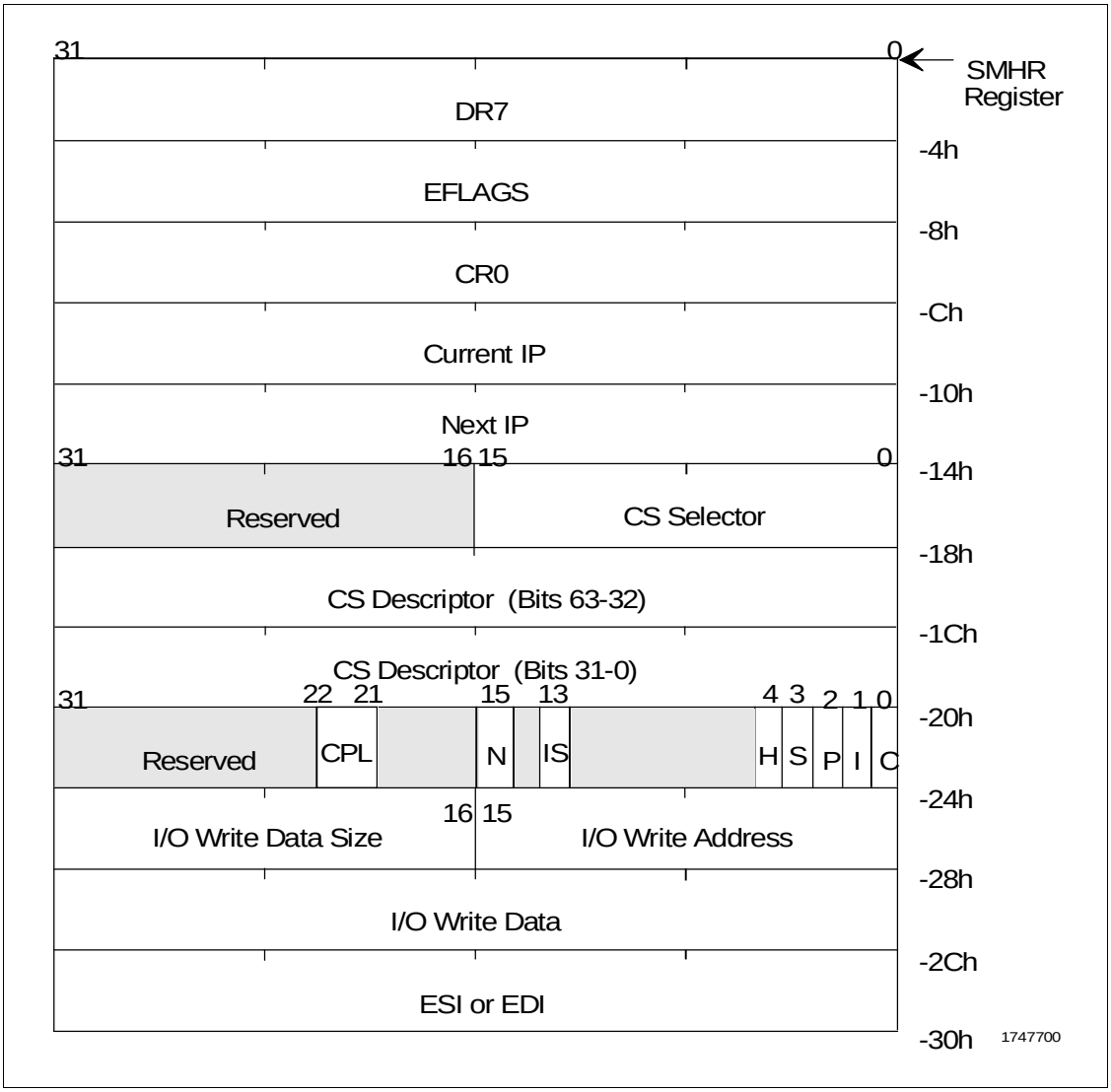


Figure 2-37. SMM Memory Space Header

Table 2-36. SMM Memory Space Header

NAME	DESCRIPTION	SIZE
DR7	The contents of Debug Register 7.	4 Bytes
EFLAGS	The contents of Extended Flags Register.	4 Bytes
CR0	The contents of Control Register 0.	4 Bytes
Current IP	The address of the instruction executed prior to servicing SMI interrupt.	4 Bytes
Next IP	The address of the next instruction that will be executed after exiting SMM mode.	4 Bytes
CS Selector	Code segment register selector for the current code segment.	2 Bytes
CS Descriptor	Code segment register descriptor for the current code segment.	8 Bytes
CPL	Current privilege level for current code segment.	2 Bits
N	Nested SMI Indicator If N = 1: current SMM is being serviced from within SMM mode. If N = 0: current SMM is not being serviced from within SMM mode.	1 Bit
IS	Internal SMI Indicator If IS =1: current SMM is the result of an internal SMI event. If IS =0: current SMM is the result of an external SMI event.	1 Bit
H	SMI during CPU HALT state indicator If H = 1: the processor was in a halt or shutdown prior to servicing the SMM interrupt.	1 Bit
S	Software SMM Entry Indicator. If S = 1: current SMM is the result of an SMINT instruction. If S = 0: current SMM is not the result of an SMINT instruction.	1 Bit
P	REP INSx/OUTSx Indicator If P = 1: current instruction has a REP prefix. If P = 0: current instruction does not have a REP prefix.	1 Bit
I	IN, INSx, OUT, or OUTSx Indicator If I = 1: if current instruction performed is an I/O WRITE. If I = 0: if current instruction performed is an I/O READ.	1 Bit
C	Code Segment writable Indicator If C = 1: the current code segment is writable. If C = 0: the current code segment is not writable.	1 Bit
I/O	Indicates size of data for the trapped I/O write: 01h = byte 03h = word 0Fh = dword	2 Bytes
I/O Write Address	I/O Write Address Processor port used for the trapped I/O write.	2 Bytes
I/O Write Data	I/O Write Data Data associated with the trapped I/O write.	4 Bytes
ESI or EDI	Restored ESI or EDI value. Used when it is necessary to repeat a REP OUTSx or REP INSx instruction when one of the I/O cycles caused an SMI# trap.	4 Bytes

Note: INSx = INS, INSB, INSW or INSD instruction.
Note: OUTSx = OUTS, OUTSB, OUTSW and OUTSD instruction.

Current and Next IP Pointers

Included in the header information are the Current and Next IP pointers. The Current IP points to the instruction executing when the SMI was detected and the Next IP points to the instruction that will be executed after exiting SMM.

Normally after an SMM routine is completed, the instruction flow begins at the Next IP address. However, if an I/O trap has occurred, instruction flow should return to the Current IP to complete the I/O instruction.

If SMM has been entered due to an I/O trap for a REP INSx or REP OUTSx instruction, the Current IP and Next IP fields contain the same address.

If an entry into SMM mode was caused by an I/O trap, the port address, data size and data value associated with that I/O operation are stored in the SMM header. Note that these values are only valid for I/O operations. The I/O data is not restored within the CPU when executing a RSM instruction.

Under these circumstances the I and P bits, as well as ESI/EDI field, contain valid information.

Also saved are the contents of debug register 7 (DR7), the extended flags register (EFLAGS), and control register 0 (CR0).

If the S bit in the SMM header is set, the SMM entry resulted from an SMINT instruction.

SMM Header Address Pointer

The SMM Header Address Pointer Register (SMHR) (Figure 2-38) contains the 32-bit SMM Header pointer. The SMHR address is dword aligned, so the two least significant bits are ignored.

The SMHR valid bit (bit 0) is cleared with every write to ARR3 and during a hardware RESET. Upon entry to SMM, the SMHR valid bit is examined before the CPU state is saved into the SMM memory space header. When the valid bit is reset, the SMM header pointer will be calculated (ARR3 base field + ARR3 size field) and loaded into the SMHR and the valid bit will be set.

If the desired SMM header location is different than the top of SMM memory space, as may be the case when nesting SMI's, then the SMHR register must be loaded with a new value and valid bit from within the SMI routine before nesting is enabled.

The SMM memory space header can be relocated using the new RDSHR and WRSHR instructions.

Figure 2-38. SMHR Register

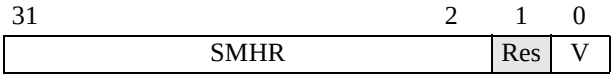


Table 2-37. SMHR Register Bits

BIT POSITION	DESCRPTION
31 - 2	SMHR header pointer address.
1	Reserved
0	Valid Bit

2.15.3 SMM Instructions

After entering the SMI service routine, the MOV, SVDC, SVLDT and SVTS instructions (Table 2-38) can be used to save the complete CPU state information. If the SMI service routine modifies more than what is automatically saved or forces the CPU to power down, the complete CPU state information must be saved. Since the CPU is a static device, its internal state is retained when the input clock is stopped. Therefore, an entire CPU state save is not necessary prior to stopping the input clock.

Table 2-38. SMM Instruction Set

INSTRUCTION	OPCODE	FORMAT	DESCRIPTION
SVDC	0F 78 [mod sreg3 r/m]	SVDC mem80, sreg3	*Save Segment Register and Descriptor* Saves reg (DS, ES, FS, GS, or SS) to mem80.
RSDC	0F 79 [mod sreg3 r/m]	RSDC sreg3, mem80	*Restore Segment Register and Descriptor* Restores reg (DS, ES, FS, GS, or SS) from mem80. Use RSM to restore CS. Note: Processing "RSDC CS, Mem80" will produce an exception.
SVLDT	0F 7A [mod 000 r/m]	SVLDT mem80	*Save LDTR and Descriptor* Saves Local Descriptor Table (LDTR) to mem80.
RSLDT	0F 7B [mod 000 r/m]	RSLDT mem80	*Restore LDTR and Descriptor* Restores Local Descriptor Table (LDTR) from mem80.
SVTS	0F 7C [mod 000 r/m]	SVTS mem80	*Save TSR and Descriptor* Saves Task State Register (TSR) to mem80.
RSTS	0F 7D [mod 000 r/m]	RSTS mem80	*Restore TSR and Descriptor* Restores Task State Register (TSR) from mem80.
SMINT	0F 7E	SMINT	*Software SMM Entry* CPU enters SMM mode. CPU state information is saved in SMM memory space header and execution begins at SMM base address.
RSM	0F AA	RSM	*Resume Normal Mode* Exits SMM mode. The CPU state is restored using the SMM memory space header and execution resumes at interrupted point.
RDSHR	0F 36	RDSHR ereg/mem32	*Read SMM Header Pointer Register* Saves SMM header pointer to extended register or memory.
WRSHR	0F 37	WRSHR ereg/mem32	*Write SMM Header Pointer Register* Load SMM header pointer register from extended register or memory.

Note: mem32 = 32-bit memory location
mem80 = 80-bit memory location

The SMM instructions listed in Table 2-38, (except the SMINT instruction) can be executed only if:

1) ARR3 Size > 0
2) Current Privilege Level =0
3) SMAC bit is set or the CPU is executing an SMI service routine.
4) USE_SMI (CCR1- bit 1) = 1
5) SM3 (CCR1-bit 7) = 1

If the above conditions are not met and an attempt is made to execute an SVDC, RSDC, SVLDT, RSLDT, SVTS, RSTS, SMINT, RSM, RDSHR, or WDSHR instruction, an invalid opcode exception is generated. These instructions can be executed outside of defined SMM space provided the above conditions are met.

The SMINT instruction allows software entry into SMM. The SVDC, RSDC, SVLDT, RSLDT, SVTS and RSTS instructions save or restore 80 bits of data, allowing the saved values to include the hidden portion of the register contents.

The WRSHR instruction loads the contents of either a 32-bit memory operand or a 32-bit register operand into the SMHR pointer register based on the value of the mod r/m instruction byte. Likewise the RDSHR instruction stores the contents of the SMHR pointer register to either a 32 bit memory operand or a 32 bit register operand based on the value of the mod r/m instruction byte.

2.15.4 SMM Operation

This section details the SMM operations.

Entering SMM

Entering SMM requires the assertion of the SMI# pin or execution of an SMINT instruction. SMI interrupts have higher priority than any interrupt including NMI interrupts.

For the SMI# or SMINT instruction to be recognized, the following configuration register bits must be set as shown in Table 2-39.

Table 2-39. Requirements for Recognizing SMI# and SMINT

REGISTER (Bit)		SMI#	SMINT
SMI	CCR1 (1)	1	1
SMAC	CCR1 (2)	0	1
ARR3	SIZE (3-0)	> 0	> 0
SM3	CCR1 (7)	1	1

Upon entry into SMM, after the SMM header has been saved, the CR0, EFLAGS, and DR7 registers are set to their reset values. The Code Segment (CS) register is loaded with the base, as defined by the ARR3 register, and a limit of 4 GBytes. The SMI service routine then begins execution at the SMM base address in real mode.

Saving the CPU State

The programmer must save the value of any registers that may be changed by the SMI service routine. For data accesses immediately after entering the SMI service routine, the programmer must use CS as a segment override. I/O port access is possible during the routine but care must be taken to save registers modified by the I/O instructions. Before using a segment register, the register and the register's descriptor cache contents should be saved using the SVDC instruction. While executing in the SMM space, execution flow can transfer to normal memory locations.

Program Execution

Hardware interrupts, (INTRs and NMIs), may be serviced during a SMI service routine. If interrupts are to be serviced while executing in the SMM memory space, the SMM memory space must be within the 0 to 1 MByte address range to guarantee proper return to the SMI service routine after handling the interrupt.

INTRs are automatically disabled when entering SMM since the IF flag is set to its reset value. Once in SMM, the INTR can be enabled by setting the IF flag. NMI is also automatically disable when entering SMM. Once in SMM, NMI can be enabled by setting NMI_EN in CCR3. If NMI is not enabled, the CPU latches one NMI event and services the interrupt after NMI has been enabled or after exiting SMM through the RSM instruction.

Within the SMI service routine, protected mode may be entered and exited as required, and real or protected mode device drivers may be called.

Exiting SMM

To exit the SMI service routine, a Resume (RSM) instruction, rather than an IRET, is executed. The RSM instruction causes the IBM 6x86MX processor to restore the CPU state using the SMM header information and resume execution at the interrupted point. If the full CPU state was saved by the programmer, the stored values should be reloaded prior to executing the RSM instruction using the MOV, RSDC, RSLDT and RSTS instructions.

When the RSM instruction is executed at the end of the SMI handler, the EIP instruction pointer is automatically read from the NEXT IP field in the SMM header.

When restarting I/O instructions, the value of NEXT IP may need modification. Before executing the RSM instruction, use a MOV instruction to move the CURRENT IP value to the NEXT IP location as the CURRENT IP value is valid if an I/O instruction was executing when the SMI interrupt occurred. Execution is then returned to the I/O instruction, rather than to the instruction after the I/O instruction.

A set H bit in the SMM header indicates that a HLT instruction was being executed when the SMI occurred. To resume execution of the HLT instruction, the NEXT IP field in the SMM header should be decremented by one before executing RSM instruction.

2.15.5 SL and Cyrix SMM Operating Modes

There are two SMM modes, SL-compatible mode (default) and Cyrix SMM mode.

2.15.5.1 SL-Compatible SMM Mode

While in SL-compatible mode, SMM memory space accesses can only occur during an SMI service routine. While executing an SMI service routine SMIACT# remains asserted regardless of the address being accessed. This includes the time when the SMI service routine accesses memory outside the defined SMM memory space.

SMM memory caching is not supported in SL-compatible SMM mode. If a cache inquiry cycle occurs while SMIACT# is active, any resulting write-back cycle is issued with SMIACT# asserted. This occurs even though the write-back cycle is intended for normal memory rather than SMM memory. To avoid this problem it is recommended that the internal caches be flushed prior to servicing an SMI event. Of course in write-back mode this could add an indeterminate delay to servicing of SMI.

An interrupt on the SMI# input pin has higher priority than the NMI input. The SMI# input pin is falling edge sensitive and is sampled on every rising edge of the processor input clock.

Asserting SMI# forces the processor to save the CPU state to memory defined by SMHR register and to begin execution of the SMI service routine at the beginning of the defined SMM memory space. After the processor internally acknowledges the SMI# interrupt, the SMIACT# output is driven low for the duration of the interrupt service routine.

When the RSM instruction is executed, the CPU negates the SMIACT# pin after the last bus cycle to SMM memory. While executing the SMM service routine, one additional SMI# can be latched for service after resuming from the first SMI.

During RESET, the USE_SMI bit in CCR1 is cleared. While USE_SMI is zero, SMACT# is always negated. SMIACT# does not float during bus hold states.

2.15.5.2 Cyrix Enhanced SMM Mode

The Cyrix SMM Mode is enabled when bit 0 in the CCR6 (SMM_MODE) is set. Only in Cyrix enhanced SMM mode can:

- SMM memory be cached
- SMM interrupts be nested

Pin Interface

The SMI# and SMIACT# pins behave differently in Cyrix Enhanced SMM mode.

In Cyrix Enhanced SMM mode SMI# is level sensitive. As a level sensitive signal software can process SMI interrupts until all sources in the chipset have been cleared.

While operating in this mode, SMIACT# output is not used to indicate that the CPU is operating in SMM mode. This is left to the SMM driver.

In Cyrix enhanced SMM, SMIACT# is asserted for every SMM memory bus cycle and is de-asserted for every non-SMM bus cycle. In this mode the SMIACT# pin meets the timing of D/C# and W/R#.

During RESET, the USE_SMI bit in CCR1 is cleared. While USE_SMI is zero, SMIACT# is always negated. SMIACT# does float during bus hold states.

Cacheability of SMM Space

In SL-compatible SMM mode, caching is not available, but in Cyrix SMM mode, both code and data caching is supported. In order to cache SMM data and avoid coherency issues the processor assumes no overlap of main memory with SMM memory. This implies that a section of main memory must be dedicated for SMM.

The on-chip cache sets a special ID bit in the cache tag block for each line that contains SMM code data. This ID bit is then used by the bus controller to regulate assertion of the SMIACT# pin for write-back of any SMM data.

Nested SMI

Only in the Cyrix enhanced SMM mode is nesting of SMI interrupts supported. This is important to allow high priority events such as audio emulation to interrupt lower priority SMI code. In the case of nesting, it is up to the SMM driver to determine which SMM event is being serviced, which to prioritize, and perform all SMM interrupt control functions.

Software enables and disables SMI interrupts while in SMM mode by setting and clearing the nest-enable bit (N bit, bit 6 of CCR6). By default the CPU automatically disables SMI interrupts (clears the N bit) on entry to SMM mode, and re-enables them (sets the N bit) when exiting SMM mode (i.e., RSM). The SMI handler can optionally enable nesting to allow higher priority SMI interrupts to occur while handling the current SMI event.

The SMI handler is responsible for managing the SMHR pointer register when processing nested SMI interrupts. Before nested SMI's can be serviced the current SMM handler must save the contents of the SMHR pointer register and then load a new value into the SMHR register for use by a subsequent nested SMI event.

Prior to execution of a RSM instruction the contents of the old SMHR pointer register must be restored for proper operation to continue. Prior to restoring the contents of old SMHR pointer register one should disable additional SMI's. This should be done so that the CPU will not inadvertently receive and service an SMI event after the old SMHR contents have been restored but before the RSM instruction is executed.

2.15.6 Maintaining the FPU and MMX States

If power will be removed from the CPU or if the SMM routine will execute MMX or FPU instructions, then the MMX or FPU state should be maintained for the application running before SMM was entered. If the MMX or FPU state is to be saved and restored from within SMM, there are certain guidelines that must be followed to make SMM completely transparent to the application program.

The complete state of the FPU can be saved and restored with the FNSAVE and FNRSTOR instructions. FNSAVE is used instead of the FSAVE because FSAVE will wait for the FPU to check for existing error conditions before

storing the FPU state. If there is a unmasked FPU exception condition pending, the FSAVE instruction will wait until the exception condition is serviced. To maintain transparency for the application program, the SMM routine should not service this exception. If the FPU state is restored with the FNRSTOR instruction before returning to normal mode, the application program can correctly service the exception. FPU instructions can be executed within SMM once the FPU state has been saved.

The information saved with the FSAVE instruction varies depending on the operating mode of the CPU. To save and restore all FPU information, the 32-bit protected mode version of the FPU save and restore instruction should be used.

CPU States Related to SMM and Suspend Mode

The state diagram shown in Figure 2-39 (Page 2-81) illustrates the various CPU states associated with SMM and suspend mode. While in the SMI service routine, the 6x86MX CPU can enter suspend mode either by (1) executing a halt (HLT) instruction or (2) by asserting the SUSP# input.

During SMM operations and while in SUSP# initiated suspend mode, an occurrence of SMI#, NMI, or INTR is latched. (In order for INTR to be latched, the IF flag must be set.) The INTR or NMI is serviced after exiting suspend mode.

If suspend mode is entered via a HLT instruction from the operating system or application software, the reception of an SMI# interrupt causes the CPU to exit suspend mode and enter SMM.

2.16 Shutdown and Halt

The **Halt Instruction** (HLT) stops program execution and prevents the processor from using the local bus until restarted. The IBM 6x86MX CPU then issues a special Stop Grant bus cycle and enters a low-power suspend mode if the SUSP_HLT bit in CCR2 is set. SMI, NMI, INTR with interrupts enabled (IF bit in EFLAGS=1), WM_RST or RESET forces the CPU out of the halt state. If interrupted, the saved code segment and instruction pointer specify the instruction following the HLT.

Shutdown occurs when a severe error is detected that prevents further processing. An NMI input can bring the processor out of shutdown if the IDT limit is large enough to contain the NMI interrupt vector and the stack has enough room to contain the vector and flag information. Otherwise, shutdown can only be exited by a processor reset.

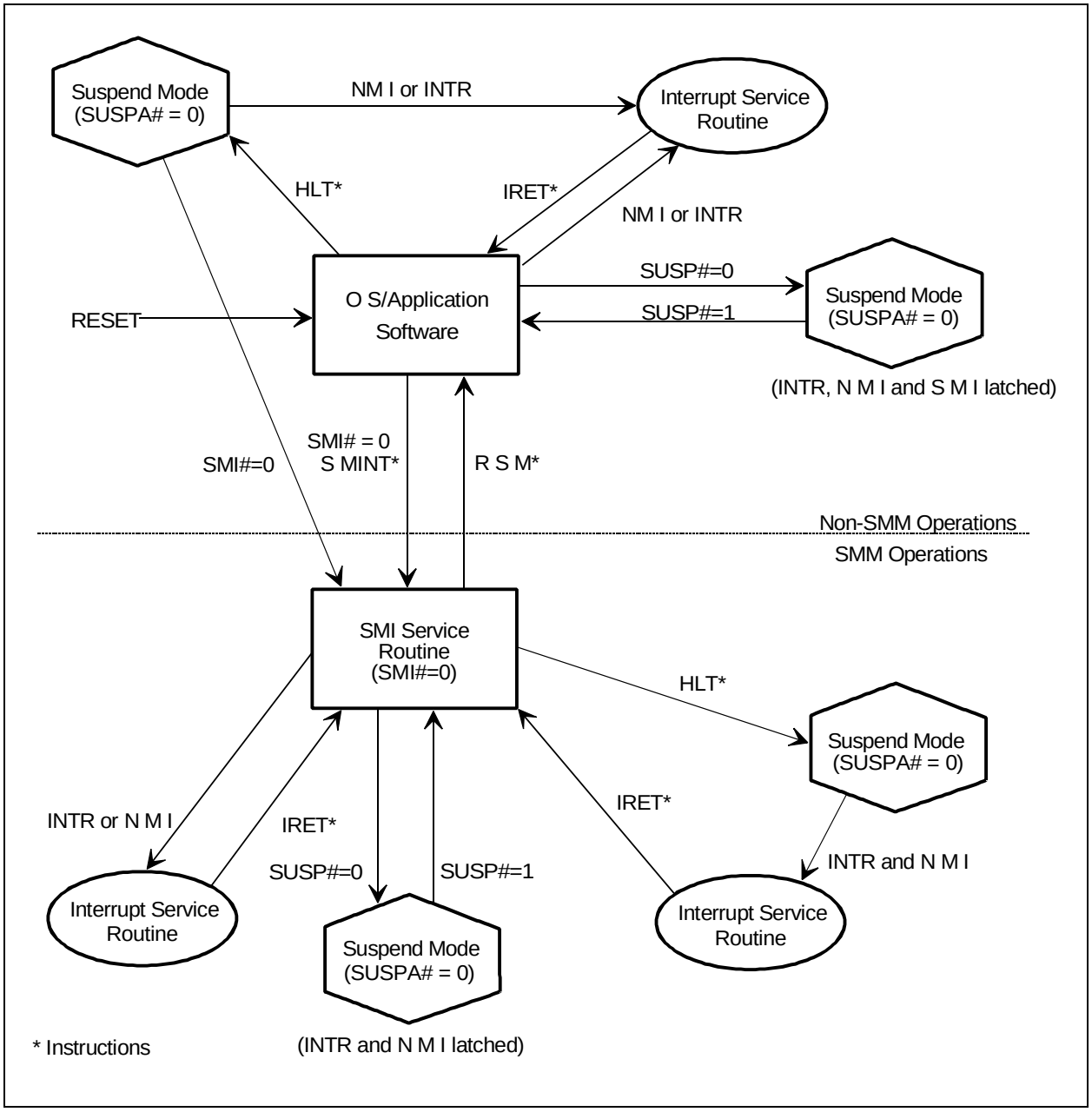


Figure 2-39. SMM and Suspend Mode State Diagram

2.17 Protection

Segment protection and page protection are safeguards built into the IBM 6x86MX CPU protected mode architecture which deny unauthorized or incorrect access to selected memory addresses. These safeguards allow multitasking programs to be isolated from each other and from the operating system. Page protection is discussed earlier in this chapter. This section concentrates on segment protection.

Selectors and descriptors are the key elements in the segment protection mechanism. The segment base address, size, and privilege level are established by a segment descriptor. Privilege levels control the use of privileged instructions, I/O instructions and access to segments and segment descriptors. Selectors are used to locate segment descriptors.

Segment accesses are divided into two basic types, those involving code segments (e.g., control transfers) and those involving data accesses. The ability of a task to access a segment depends on the:

- segment type
- instruction requesting access
- type of descriptor used to define the segment
- associated privilege levels (described below).

Data stored in a segment can be accessed only by code executing at the same or a more privileged level. A code segment or procedure can only be called by a task executing at the same or a less privileged level.

2.17.1 Privilege Levels

The values for privilege levels range between 0 and 3. Level 0 is the highest privilege level (most privileged), and level 3 is the lowest privilege level (least privileged). The privilege level in real mode is effectively 0.

The **Descriptor Privilege Level** (DPL) is the privilege level defined for a segment in the segment descriptor. The DPL field specifies the minimum privilege level needed to access the memory segment pointed to by the descriptor.

The **Current Privilege Level** (CPL) is defined as the current task's privilege level. The CPL of an executing task is stored in the hidden portion of the code segment register and essentially is the DPL for the current code segment.

The **Requested Privilege Level** (RPL) specifies a selector's privilege level and is used to distinguish between the privilege level of a routine actually accessing memory (the CPL), and the privilege level of the original requestor (the RPL) of the memory access. The lesser of the RPL and CPL is called the effective privilege level (EPL). Therefore, if RPL = 0 in a segment selector, the effective privilege level is always determined by the CPL. If RPL = 3, the effective privilege level is always 3 regardless of the CPL.

For a memory access to succeed, the effective privilege level (EPL) must be at least as privileged as the descriptor privilege level (EPL ≤ DPL). If the EPL is less privileged than the DPL (EPL > DPL), a general protection fault is generated. For example, if a segment has a DPL = 2, an instruction accessing the segment only succeeds if executed with an EPL ≤ 2.

2.17.2 I/O Privilege Levels

The I/O Privilege Level (IOPL) allows the operating system executing at CPL=0 to define the least privileged level at which IOPL-sensitive instructions can unconditionally be used. The IOPL-sensitive instructions include CLI, IN, OUT, INS, OUTS, REP INS, REP OUTS, and STI. Modification of the IF bit in the EFLAGS register is also sensitive to the I/O privilege level. The IOPL is stored in the EFLAGS register.

An I/O permission bit map is available as defined by the 32-bit Task State Segment (TSS). Since each task can have its own TSS, access to individual processor I/O ports can be granted through separate I/O permission bit maps.

If CPL $\leq$ IOPL, IOPL-sensitive operations can be performed. If CPL > IOPL, a general protection fault is generated if the current task is associated with a 16-bit TSS. If the current task is associated with a 32-bit TSS and CPL > IOPL, the CPU consults the I/O permission bitmap in the TSS to determine on a port-by-port basis whether or not I/O instructions (IN, OUT, INS, OUTS, REP INS, REP OUTS) are permitted, and the remaining IOPL-sensitive operations generate a general protection fault.

2.17.3 Privilege Level Transfers

A task's CPL can be changed only through intersegment control transfers using gates or task switches to a code segment with a different privilege level. Control transfers result from exception and interrupt servicing and from execution of the CALL, JMP, INT, IRET and RET instructions.

There are five types of control transfers that are summarized in Table 2-40 (Page 2-84). Control transfers can be made only when the operation causing the control transfer references the correct descriptor type. Any violation of these descriptor usage rules causes a general protection fault.

Any control transfer that changes the CPL within a task results in a change of stack. The initial values for the stack segment (SS) and stack pointer (ESP) for privilege levels 0, 1, and 2 are stored in the TSS. During a CALL control transfer, the SS and ESP are loaded with the new stack pointer and the previous stack pointer is saved on the new stack. When returning to the original privilege level, the RET or IRET instruction restores the less-privileged stack

Table 2-40. Descriptor Types Used for Control Transfer

TYPE OF CONTROL TRANSFER	OPERATION TYPES	DESCRIPTOR REFERENCED	DESCRIPTOR TABLE
Intersegment within the same privilege level.	JMP, CALL, RET, IRET*	Code Segment	GDT or LDT
Intersegment to the same or a more privileged level. Interrupt within task (could change CPL level).	CALL	Gate Call	GDT or LDT
	Interrupt Instruction, Exception, External Interrupt	Trap or Interrupt Gate	IDT
Intersegment to a less privileged level (changes task CPL).	RET, IRET*	Code Segment	GDT or LDT
Task Switch via TSS	CALL, JMP	Task State Segment	GDT
Task Switch via Task Gate	CALL, JMP	Task Gate	GDT or LDT
	IRET**, Interrupt Instruction, Exception, External Interrupt	Task Gate	IDT

* NT (Nested Task bit in EFLAGS) = 0
** NT (Nested Task bit in EFLAGS) = 1

Gates

Gate descriptors provide protection for privilege transfers among executable segments. Gates are used to transition to routines of the same or a more privileged level. Call gates, interrupt gates and trap gates are used for privilege transfers within a task. Task gates are used to transfer between tasks.

Gates conform to the standard rules of privilege. In other words, gates can be accessed by a task if the effective privilege level (EPL) is the same or more privileged than the gate descriptor's privilege level (DPL).

2.17.4 Initialization and Transition to Protected Mode

The IBM 6x86MX microprocessor switches to real mode immediately after RESET. While operating in real mode, the system tables and registers should be initialized. The GDTR and IDTR must point to a valid GDT and IDT, respectively. The GDT must contain descriptors which describe the initial code and data segments.

The processor can be placed in protected mode by setting the PE bit in the CR0 register. After enabling protected mode, the CS register should be loaded and the instruction decode queue should be flushed by executing an intersegment JMP. Finally, all data segment registers should be initialized with appropriate selector values.

2.18 Virtual 8086 Mode

Both real mode and virtual 8086 (V86) mode are supported by the IBM 6x86MX CPU allowing execution of 8086 application programs and 8086 operating systems. V86 mode allows the execution of 8086-type applications, yet still permits use of the IBM 6x86MX CPU paging mechanism. V86 tasks run at privilege level 3. When loaded, all segment limits are set to FFFFh (64K) as in real mode.

2.18.1 V86 Memory Addressing

While in V86 mode, segment registers are used in an identical fashion to real mode. The contents of the segment register are multiplied by 16 and added to the offset to form the segment base linear address. The IBM 6x86MX CPU permits the operating system to select which programs use the V86 address mechanism and which programs use protected mode addressing for each task.

The IBM 6x86MX CPU also permits the use of paging when operating in V86 mode. Using paging, the 1-MByte memory space of the V86 task can be mapped to anywhere in the 4-GByte linear memory space of the IBM 6x86MX CPU.

The paging hardware allows multiple V86 tasks to run concurrently, and provides protection and operating system isolation. The paging hardware must be enabled to run multiple V86 tasks or to relocate the address space of a V86 task to physical address space greater than 1 MByte.

2.18.2 V86 Protection

All V86 tasks operate with the least amount of privilege (level 3) and are subject to all of the IBM 6x86MX CPU protected mode protection checks. As a result, any attempt to execute a privileged instruction within a V86 task results in a general protection fault.

In V86 mode, a slightly different set of instructions are sensitive to the I/O privilege level (IOPL) than in protected mode. These instructions are: CLI, INT n, IRET, POPF, PUSHF, and STI. The INT3, INTO and BOUND variations of the INT instruction are not IOPL sensitive.

2.18.3 V86 Interrupt Handling

To fully support the emulation of an 8086-type machine, interrupts in V86 mode are handled as follows. When an interrupt or exception is serviced in V86 mode, program execution transfers to the interrupt service routine at privilege level 0 (i.e., transition from V86 to protected mode occurs) and the VM bit in the EFLAGS register is cleared. The protected mode interrupt service routine then determines if the interrupt came from a protected mode or V86 application by examining the VM bit in the EFLAGS image stored on the stack. The interrupt service routine may then choose to allow the 8086 operating system to handle the interrupt or may emulate the function of the interrupt handler. Following completion of the interrupt service routine, an IRET instruction restores the EFLAGS register (restores VM=1) and segment selectors and control returns to the interrupted V86 task.

2.18.4 Entering and Leaving V86 Mode

V86 mode is entered from protected mode by either executing an IRET instruction at CPL = 0 or by task switching. If an IRET is used, the stack must contain an EFLAGS image with VM = 1. If a task switch is used, the TSS must contain an EFLAGS image containing a 1 in the VM bit position. The POPF instruction cannot be used to enter V86 mode since the state of the VM bit is not affected. V86 mode can only be exited as the result of an interrupt or exception. The transition out must use a 32-bit trap or interrupt gate which must point to a non-conforming privilege level 0 segment (DPL = 0), or a 32-bit TSS. These restrictions are required to permit the trap handler to IRET back to the V86 program.

2.19 Floating Point Unit Operations

The 6x86MX CPU includes an on-chip FPU that provides the user access to a complete set of floating point instructions (see Chapter 6). Information is passed to and from the FPU using eight data registers accessed in a stack-like manner, a control register, and a status register. The IBM 6x86MX CPU also provides a data register tag word which improves context switching and performance by maintaining empty/non-empty status for each of the eight data registers. In addition, registers in the CPU contain pointers to (a) the memory location containing the current instruction word and (b) the memory location containing the operand associated with the current instruction word (if any).

FPU Tag Word Register. The IBM 6x86MX CPU maintains a tag word register (Figure 2-40 (Page 2-87)) comprised of two bits for each physical data register. Tag Word fields assume one of four values depending on the contents of their associated data registers, Valid (00), Zero (01), Special (10), and Empty (11). Note: Denormal, Infinity, QNaN, SNaN and unsupported formats are tagged as “Special”. Tag values are maintained transparently by the IBM 6x86MX CPU and are only available to the programmer indirectly through the FSTENV and FSAVE instructions.

FPU Control and Status Registers. The FPU circuitry communicates information about its status and the results of operations to the programmer via the status register. The FPU status register is comprised of bit fields that reflect exception status, operation execution status, register status, operand class, and comparison results. The FPU status register bit definitions are shown in Figure 2-41 (Page 2-87) and Table 2-41 (Page 2-87).

The FPU Mode Control Register (MCR) is used by the CPU to specify the operating mode of the FPU. The MCR contains bit fields which specify the rounding mode to be used, the precision by which to calculate results, and the exception conditions which should be reported to the CPU via traps. The user controls precision, rounding, and exception reporting by setting or clearing appropriate bits in the MCR. The FPU mode control register bit definitions are shown in Figure 2-42 (Page 2-88) and Table 2-42 (Page 2-88).

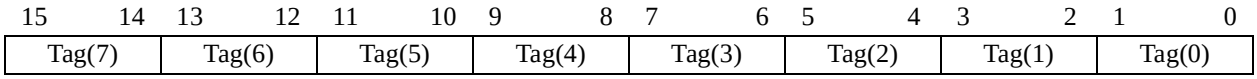


Figure 2-40. FPU Tag Word Register

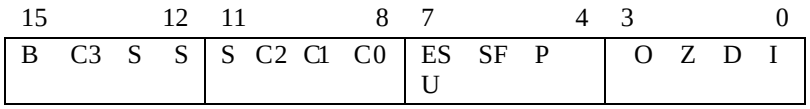


Figure 2-41. FPU Status Register

Table 2-41. FPU Status Register Bit Definitions

BIT POSITION	NAME	DESCRIPTION
15	B	Copy of the ES bit. (ES is bit 7 in this table.)
14, 10 - 8	C3 - C0	Condition code bits.
13 - 11	SSS	Top of stack register number which points to the current TOS.
7	ES	Error indicator. Set to 1 if an unmasked exception is detected.
6	SF	Stack Fault or invalid register operation bit.
5	P	Precision error exception bit.
4	U	Underflow error exception bit.
3	O	Overflow error exception bit.
2	Z	Divide by zero exception bit.
1	D	Denormalized operand error exception bit.
0	I	Invalid operation exception bit.

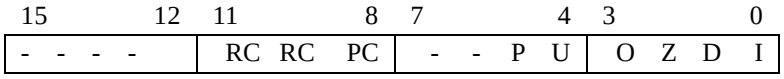


Figure 2-42. FPU Mode Control Register

Table 2-42. FPU Mode Control Register Bit Definitions

BIT POSITION	NAME	DESCRIPTION
11 - 10	RC	Rounding Control bits: 00 Round to nearest or even 01 Round towards minus infinity 10 Round towards plus infinity 11 Truncate
9 - 8	PC	Precision Control bits: 00 24-bit mantissa 01 Reserved 10 53-bit mantissa 11 64-bit mantissa
5	P	Precision error exception bit mask.
4	U	Underflow error exception bit mask.
3	O	Overflow error exception bit mask.
2	Z	Divide by zero exception bit mask.
1	D	Denormalized operand error exception bit mask.
0	I	Invalid operation exception bit mask.

2.20 MMX Operations

The IBM 6x86MX CPU provides user access to the MMX instruction set. MMX data is configured in one of four MMX data formats. During operations eight 64-bit MMX registers are utilized.

2.20.1 MMX Data Formats

The MMX instructions operate on 64-bit data groups called “packed data.” A single packed data group can be interpreted as a:

- Packed byte (8 bytes)
- Packed word (4 words)
- Packed doubleword (2 doublewords)
- Quadword (1 quadword)

The packed data types supported are signed and unsigned integer.

2.20.2 MMX Registers

The MMX instruction set operates on eight 64-bit, general-purpose registers (MM0-MM7). These registers are overlayed with the floating point register stack, so no new architectural state is defined by the MMX instruction set. Existing mechanisms for saving and restoring floating point state automatically work for saving and restoring MMX state.

2.20.3 MMX Instruction Set

The MMX instructions operate on all the elements of a signed or unsigned packed data group. All data elements (bytes, words, doublewords or a quadword) are operated on separately in parallel. For example, eight bytes in one packed data group can be added to another packed data group, such that eight independent byte additions are performed in parallel.

2.20.4 Instruction Group Overview

The 57 MMX instructions are grouped into seven categories:

- Arithmetic Instructions
- Comparison Instructions
- Conversion Instructions
- Logical Instructions
- Shift Instructions
- Data Transfer Instructions
- Empty MMX State (EMMS) Instruction

2.20.5 Saturation Arithmetic

For saturating MMX instructions, a ceiling is placed on an overflow and a floor is placed on an underflow. When the result of an operation exceeds the range of the data-type it saturates to the maximum value of the range. Conversely, when a result that is less than the range of a data type, the result saturates to the minimum value of the range.

The saturation limits are shown in Table 2-43.

MMX instructions do not indicate overflow or underflow occurrence by generating exceptions or setting flags.

Table 2-43. Saturation Limits

DATA TYPE	LOWER LIMIT		UPPER LIMIT	
Signed Byte	80h	-128	7Fh	127
Signed Word	8000h	-32,768	7FFFh	32,767
Unsigned Byte	00h	0	FFh	256
Unsigned Word	0000h	0	FFFFh	65,535

2.20.6 EMMS Instruction

The EMMS Instruction clears the TOS pointer and sets the entire FPU tag word as empty. An EMMS instruction should be executed at the end of each MMX routine.

[redacted]